



**HAL**  
open science

## Identification de nouvelles bases moléculaires des cancers précoces par séquençage à haut débit.

Pierre Fermey

► **To cite this version:**

Pierre Fermey. Identification de nouvelles bases moléculaires des cancers précoces par séquençage à haut débit.. Génétique humaine. Normandie Université, 2017. Français. NNT : 2017NORMR110 . tel-01744221

**HAL Id: tel-01744221**

**<https://theses.hal.science/tel-01744221v1>**

Submitted on 27 Mar 2018

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

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



Normandie Université

## THÈSE

**Pour obtenir le diplôme de doctorat**

**Spécialité Aspects Moléculaires et Cellulaires de la Biologie**

**Préparée au sein de l'Université de Rouen Normandie**

**Identification de nouvelles bases moléculaires  
des cancers précoces par séquençage à haut-débit.**

**Présentée et soutenue par**

**Pierre FERMEY**

**Thèse soutenue publiquement le 13 décembre 2017  
devant le jury composé de**

Mme Marie-Pierre BUISINE	PU-PH, Université de Lille 2	Rapporteur
M. Nicolas SEVENET	PU-PH, Université de Bordeaux	Rapporteur
M. Martin FIGEAC	IR, Université de Lille 2	Examinateur
Mme Catherine BAUGÉ	MCU, Université de Caen Normandie	Examinateur
Mme Isabelle TOURNIER	MCU, Université de Rouen Normandie	Co-encadrante de thèse
M. Thierry FREBOURG	PU-PH, Université de Rouen Normandie	Directeur de thèse
Mme Hélène CASTEL	DR, Université de Rouen Normandie	Membre invité

**Thèse dirigée par le Professeur Thierry FREBOURG, Unité Inserm U1245**







Normandie Université

# THÈSE

**Pour obtenir le diplôme de doctorat**

**Spécialité Aspects Moléculaires et Cellulaires de la Biologie**

**Préparée au sein de l'Université de Rouen Normandie**

## **Identification de nouvelles bases moléculaires des cancers précoces par séquençage à haut-débit.**

**Présentée et soutenue par**

**Pierre FERMEY**

**Thèse soutenue publiquement le 13 décembre 2017  
devant le jury composé de**

Mme Marie-Pierre BUISINE	PU-PH, Université de Lille 2	Rapporteur
M. Nicolas SEVENET	PU-PH, Université de Bordeaux	Rapporteur
M. Martin FIGEAC	IR, Université de Lille 2	Examinateur
Mme Catherine BAUGÉ	MCU, Université de Caen Normandie	Examinateur
Mme Isabelle TOURNIER	MCU, Université de Rouen Normandie	Co-encadrante de thèse
M. Thierry FREBOURG	PU-PH, Université de Rouen Normandie	Directeur de thèse
Mme Hélène CASTEL	DR, Université de Rouen Normandie	Membre invité

**Thèse dirigée par le Professeur Thierry FREBOURG, Unité Inserm U1245**



## Résumé

Une des plus grandes avancées en cancérologie et en génétique au cours des vingt dernières années fût l'identification des formes héréditaires de cancer et des gènes de prédisposition impliqués. Chez une majorité de patients soupçonnés de présenter une forme héréditaire de cancer, les analyses centrées sur les gènes connus pour être impliqués dans les prédispositions mendéliennes au cancer restent bien souvent négatives. Aujourd'hui, grâce à l'émergence du séquençage à haut-débit (NGS), il est possible de séquencer l'ensemble des exons (exome) d'un individu ou plusieurs centaines de gènes dans un laps de temps court et à des coûts raisonnables. Dans ce contexte, nous avons appliqué plusieurs stratégies basées sur ces nouveaux outils, avec l'objectif **d'identifier de nouvelles bases moléculaires des cancers héréditaires à survenue précoce**.

Tout d'abord, nous avons employé une **stratégie d'analyse exomique intrafamiliale** dans une famille atypique présentant des chondrosarcomes de localisation thoracique pour lesquels aucune base moléculaire n'avait pu être mise en évidence. Grâce à cette stratégie, nous avons pu identifier une altération tronquante du gène *EXT2* (NM\_000401.3; c.237G>A; p.Trp79\*). Les altérations perte de fonction documentées pour ce gène sont impliquées dans la maladie des ostéochondromes multiples (OM), des tumeurs bénignes. Or, dans cette famille, aucun signe clinique d'OM n'était présent. Ces travaux nous ont donc permis d'étendre le spectre phénotypique des mutations *EXT2* et de modifier la prise en charge clinique de cette famille.

Nous avons ensuite employé une **stratégie d'analyse exomique soustractive** de trio enfant malade / parents sains dans le but d'identifier des mutations *de novo* potentiellement responsables de la prédisposition génétique au cancer observée chez un jeune patient ayant développé un médulloblastome du cervelet à l'âge de 8 ans, suivi d'un méningiome à 22 ans. L'analyse exomique du trio a révélé l'existence chez ce patient d'une mutation *de novo* faux-sens affectant un acide aminé très conservé de la protéine HID-1. Cette dernière est particulièrement exprimée dans les cellules neuronales et sécrétrices, et semble fonctionner autour de l'appareil de Golgi pour réguler le tri des vésicules nouvellement formées. Ainsi, notre hypothèse est qu'un défaut de la protéine HID-1, lié à une mutation du gène *HID-1*, perturberait la voie de sécrétion et participerait à la genèse du médulloblastome. Ces travaux, toujours en cours, démontrent à la fois la force de la stratégie exomique de trio pour identifier rapidement des mutations *de novo* et illustre toute la difficulté d'interprétation des variants détectés dans des gènes non impliqués dans le cancer.

Par ailleurs, nous avons appliqué une stratégie **exomique soustractive et interfamiliale** à une cohorte de dix patients ayant développé un corticosurréalome à un âge très précoce et pour lesquels aucune base moléculaire n'a pu être mise en évidence. Malheureusement, nous n'avons pas pu identifier de nouvelles bases moléculaires du corticosurréalome de l'enfant par ces techniques.

Enfin, sous l'hypothèse que des mutations rares ou privées dans un nombre limité de gènes impliqués dans le cancer contribueraient à des formes héréditaires de cancer, nous avons entrepris un projet visant à **séquencer à haut débit 201 gènes fortement impliqués dans le cancer** chez des patients ayant développé des tumeurs à un âge pédiatrique. Les premiers résultats de ce projet toujours en cours ont permis de confirmer la robustesse de cette technique et suggèrent une extension phénotypique du spectre des mutations *DICER1* ainsi qu'une contribution oligogénique des gènes de réparation de l'ADN dans les tumeurs pédiatriques. L'ensemble de ces résultats seront bientôt compilés au sein d'une base de données et bénéficieront d'une analyse statistique fine avec l'objectif d'identifier des enrichissements en variants rares dans des gènes ou voies biologiques.

Mots-clés : Séquençage à haut-débit, Cancer héréditaire, Tumeurs pédiatriques, Séquençage de l'exome, Séquençage de panels de gènes.



## Abstract

One of the greatest advances in oncology and genetics over the past 20 years has been the identification of hereditary forms of cancer and of the cancer genes. Nevertheless, in a majority of patients suspected to present an inherited form of cancer, analyses of the genes known to be involved in the Mendelian predispositions to cancer often remain negative. Today, thanks to the emergence of high-throughput sequencing (NGS), it is now possible to sequence all exons of an individual (exome) or several hundred genes in a short period of time and for a reasonable cost. In this context, we have applied several strategies based on these new tools in order **to identify new molecular basis of early-onset cancers.**

First, we applied an **intra-familial exome analysis strategy** to an atypical family with chondrosarcomas of the chest, for which no molecular basis could be identified. Using this strategy, we were able to identify a truncating alteration of the *EXT2* gene (NM\_000401.3; c.237G> A; p.Trp79 \*). The documented loss of function alterations of this gene are implicated in a disease called multiple osteochondromas (OM), associated with benign lesions. Interestingly, these patients showed no clinical signs of OM indicating a potential phenotypic extension of *EXT2* mutations. Plus, this work allowed us to change the clinical management of this family.

We then used a **strategy of subtractive exomic analysis of trio** sick child/healthy parents in order to identify *de novo* mutations in a young patient who developed a medulloblastoma of the cerebellum at 8 years-old followed by a meningioma at 22 years-old. The analysis of the trio revealed the existence of a *de novo* mutation affecting a highly conserved amino acid of the *HID-1* protein. *HID-1* is specifically expressed in neuronal and secretory cells, and seems to function around the Golgi apparatus to regulate the sorting of newly formed vesicles. Our hypothesis is that a defect of the *HID-1* protein linked to a mutation of the *HID-1* gene, could alter the secretory pathway therefore contributing to the development of the tumor. This work, which is still ongoing, demonstrates ~~both~~ the strength of the trio strategy for the rapid identification of *de novo* mutations and illustrates all the difficulty of interpreting variants detected in genes not yet involved in cancer.

Then, thanks to the recruitment of the Laboratory of Molecular Genetics of the CHU of Rouen, we have collected a cohort of 10 patients who developed an adrenocortical carcinoma (ACC) at a very early age and for which no molecular basis could be identified. Despite **subtractive and inter-familial exomic analyses**, we were unable to highlight new molecular bases for these cases of pediatric ACC.

Finally, under the assumption that rare or private mutations in a limited number of genes involved in cancer could contribute to inherited forms of cancer, we undertook a project to **sequence 201 genes involved in cancer** in patients who developed tumors at a pediatric age. The first results of this project confirmed the robustness of this technique and suggested a phenotypic extension of the *DICER1* mutation spectrum as well as an oligogenic contribution of DNA repair genes in pediatric tumors. Soon, these results will be compiled in a database and will benefit from a statistical analysis with the objective to identify enrichment of rare variants in specific genes or biological pathways in these patients compared to control individuals.

Keywords: High-throughput sequencing, Hereditary cancer, Pediatric tumors, Exome sequencing, Gene panel sequencing.





## Liste des abréviations

Par ordre d'alphabétique :

ADN	Acide DésoxyriboNucléique
BAC	<i>Bacterial Artificial Chromosome</i>
BD	<i>Broadly Damaging</i>
BER	<i>Base Excision Repair</i>
BSA	<i>Bovine Serum Albumin</i>
CCD	<i>Charge-Coupled Device</i>
CES	<i>Clinical Exome Sequencing</i>
CgA	Chromogranine A
CHU	Centre Hospitalo-Universitaire
CMT	Cancer Médullaire de la Thyroïde
CNV	<i>Copy Number Variation</i>
cPAL	<i>combinatorial Probe-Anchor Ligation</i>
CPG	<i>Cancer Predisposition Genes</i>
CRT	<i>Cyclic Reversible Termination</i>
CS	ChondroSarcome
DCV	<i>Dense Core Vesicle</i>
DFNA11	<i>DeaFNess Autosomal dominant 11</i>
DP	<i>Depth (Profondeur)</i>
DSD	<i>Disruptive and Strictly Damaging</i>
EGL	<i>External Granule Layer</i>
EMBL	<i>European Molecular Biology Laboratory</i>
emPCR	<i>emulsion PCR</i>
EMT	<i>Epithelial to Mesenchymal Transition</i>
ESP	<i>Exome Sequencing Project</i>
ExAC	<i>Exome Aggregation Consortium</i>
FISSEQ	<i>Fluorescent In Situ SEQuencing</i>
FREX	<i>FRench EXome project</i>
Gb	Gigabase
GCP	<i>Granule Cell Precursors</i>
GWAS	<i>Genome-Wide Association Study</i>
HBOC	<i>Hereditary Breast and Ovarian Cancer</i>
HGP	<i>Human Genome Project</i>
HNPCC	<i>Hereditary Non-Polyposis Colorectal Cancer</i>
HR	<i>Homologous Repair</i>
HRDSB	<i>Homologous Recombination of Double Strand DNA Breaks repair</i>
HTS	<i>High-Throughtput Sequencing</i>
IGL	<i>Internal Granule Layer</i>
kb	kilobase
LFS	<i>Li-Fraumeni Syndrome</i>
LMMJ	Leucémie MyéloMonocytaire Juvénile

MAF	<i>Minor Allele Frequency</i>
MET	<i>Mesenchymal to Epithelial Transition</i>
MIP	<i>Molecular Inversion Probe</i>
MMR	<i>MisMatch Repair</i>
MSI	<i>MicroSatellite Instability</i>
NEM1	<i>Néoplasie Endocrinienne Multiple de type 1</i>
NEM2	<i>Néoplasie Endocrinienne Multiple de type 2</i>
NER	<i>Nucleotide Excision Repair</i>
ng	<i>nanogramme</i>
NGS	<i>Next Generation Sequencing</i>
NHEJ	<i>NonHomologous End-Joining</i>
OM	<i>Ostéochondromes Multiples</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
ONT	<i>Oxford Nanopore Technologies</i>
PARP	<i>poly(ADP-ribose) polymerase</i>
pb	<i>paire de base</i>
PCR	<i>Polymerase Chain Reaction</i>
PPi	<i>PyroPhosphate inorganique</i>
PTV	<i>Protein Truncating Variant</i>
QMPSF	<i>Quantitative Multiplex PCR of Short Fluorescent fragments</i>
RCA	<i>Rolling Circle Amplification</i>
SBL	<i>Sequencing By Ligation</i>
SBS	<i>Sequencing By Synthesis</i>
SD	<i>Strictly Damaging</i>
SHH	<i>Sonic Hedgehog</i>
SMRT	<i>Single Molecule Real Time</i>
SMS	<i>Single Molecule Sequencing</i>
SNA	<i>Single Nucleotide Addition</i>
SNP	<i>Single Nucleotide Polymorphism</i>
SoLiD	<i>Sequencing by oligonucleotide Ligation and Detection</i>
TCGA	<i>The Cancer Genome Atlas</i>
TIRF	<i>Total Internal Reflection Fluorescence</i>
UTR	<i>UnTranslated Region</i>
VSI	<i>Variant de Signification Inconnue</i>
WES	<i>Whole Exome Sequencing</i>
WGS	<i>Whole Genome Sequencing</i>
WXS	<i>Whole eXome Sequencing</i>
ZMW	<i>Zero-Mode Waveguides</i>

# Remerciements

En premier lieu, je tiens sincèrement à remercier l'ensemble des membres de mon jury : Madame Marie-Pierre Buisine, Madame Hélène Castel, Madame Catherine Baugé, Monsieur Martin Figeac et Monsieur Nicolas Sevenet. Merci pour votre investissement et de me faire l'honneur d'avoir accepté de juger ce travail.

J'adresse mes remerciements particuliers au Pr Thierry Frébourg, directeur de l'unité Inserm U1245, pour m'avoir accepté au sein du laboratoire depuis 2011, en tant que stagiaire de Master 1 jusqu'à aujourd'hui en tant que doctorant. Merci d'avoir tout mis en œuvre pour que mon passage dans l'unité se déroule dans des conditions optimales tant sur le plan matériel que sur le plan de l'encadrement. Merci sincèrement pour votre engagement et pour vos conseils précieux.

Isabelle, je tiens tout particulièrement à te témoigner ici toute ma reconnaissance à la fois pour ton encadrement et pour ta confiance depuis si longtemps. Je me rappellerai toujours pourquoi j'ai décidé de faire mon stage de Master à tes côtés : outre les aspects attractifs du stage et de l'unité, tu as pris le temps de m'expliquer chaque détail du projet et de toujours t'assurer de ma bonne compréhension. C'est cette bienveillance et cette générosité qui m'ont tout de suite attiré et qui d'ailleurs ont perduré durant tout le long de ma thèse. Tout au long des années passées ensemble, tu m'as toujours épaulé, poussé dans mes retranchements et guidé. Je suis réellement désolé pour toi d'être tombé sur une tête dure comme moi comme premier thésard mais sache que j'ai énormément appris avec et grâce à toi. Tes qualités humaines et scientifiques font de toi une personne que l'on ne peut pas oublier. Merci sincèrement pour ton investissement professionnel et personnel à mon égard. J'espère réellement que la suite fera que nos chemins se croiseront, qu'il s'agisse de situations professionnelles ou d'autres moments. As-tu remarqué que j'ai réussi à ne pas parler de bleu de Coomassie, ni de montage de meubles ? Oups...

Françoise, j'aimerais t'adresser un clin d'œil particulier dans ces remerciements. Je te remercie sincèrement pour l'ensemble des conseils scientifiques, du temps que tu m'as consacré pour m'apprendre sans concession tout ce que tu sais et de m'avoir transmis un certain goût de la résolution de problème ! Au-delà des nombreuses discussions scientifiques nourries que nous avons eu, j'aimerais que tu saches à quel point j'apprécie ton amitié et les conseils sur la vie en général que tu as su me prodiguer. Je souhaite autant de complicité à tous les étudiants qui te croiseront. Merci sincèrement.

Sophie, merci pour ton aide sur l'ensemble des projets notamment en faisant en sorte que ces fameux fichiers FASTQ terminent en fichier de variants comme par magie ! L'ensemble de mon travail n'aurait pas été possible sans ton aide et ta rigueur. Heureusement pour moi que tu es particulièrement organisée et accessible ! Tu as toujours le sourire y compris quand inlassablement, je viens t'embêter pour que tu m'expliques tel ou tel pipeline ! Merci beaucoup, c'était un réel plaisir de travailler avec toi.

Gaëlle et Stéphanie, ces projets sont également les vôtres. Vous m'avez permis, grâce à votre organisation sans faille et votre rigueur, d'avoir accès aux cohortes de patients. Mais au-delà de cet aspect pratique, vous m'avez fait confiance et avez toujours répondu présentes pour m'aider. Merci sincèrement pour tout cela.

Ma chère Sabine, je tiens à t'adresser ici mes remerciements les plus sincères. Je me considère comme particulièrement chanceux d'avoir fait ta connaissance, tellement tu es une vraie bulle d'air frais à chaque instant ! Durant cette thèse, tu as été d'une aide plus que précieuse. Tu es toujours volontaire et présente pour aider les gens qu'il s'agisse de travail ou à l'extérieur. Tu as été ma vraie première "stagiaire" et je dois avouer que j'ai lié avec toi une relation bien plus que professionnelle, presque fraternelle. Je voudrais te remercier du fond du cœur pour m'avoir fait confiance, pour ta présence sans faille et pour tous les fous-rires que nous

avons eu ensemble. Tu es travailleuse, rigoureuse, drôle, et par-dessus-tout tu es une vraie gentille. Je te souhaite le meilleur pour la suite chef car c'est ce que tu mérites !

Mon cher Omar, je ne pourrai écrire des remerciements sans que tu y figures. Ta gentillesse, ta générosité, ta sensibilité, ton sens du travail sont des qualités que j'ai appréciées durant ma thèse et que j'apprécie aujourd'hui au quotidien en dehors du laboratoire. Tu es un excellent camarade de travail, de congrès et de sorties également. Au fil des années, tu es devenu un ami fidèle et je t'en remercie vraiment (et non je n'écris pas cela pour obtenir surnoisement des kilos d'épices libanaises...). Merci mon ami, merci pour ce que tu es et ce que tu m'apportes !

Mon cher Hafid, merci pour les plus nombreuses discussions nourries autour de l'entrepreneuriat, sujet qui nous anime tous les deux. Merci pour ta sagesse et pour réussir constamment à faire ressortir le meilleur des gens (je t'envie jalousement cette qualité). Je ne doute pas que nous garderons contact mais surtout que tu iras loin !

J'aimerais également remercier mes collègues de bureau. Vous avez chacun un quelque chose de particulier qui fait que vous êtes uniques et tellement différents. Ma chère Edwige, tu es quelqu'un d'extrêmement volontaire et organisé. J'admire ta détermination, les efforts que tu déploies quand un sujet te tient à cœur et bien-sûr aussi ta "bonhomie". Mon cher Alexandre, merci pour ta gentillesse et ta passion. Ce sont des qualités tellement nobles qu'il ne faut pas que tu changes. Et au passage, je te confesse que je ne détiens pas le record de bilboquet du bureau. Ma chère Estelle, merci d'être si discrète et en même temps d'avoir un rire si communicatif. Ma chère Marine, je suis désolé mais tu resteras pour moi la nouvelle même si d'autres sont arrivés depuis. Merci d'avoir une énergie si communicative, ton peps est un véritable remontant au quotidien. Ma chère Hélène, ton organisation et ta capacité à abattre autant de travail en un laps de temps aussi court m'impressionneront toujours. Mon cher Kilan, merci pour nos discussions pertinentes autour d'un café/thé. Ma chère Ludivine, merci pour ta bonne humeur constante et tes moelleux chocolats.

Je vous souhaite à tous la meilleure des réussites tant sur le plan professionnel que personnel. N'oubliez jamais que vous êtes chacun magique à votre façon !

Dans le désordre : Myriam, Raphael, Olivier, Camille, Anne, Alexandra, Jean-Batiste, Magalie, Pascale, Aurélie, Sébastien, Stéphane, Yasmine, Chirine, Dominique, Anne-Claire, Laetitia, même si nous avons peu d'interaction directe ensemble, vous avez toujours été bienveillants à mon égard et prêts à m'apporter vos expertises respectives quand j'en avais besoin. Merci d'être des gens passionnés et de transmettre tout ça vigoureusement ! Vos énergies sont contagieuses ! J'ai également une pensée pour les anciens doctorants devenus docteurs qui ne sont plus au laboratoire. Merci Popo, Momo, Juliette et Cyril !

Chère Nathalie, merci d'être là pour soigner nos petits problèmes quotidiens qu'il s'agisse de commandes, de stylos, de trombones, d'ordres de mission. En plus, d'être d'une efficacité redoutable, tu arbores toujours un large sourire bienveillant.

Merci à vous les copains. Sachez que je me considère comme particulièrement chanceux d'être si bien entouré. Vous avez tous fait quelque chose qui, un jour, a changé ma vie sans que vous le sachiez peut-être : une discussion, une morale, un rire, un regard, l'annonce d'une grossesse ou d'un achat immobilier... Merci pour tout ça, sachez que je suis tellement fier de vous !

Je tiens sincèrement à remercier mes parents et ma famille. Merci d'avoir toujours cru en moi et encouragé inlassablement. Sachez que vous êtes une véritable source d'inspiration au quotidien. Merci des sacrifices pour vous avez fait pour moi / nous, je vous en suis profondément reconnaissant. Merci à mes deux petits frères, d'être aussi géniaux et inspirants chacun à leurs façons. Je vous aime.

Inès, ma moitié, merci pour ta sensibilité, ta générosité, ta détermination, ta tendresse et ta présence infaillible. Je suis particulièrement admiratif et fier de ce que tu entreprends au quotidien et mène à bout dans ta vie professionnelle et personnelle. Merci de me laisser partager

ton quotidien. Chaque jour passé avec toi est une grande source d'épanouissement et de bonheur, sache-le !

Enfin, merci à toi petit lecteur curieux qui est en train de lire mes remerciements. On ne se connaît peut-être pas, mais si tu as ouvert ce document t'es forcément quelqu'un de bien !



*" Any sufficiently advanced technology is indistinguishable from magic"*

Arthur C. Clarke





# TABLE DES MATIERES

<b>CHAPITRE I :</b>	<b>25</b>
<b>LES BASES MOLECULAIRES DES PREDISPOSITIONS HEREDITAIRES AU CANCER</b>	<b>25</b>
I. LE CANCER : UNE MALADIE GENETIQUE TRES PARTICULIERE	26
II. LES BASES MOLECULAIRES DES PREDISPOSITIONS HEREDITAIRES AU CANCER	29
a. Définition d'un gène de prédisposition héréditaire au cancer	29
b. Catégories de gènes prédisposant au cancer	30
c. L'oncogénétique en pratique	35
d. Méthodes d'identification des CPG	38
III. L'EMERGENCE DE NOUVEAUX CONCEPTS	40
a. Le nombre de gènes prédisposant au cancer serait limité	40
b. Parallélisme entre les gènes du développement et les gènes de l'oncogénèse	42
c. Oligogénisme/Polygénisme	44
IV. DU DIAGNOSTIC MOLECULAIRE AU DEVELOPPEMENT DE NOUVELLES THERAPIES	44
a. Déroulement des consultations d'oncogénétique	44
b. Prise en charge médicale : exemple des familles avec mutation des gènes BRCA	45
c. Les nouveaux traitements basés sur les inhibiteurs de PARP	47
<b>CHAPITRE II :</b>	<b>49</b>
<b>LA REVOLUTION DU SEQUENÇAGE DE NOUVELLE GENERATION</b>	<b>49</b>
I. DU SEQUENÇAGE SANGER AU SEQUENÇAGE DU GENOME HUMAIN	50
II. LE SEQUENÇAGE DE NOUVELLE GENERATION	53
a. L'amplification clonale : une étape nécessaire au séquençage de nouvelle génération	53
b. Le Séquençage Nouvelle Génération (NGS)	56
c. La révolution du séquençage de molécule unique : The next-next generation sequencing	67
d. Evolution rapide des technologies et des coûts de séquençage	70
III. DEVELOPPEMENT DE TECHNOLOGIES D'ENRICHISSEMENT	75
a. Enrichissement par hybridation	75
b. Enrichissement par PCR amplicon	77
c. Enrichissement par Molecular Inversion Probe (MIP)	78
d. Comparaison de l'enrichissement par amplicon et par hybridation	80
<b>CHAPITRE III :</b>	<b>81</b>

## LES ENSEIGNEMENTS ET APPLICATIONS DU SEQUENÇAGE NOUVELLE

<b>GENERATION</b> .....	<b>81</b>
I. LES ENSEIGNEMENTS DU NGS : LES CHANGEMENTS DE PARADIGMES .....	82
a. <i>La variabilité insoupçonnée du génome humain</i> .....	83
b. <i>Taux de variations de novo</i> .....	84
c. <i>Notions de singletons et de variants rares</i> .....	84
d. <i>La tolérance aux variants perte de fonction</i> .....	85
II. APPLICATION DU NGS : LE SEQUENÇAGE DE PANELS DE GENES .....	89
a. <i>Intérêts du séquençage de panel de gènes</i> .....	89
b. <i>L'utilisation des panels de gènes dans le contexte du diagnostic des prédispositions héréditaires aux cancers</i> .....	90
c. <i>Applications du séquençage de panels de gènes : paradigme des prédispositions héréditaires aux cancers du sein et de l'ovaire</i> .....	92
d. <i>Les variations de signification biologique inconnue</i> .....	96
III. LE SEQUENÇAGE EXOMIQUE ET SES APPLICATIONS .....	98
a. <i>Intérêts de l'exome</i> .....	98
b. <i>Stratégies d'analyses employant l'exome</i> .....	99
c. <i>L'exome comme outil diagnostique</i> .....	105
d. <i>Analyse de panel in silico à partir de l'exome</i> .....	107
e. <i>Remarques sur l'utilisation de l'exome et de ses approches</i> .....	107
<b>PROJET 1 : IDENTIFICATION PAR LA STRATEGIE D'ANALYSE EXOMIQUE</b>	
<b>INTRAFAMILIALE DES BASES MOLECULAIRES A L'ORIGINE D'UNE AGREGATION</b>	
<b>FAMILIALE ATYPIQUE DE CHONDROSARCOMES COSTAUX SOLITAIRES</b> .....	<b>117</b>
I. CONTEXTE SCIENTIFIQUE .....	119
II. RESULTATS : ARTICLE .....	120
III. DISCUSSION .....	127
<b>PROJET 2 : IDENTIFICATION DE MUTATIONS DE NOVO PAR ANALYSE EXOMIQUE</b>	
<b>DE TRIO CHEZ UN JEUNE PATIENT AYANT DEVELOPPE SPORADIQUEMENT UN</b>	
<b>MEDULLOBLASTOME ET UN MENINGIOME</b> .....	<b>129</b>
I. CONTEXTE SCIENTIFIQUE DE L'ETUDE .....	130
II. MATERIELS ET METHODES .....	131
III. RESULTATS .....	137
IV. DISCUSSION .....	148
<b>PROJET 3 : RECHERCHE DE NOUVELLES BASES MOLECULAIRES DU</b>	
<b>CORTICOSURRENALOME DE L'ENFANT PAR ANALYSE D'EXOMES</b> .....	<b>151</b>
I. CONTEXTE .....	152
II. MATERIELS ET METHODES .....	152

III. RESULTATS .....	155
IV. CONCLUSION ET DISCUSSION .....	161
<b>PROJET 4 : RECHERCHE DE NOUVELLES BASES MOLECULAIRES DES TUMEURS PEDIATRIQUES PAR SEQUENÇAGE A HAUT-DEBIT DE 201 GENES IMPLIQUES DANS LE CANCER. ....</b>	<b>165</b>
I. CONTEXTE ET OBJECTIFS DE L'ETUDE .....	166
II. MATERIEL ET METHODES .....	168
III. RESULTATS .....	174
IV. CONCLUSION, DISCUSSION, PERSPECTIVES .....	182
<b>DISCUSSION GENERALE ET PERSPECTIVES .....</b>	<b>186</b>
<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>194</b>



## Liste des figures

<b>Figure 1</b> : Différence entre un cancer sporadique et un cancer familial.....	28
<b>Figure 2</b> : Altérations des gènes impliqués dans le Cancer .....	40
<b>Figure 3</b> : Voies biologiques impliquées dans la cancérogenèse selon Vogelstein <i>et al</i> <sup>24</sup> .....	41
<b>Figure 4</b> : Le séquençage par la méthode de Sanger.....	51
<b>Figure 5</b> : L'amplication clonale à l'aide de la PCR émulsion.....	54
<b>Figure 6</b> : L'amplication clonale en pont sur support solide .....	55
<b>Figure 7</b> : Amplification clonale en phase liquide à l'aide d'une <i>Rolling Circle Amplification PCR</i> (RCA-PCR) .....	56
<b>Figure 8</b> : Système de codage du dibase employé dans la technologie SoLiD .....	57
<b>Figure 9</b> : Le séquençage SBL de la société SoLiD (Thermo Fisher).....	58
<b>Figure 10</b> : Le séquençage développé par Complete Genomics basé sur la technologie cPAL .....	59
<b>Figure 11</b> : Le séquençage basé sur l'utilisation de nucléotides terminateurs de séquences réversibles .....	60
<b>Figure 12</b> : Le séquençage à haut-débit par pyroséquençage.....	61
<b>Figure 13</b> : Le séquençage à haut-débit basé sur la variation de pH : l'ion Torrent.....	62
<b>Figure 14</b> : Séquençage de longs reads synthétiques d'Illumina : Moleculo.....	65
<b>Figure 15</b> : Le séquençage synthétique développé par 10X Genomics.....	66
<b>Figure 16</b> : Séquençage d'une seule molécule en temps réel (SMRT) de Pacific Biosciences (PacBio) .....	68
<b>Figure 17</b> : Séquençage par nanopore d'Oxford Nanopore Technologies (ONT).....	69
<b>Figure 18</b> : Photographie du MinION d'Oxford Nanopore Technologies.....	70
<b>Figure 19</b> : Images illustrant un séquençage FISSEQ .....	71
<b>Figure 20</b> : Graphique illustrant le coût d'un génome humain en fonction du temps en comparaison avec la conjecture de Moore ( <a href="http://genome.gov/sequencingcosts">http://genome.gov/sequencingcosts</a> ).....	73
<b>Figure 21</b> : Graphique illustrant la capacité de séquençage mondiale en térabases en fonction du temps .....	74
<b>Figure 22</b> : Enrichissement en régions d'intérêt par capture.....	76
<b>Figure 23</b> : Technique d'enrichissement par PCR amplicon .....	78
<b>Figure 24</b> : Technique d'enrichissement par sondes moléculaires inversées ( <i>Molecular Probes Inversion</i> ) .....	79
<b>Figure 25</b> : Nombre moyen de variations génétiques par exome .....	83
<b>Figure 26</b> : Distribution relative des variants détectés par séquençage de panel de gènes chez 708 patients présentant un cancer du sein et/ou de l'ovaire dans la cohorte Caennaise.....	95
<b>Figure 27</b> : Représentation des variants identifiés par l'équipe de Salvin par séquençage à haut-débit d'un panel de gènes consacré à l'étude des prédispositions héréditaires au cancer du sein et de l'ovaire.....	97
<b>Figure 28</b> : Stratégie d'analyse exomique intrafamiliale .....	100
<b>Figure 29</b> : Stratégie d'analyse exomique interfamiliale.....	102

<b>Figure 30</b> : Stratégie d'analyse exomique soustractive de trios .....	104
<b>Figure 31</b> : Génétique des chondrosarcomes .....	118
<b>Figure 32</b> : Identification d'une mutation du gène <i>EXT2</i> par la stratégie d'analyse exomique intra-familiale .....	119
<b>Figure 33</b> : Arbre génétique de la famille analysée et la stratégie exomique employée dans cette famille .....	131
<b>Figure 34</b> : La voie biologique Sonic Hedgehog et Wntless .....	132
<b>Figure 35</b> : Schéma des différentes constructions plasmidiques utilisées dans ce projet .....	135
<b>Figure 36</b> : Position du variant de novo dans le gène <i>MYO7A</i> identifié chez le patient analysé.....	137
<b>Figure 37</b> : Schéma représentant la localisation de la protéine <i>HID-1</i> au sein du trafic vésiculaire cellulaire .....	139
<b>Figure 38</b> : Schéma illustrant la communication moléculaire fine entre les différentes couches cellulaires du cervelet médiée par la molécule <i>SHH</i> pendant le développement et dans le contexte du médulloblastome .....	140
<b>Figure 39</b> : Arbre génétique du second patient présentant une altération constitutionnelle du gène <i>HID-1</i> (NM_030630.2; c.1532A>G; p.Glu511Gly).....	142
<b>Figure 40</b> : Analyse de l'effet des mutations sur l'expression des transcrits de <i>HID-1</i> par RT-PCR... 145	145
<b>Figure 41</b> : Impact des mutations sur l'expression protéique du gène <i>HID-1</i> .....	146
<b>Figure 42</b> : Analyse de l'impact des mutations G2N, E511G, I713M sur la localisation subcellulaire d' <i>HID-1</i> .....	147
<b>Figure 43</b> : Arbre génétique de la famille du cas index #01 et la stratégie exomique employée .....	156
<b>Figure 44</b> : Arbre génétique de la famille du cas index #02 et la stratégie exomique employée dans cette famille .....	158
<b>Figure 45</b> : Résultats de l'analyse interfamiliale et stratégie d'analyse employée par la suite .....	160
<b>Figure 46</b> : Etapes de sélection des 201 gènes impliqués dans le Cancer.....	169
<b>Figure 47</b> : Arbre génétique du patient #09 porteur de l'altération <i>SOS1</i> (NM_005633; exon21; c.3347-1G>A).....	177
<b>Figure 48</b> : Illustration des effets probables de l'altération <i>SOS1</i> identifiée .....	177
<b>Figure 49</b> : Confirmation de la mutation <i>SOS1</i> chez le patient et recherche de cette mutation chez les parents .....	178
<b>Figure 50</b> : Arbre génétique du patient analysé .....	178
<b>Figure 51</b> : Biogenèse et maturation des microARN .....	180

## Liste des tableaux

<b>Tableau 1</b> : Bases moléculaires des principales formes héréditaires de cancer.....	34
<b>Tableau 2</b> : Présentation des quatre portes d'entrée du syndrome de Li-Fraumeni d'après les critères de Chompret version réactualisée en 2015. ....	37
<b>Tableau 3</b> : Prise en charge des femmes porteuses d'une mutation d'un gène <i>BRCA</i> .....	47
<b>Tableau 4</b> : Les différents panels de gènes disponibles consacrés à l'étude des prédispositions héréditaires au cancer du sein et de l'ovaire.....	94
<b>Tableau 5</b> : Liste non exhaustive des études rapportant le pourcentage de mutations pathogènes et de variants de signification biologique inconnue identifiés par séquençage à haut-débit de panels de gènes. ....	96
<b>Tableau 6</b> : Récapitulatif des informations concernant les familles criblées dans le projet 3.....	153
<b>Tableau 7</b> : Liste des 201 gènes sélectionnés .....	170





# **INTRODUCTION**



## **CHAPITRE I :**

### **Les bases moléculaires des prédispositions héréditaires au cancer**

## I. Le Cancer : une maladie génétique très particulière

Le cancer est défini comme une maladie caractérisée par une prolifération cellulaire anormale au sein d'un tissu de l'organisme. Le terme cancer peut être utilisé de façon générique pour toute maladie dans laquelle certaines cellules se divisent d'une manière incontrôlée et inappropriée. Les nouvelles cellules résultantes forment alors un néoplasme. Dans le cas d'une tumeur maligne, ce néoplasme acquerra la propriété d'envahir les tissus adjacents et de se propager à travers le corps. Ce processus de transformation clonale est conduit par l'acquisition de nouvelles capacités comme :

- l'augmentation de la prolifération cellulaire et de la survie,
- la perte du contrôle du cycle cellulaire,
- l'acquisition d'un potentiel répliatif illimité,
- l'inhibition de l'apoptose et de la différenciation,
- l'acquisition de capacités angiogéniques et métastatiques,
- la perturbation du système immunitaire,
- la modification de l'environnement extracellulaire.

Ces nouvelles capacités sont acquises par la future cellule tumorale par une altération des gènes codant ces fonctions. Ainsi, d'un point de vue strictement moléculaire, le cancer est une maladie génétique puisqu'il survient en affectant les gènes quantitativement et qualitativement.

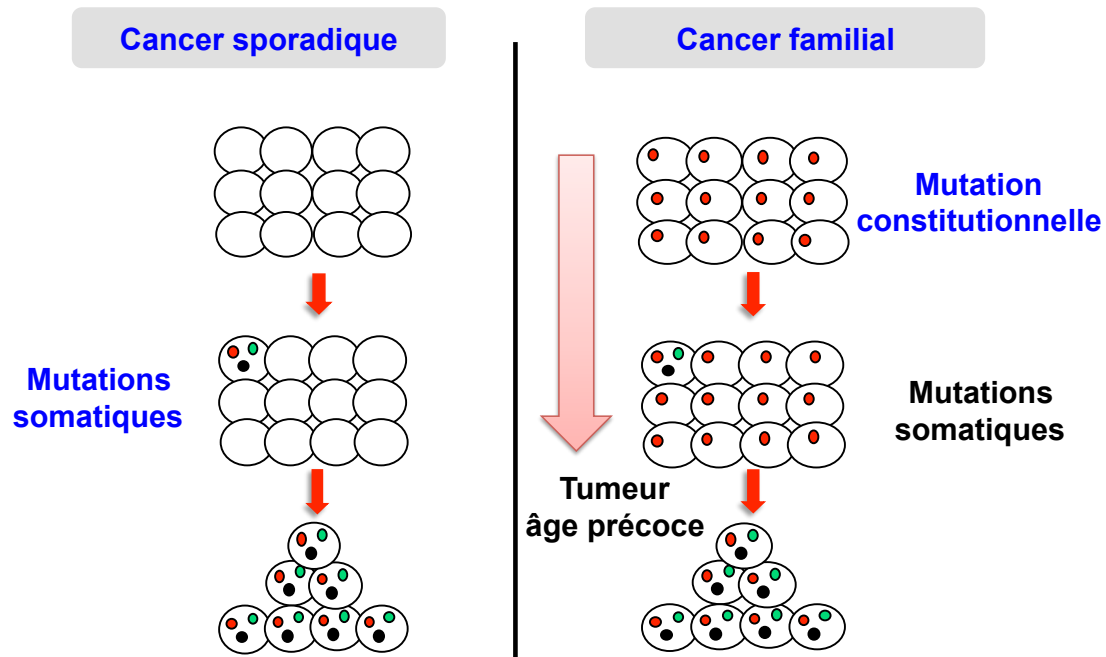
Le cancer est, sur le plan génétique, une maladie extraordinaire car si la transformation cancéreuse d'une cellule résulte toujours d'altérations génétiques, ces altérations sont, dans la majorité des cancers, restreintes aux cellules tumorales et ces altérations sont dites **somatiques** (*sôma* en grec : corps). Le cancer est donc une **maladie génétique somatique** et, sur le plan épidémiologique, le plus souvent sporadique, survenant de façon isolée dans une famille. Ainsi, le cancer correspond à la forme la plus extrême du mosaïcisme génétique, défini par la présence chez un individu de tissus génétiquement différents mais provenant du même zygote.

Les analyses extensives des exomes (ensemble des séquences codantes) de tumeurs réalisées par des approches de séquençage nouvelle génération ont révélé que le nombre d'altérations génétiques présentes dans une cellule tumorale était de l'ordre de la centaine. Une minorité de ces altérations jouent un rôle causal dans la transformation maligne : ces altérations sont qualifiées de **driver mutations** et le nombre de *driver mutations* nécessaires à la transformation maligne est de l'ordre de la dizaine. La plupart des altérations génétiques détectables dans le génome tumoral reflète simplement l'instabilité

génétique intrinsèque de la cellule cancéreuse. Ces mutations sont qualifiées de passagères ou de **passenger mutations**. La caractéristique essentielle de la cellule cancéreuse est en effet son **instabilité génétique** et c'est cette instabilité qui fait toute la gravité du cancer. Tel un génome viral ou bactérien soumis à la pression d'antiviraux ou d'antibiotiques, le génome tumoral sera progressivement sélectionné pour échapper au système immunitaire de l'hôte et aux contraintes mécaniques des tissus environnants, pour s'adapter aux différentes localisations tissulaires lors de l'essaimage métastatique et résister progressivement aux chimiothérapies. Cette instabilité explique l'**hétérogénéité génétique** de toute tumeur maligne, les différentes zones tumorales ne présentant pas toutes les mêmes altérations. Les analyses exomiques comparatives de différents territoires d'une tumeur permettent de reconstituer la séquence des événements mutationnels intervenant de façon causale dans la transformation maligne et d'établir ainsi une véritable **phylogénie mutationnelle tumorale**.

Si le cancer est le plus souvent une maladie génétique somatique de présentation sporadique, environ 5 % des cancers correspondent à de véritables maladies héréditaires se transmettant selon un mode Mendélien<sup>1</sup>. Ces formes héréditaires de cancer représenteraient environ 300 000 nouveaux cas par an dans le monde <sup>2</sup>. **L'oncogénétique** est le secteur de la génétique consacré aux cancers héréditaires. En prenant comme exemple le cancer du sein dont la fréquence en population générale est de l'ordre de 1/9, la fréquence des cancers héréditaires du sein peut être estimée à 1/200. Les cancers héréditaires sont les maladies Mendéliennes les plus fréquentes et représentent aujourd'hui un des principaux motifs de consultation en génétique médicale. La caractérisation de leurs bases génétiques et l'élaboration de protocoles de prise en charge médicale personnalisée représentent une des plus grandes avancées de la génétique et de la cancérologie au cours des 20 dernières années.

Les cancers héréditaires résultent d'altérations **constitutionnelles** (c'est-à-dire présentes dans toutes nos cellules par opposition aux altérations somatiques), transmises de générations en générations. Les cancers héréditaires se transmettent le plus souvent selon un mode autosomique dominant à pénétrance incomplète et plus rarement selon un mode autosomique récessif. Les altérations constitutionnelles à l'origine des formes Mendéliennes de cancer ont pour effet commun de provoquer différents types d'instabilité génétique qui favoriseront la survenue d'altérations somatiques secondaires. Ainsi le cancer héréditaire est la résultante d'altérations constitutionnelles suivies d'altérations somatiques, alors que les cancers sporadiques résultent de l'accumulation séquentielle d'altérations somatiques (Figure 1).



**Figure 1** : Différence entre un cancer sporadique et un cancer familial

Dans la majorité des cas, les cancers sont sporadiques, c'est à dire résultant de l'accumulation de mutations somatiques dans une cellule. A l'inverse, les cancers familiaux sont la résultante de la présence d'une altération constitutionnelle dans un gène de prédisposition au cancer suivie de l'accumulation de mutations somatiques dans une cellule. De fait, les cancers familiaux ont généralement un âge de survenue plus précoce.

## II. Les bases moléculaires des prédispositions héréditaires au cancer

### a. Définition d'un gène de prédisposition héréditaire au cancer

Comme annoncé succinctement ci-dessus, les prédispositions héréditaires au cancer sont la conséquence **d'une altération délétère constitutionnelle dans un gène de prédisposition**. Un gène de prédisposition au cancer (*Cancer Predisposition Gene; CPG*) peut être défini comme un gène dont les altérations constitutionnelles conduisent à une augmentation du risque relatif de développer un ou plusieurs cancers chez le sujet porteur, par rapport au risque moyen observé dans la population générale. En pratique médicale, l'élément central est la pénétrance des variations, c'est à dire le risque absolu de développer une tumeur. Ainsi, les formes héréditaires de cancer correspondent aux formes Mendéliennes monogéniques dûes à l'altération d'un gène majeur de prédisposition avec une pénétrance de l'ordre de 50 à 100 % le plus souvent. Cependant, les formes monogéniques n'expliquent pas la majorité des agrégations familiales de cancer qui pourraient résulter d'un déterminisme oligogénique, c'est à dire de la contribution d'une combinaison de quelques variations rares.

En 2009, Stratton a adjoint la notion de l'environnement à la définition d'un CPG, en proposant qu'un gène du cancer est un gène dont les altérations confèrent **un avantage prolifératif à la cellule tumorale** et qui **seront sélectionnées positivement par le microenvironnement** tissulaire dans lequel le cancer survient<sup>3</sup>. En 2014, Nazneen Rahman a proposé la définition d'un CPG comme étant un gène pour lequel des altérations constitutionnelles rares confèrent un risque modéré à élevé de développer un cancer (de deux à cinq fois supérieur par rapport à un individu non porteur) et pour lequel 5 % des individus porteurs vont développer la maladie<sup>2</sup>. Même si la plupart des CPG confèrent un risque et une pénétrance en réalité bien plus élevés, l'avantage de cette définition est d'exclure les gènes associés à un risque faible de développer un cancer, découverts généralement par étude d'association du type *Genome-Wide Association Study (GWAS)*. Rahman a ainsi recensé **114 gènes** répondant à cette définition.

Il est important de noter d'emblée que contrairement à de nombreuses maladies Mendéliennes où une seule mutation contribue au phénotype, le cancer n'est généralement pas causé par le défaut d'un seul gène mais par une succession d'évènements moléculaires somatiques<sup>4</sup>. En effet, comme cela va être développé ci-dessous, les cellules de mammifères sont dotées de multiples systèmes de protection leur permettant de lutter contre



le développement tumoral. Ainsi, le terme "gène de prédisposition au cancer" doit plutôt sous-entendre "gène contribuant au cancer" et non "gène causant le cancer".

## **b. Catégories de gènes prédisposant au cancer**

Les altérations géniques à l'origine de la transformation maligne, qu'elles soient d'origine somatique ou constitutionnelle, ont deux types de conséquences majeures : l'hyperactivation inappropriée d'un gène ou son inactivation. Les gènes impactés par les altérations activatrices regroupent des gènes activant les cascades de transduction du signal, et sont dénommées "proto-oncogènes". A l'inverse, les altérations inactivatrices altèrent, quant à elles, les gènes qualifiés de "suppresseurs de tumeurs" et les gènes impliqués dans la réparation de l'ADN.

### **Activation de proto-oncogènes :**

Les **proto-oncogènes** furent la première catégorie de gènes à être mise en cause dans l'oncogenèse, avec l'identification de l'oncogène *src* en 1976<sup>5</sup>. Les proto-oncogènes peuvent être caractérisés en cinq ensembles : (i) les gènes codant pour les facteurs de croissance, (ii) les gènes codant pour les récepteurs aux facteurs de croissance, (iii) les gènes codant pour la transduction du signal, (iv) les gènes codant pour les facteurs de transcription et enfin (v) les gènes codant pour les régulateurs de l'apoptose. Ces gènes sont donc centrés autour de la transduction des différents signaux reçus par la cellule, nécessaires au bon déroulement du cycle cellulaire, en fonction de l'environnement notamment. L'altération délétère d'un proto-oncogène entraîne la formation de protéines appelées oncogènes hyperactifs et favorisant la néoplasie (augmentation de la croissance et diminution de la mort cellulaire programmée). Les activations des proto-oncogènes peuvent résulter (i) soit d'une mutation ponctuelle responsable d'un gain de fonction délétère de la protéine, (ii) soit d'une amplification génique, (iii) soit de translocations chromosomiques. La translocation d'un proto-oncogène à proximité d'une zone transcriptionnellement active a pour conséquence de déréguler l'expression du gène ou de créer un gène de fusion codant pour une nouvelle protéine ayant une fonction aberrante. Le gène *RET* est l'exemple emblématique d'activation d'un proto-oncogène. Ses mutations constitutionnelles activatrices prédisposent au cancer médullaire de la thyroïde (CMT) caractéristique de la néoplasie endocrinienne multiple de type 2 (NEM2)<sup>6,7</sup>. En 1993, deux équipes montrèrent l'implication du proto-oncogène *RET* en démontrant son activation aberrante et constitutive dans les formes héréditaires du CMT et dans la NEM2<sup>8,9</sup>. En terme d'hérédité, les altérations constitutionnelles de proto-oncogènes agissent selon un mode dominant. En effet, l'altération

d'un seul allèle peut conduire à la dérégulation du produit de ce gène et sera à l'origine d'une prolifération cellulaire incontrôlée <sup>10</sup>. Cela explique en partie pourquoi les altérations délétères constitutionnelles des proto-oncogènes ont le plus souvent une pénétrance complète et un âge de survenue précoce. Les proto-oncogènes représentent environ 10 % des CPG (11 proto-oncogènes décrits parmi les 114 CPG proposés par N.Rahman).

### **Inactivation de gènes suppresseurs de tumeurs**

A l'inverse des proto-oncogènes, les gènes suppresseurs de tumeurs appelés également "*gatekeepers*", régulent négativement la division cellulaire en inhibant la progression du cycle cellulaire ou en induisant l'apoptose <sup>10</sup>. Ces gènes doivent être inactivés ou délétés pour initier le processus tumoral. Physiologiquement, les *gatekeepers* interviennent dans le contrôle de la division cellulaire. Ils sont capables d'orienter la cellule vers l'apoptose devant un nombre important d'altérations de l'ADN et de prévenir ainsi la tumorigenèse. En règle générale, les protéines codées par les gènes suppresseurs de tumeurs inhibent les voies biologiques activées par les oncogènes au niveau cellulaire. Pour inactiver ces gènes, deux événements moléculaires sont généralement nécessaires. Dans les cas de cancers sporadiques, deux événements moléculaires doivent se produire au sein de la même cellule pour initier le processus tumoral alors que dans les cancers héréditaires la première altération est déjà présente dans l'ensemble des cellules de l'organisme. Le premier gène suppresseur de tumeur ainsi caractérisé fut le gène *RB1*, impliqué dans le développement des rétinoblastomes, une tumeur rare de la rétine <sup>11 12</sup>. Ce type de tumeur survient généralement chez le jeune enfant et leur incidence est estimée aujourd'hui entre 1/15 000 à 1/20 000 naissances <sup>13</sup>. Un autre modèle emblématique de gène suppresseur de tumeur impliqué dans le contrôle du cycle cellulaire et de l'apoptose est celui du gène *TP53* dont les altérations constitutionnelles sont responsables du syndrome de Li-Fraumeni <sup>14,15</sup>. Ce syndrome rare est une affection prédisposant à un risque important de développer de nombreux types de cancers, en particulier des sarcomes des tissus mous, des ostéosarcomes, des cancers du sein préménopausiques, ainsi que des tumeurs particulièrement rares comme des tumeurs des plexus choroïdes ou des corticosurrénales. En conclusion, les altérations des gènes suppresseurs de tumeurs permettent aux cellules de continuer une progression tumorale malgré la présence d'anomalies de l'ADN qui normalement auraient conduit à l'arrêt du cycle cellulaire. Le contrôle du cycle cellulaire est donc un élément important de régulation de la prolifération cellulaire et ses altérations qu'elles soient d'origine constitutionnelle ou somatique contribuent de façon importante à l'oncogénèse.

## Inactivation de gènes de stabilité du génome ou *caretakers*

La troisième catégorie de CPG concerne des gènes appelés gènes de stabilité du génome (*caretakers*). Ils interviennent dans la voie de réponse aux lésions de l'ADN incluant les différentes voies de réparation de l'ADN et donc dans le maintien de la stabilité du génome. Les *caretakers* assurent l'intégrité du génome lors de la réplication de l'ADN, en réparant les lésions de l'ADN produites sous l'effet de mutagènes externes ou endogènes. Au même titre que les *gatekeepers*, une perte de fonction des protéines codées par ces gènes nécessite l'inactivation des deux allèles du gène. L'inactivation des *caretakers* n'a pas pour conséquence une prolifération cellulaire accrue mais une accumulation d'altérations génétiques non réparées<sup>16</sup>. Ceci a pour conséquence d'augmenter drastiquement l'instabilité génomique et, *in fine* l'apparition d'altérations secondaires susceptibles d'activer les proto-oncogènes, d'inactiver les *gatekeepers* voire même d'autres gènes de stabilité<sup>10</sup>. La perte de fonction des *caretakers* peut avoir différentes origines. Il peut s'agir de délétions emportant le gène suppresseur de tumeur, de mutations ponctuelles altérant sa fonction, d'une interaction avec un oncogène viral ou encore d'altérations épigénétiques qui vont diminuer l'expression du gène. Comme énoncé précédemment, les *caretakers* codent pour les protéines incluses dans les différents systèmes de réparation de l'ADN :

- le système de réparation par excision de base (*Base Excision Repair* ; BER) qui comme son nom l'indique est la voie biologique impliquée dans la réparation des bases, affectées par l'oxydation, les alkylations, les déaminations...

- le système de correction des bases mal appariées résultant notamment d'erreurs de réplication (*Mismatch Repair system* ou système MMR), qui permet de réparer post-réplication les erreurs échappant à l'activité *proof-reading* des polymérases de la réplication. Ce système est impliqué dans la réparation des bases mal-appariées et des indels.

- le système de réparation par excision de nucléotides (*Nucleotide Excision Repair* ; NER), système capable de reconnaître et de supprimer des lésions de l'ADN volumineuses comme des pontages interbrin (*interstrand crosslinks*) ou des torsions de l'hélice bloquant la division cellulaire.

- le système de réparation des cassures double brin (*Double-Strand breaks* ; DSB). Ce système implique la réparation des cassures double brin produites lors d'arrêt de la réplication ou d'altérations exogènes telles que les rayonnements ionisants. Pour réparer ces lésions, il existe à minima deux voies de réparation : la recombinaison homologue (*Homologous Recombinaison* ; HR) et la recombinaison non homologue (*Non-Homologous End-Joining*; NHEJ).

- la voie de l'anémie de Fanconi (*Fanconi Anemia* ; FA). Ce système agit principalement sur les pontages intra-brin et inter-brins. Ces gènes sont capables de

reconnaître ces lésions et d'entraîner des cascades biologiques pour les réparer. À ce jour, plus d'une quinzaine de gènes ont été décrits comme participant à cette voie (de *FANCA* à *FANCG*).

Les *caretakers* incluent également des gènes codant pour des protéines impliquées dans la détection et la signalisation des lésions de l'ADN, dans le remodelage de la chromatine ou encore dans des processus d'*editing* <sup>17</sup>.

Le tableau 1 présente différentes formes Mendéliennes de cancers et leurs bases moléculaires.

<b>Forme héréditaire de cancer</b>	<b>Gène</b>	<b>Voie(s) biologique(s)</b>
Cancer Médullaire de la thyroïde (CMT)	<i>RET</i>	ERK, AKT, MAPK, JNK, STAT
Cancer papillaire du rein	<i>MET</i>	ERK, PI3K/AKT, MAPK, RAS
Maladie de Cowden	<i>PTEN</i>	PI3K, AKT, VEGF, PTEN
Syndrome de Gorlin	<i>PTCH</i>	<i>Sonic Hedgehog</i>
Médulloblastome	<i>SUFU</i>	<i>Sonic Hedgehog</i>
Neurofibromatose de type I	<i>NF1</i>	RAS
Maladie de Von Hippel Lindau	<i>VHL</i>	<i>Hypoxia-inducible factor</i>
Polypose adénomateuse familiale	<i>APC</i>	Voie de la b-caténine
Rétinoblastome	<i>RB1</i>	G1/S, Mitose
Mélanome malin	<i>CDKN2A</i>	Voie RB
Syndrome de Li-Fraumeni	<i>TP53</i>	G1/S, G2/M
Syndrome de Lynch	<i>MSH2, MLH1, MSH6, PMS2</i>	<i>MisMatch Repair (MMR)</i>
Cancer du sein et de l'ovaire	<i>BRCA1, BRCA2</i>	<i>Homologous Recombination mediated Double Strand Break Repair (HR.DSBR)</i>
Polypose adénomateuse autosomique récessive	<i>MUTYH</i>	<i>Base-Excision Repair (BER)</i>
<i>Xeroderma Pigmentosum</i>	<i>XP</i>	<i>Nucleotide-Excision Repair (NER)</i>

**Tableau 1** : Bases moléculaires des principales formes héréditaires de cancer

### c. L'oncogénétique en pratique

La France sous, l'égide de l'Institut National du Cancer (INCa), a déployé un réseau de 126 sites de consultations d'oncogénétique bénéficiant de l'expertise de 25 laboratoires spécialisés en oncogénétique moléculaire. Cette organisation nationale et ce maillage territorial permettent de garantir une expertise et une équité de prise en charge (voir <http://www.e-cancer.fr/soins/prises-en-charge-specifiques/oncogenetique>).

À ce jour, il a été dénombré plus de cent syndromes associés à des prédispositions génétiques au cancer <sup>2,18</sup>. En terme de fréquence, les cancers familiaux du sein et de l'ovaire (*Hereditary Breast and Ovarian Cancer; HBOC*) et les cancers colorectaux héréditaires sans polype (*Hereditary Non-Polyposis Colorectal Cancer ou syndrome de Lynch; HNPCC*) sont les deux formes de cancers héréditaires les plus fréquentes <sup>18</sup>.

#### Critères évocateurs d'une prédisposition héréditaire au cancer

Les prédispositions héréditaires au cancer présentent, au même titre que les maladies Mendéliennes, un certain nombre de caractéristiques communes. En pratique clinique, quatre éléments majeurs sont fortement évocateurs d'une prédisposition héréditaire au cancer :

1. **La survenue d'un cancer à un âge précoce**, généralement avant l'âge de 51 ans. La majorité des cancers héréditaires se développent avant l'âge de 51 ans et même s'il existe une variabilité intrafamiliale dans les âges au diagnostic, il existe le plus souvent dans les familles atteintes de formes Mendéliennes de cancer des cas précoces.
2. **Une histoire familiale de cancer**, c'est à dire l'existence de plusieurs cas de cancers sur la même branche parentale. La pénétrance incomplète souvent associée à ces pathologies impose néanmoins de considérer les antécédents de cancer sur plusieurs générations. La taille de la famille est également un facteur à prendre en considération. En effet, une agrégation familiale d'un cancer fréquent en population générale ne signe pas forcément une forme Mendélienne de cancer.
3. **L'existence de tumeurs primitives multiples chez un individu** avec notamment les formes multifocales et bilatérales pour les organes doubles. En effet, les patients présentant une altération constitutionnelle prédisposant au cancer ont un risque élevé de développer plusieurs tumeurs indépendamment puisque l'altération préexiste dans toutes les cellules de l'organisme. L'âge de survenue du premier cancer est important et doit être inférieur à 51 ans. En effet, ces agrégations Mendéliennes ne doivent pas

être confondues avec le développement en rafale de cancers primitifs multiples chez les personnes âgées dont l'origine est très probablement due à des altérations épigénétiques associées à l'âge.

4. **Une association d'une pathologie tumorale avec des malformations, et/ou problèmes développementaux et/ou, neurologiques, fortement évocatrice d'une forme syndromique de cancer héréditaire.**

L'ensemble de ces observations phénotypiques est généralement affiné grâce à l'existence de critères propres à chaque prédisposition au cancer. A titre d'illustration, on peut citer les critères de Chompret dans le contexte du syndrome de Li-Fraumeni (*Li-Fraumeni Syndrome*; LFS) (Tableau 2) et les critères d'Amsterdam dans le contexte du HNPCC. D'autres critères peuvent également émaner de sociétés savantes ou de groupe de travail associé à ces dernières, comme c'est le cas pour les cancers pédiatriques par exemple. Le groupe "*Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology*" propose six portes d'entrée dans le contexte des prédispositions héréditaires aux tumeurs pédiatriques, à savoir : (i) la porte d'entrée basée sur l'histoire familiale, (ii) la survenue de tumeurs rares et/ou caractéristiques d'une altération d'un gène de prédisposition au cancer, (iii) la survenue de tumeurs primitives multiples, (iv) l'association avec des anomalies congénitales, (v) une résistance thérapeutique aux traitements anti-tumoraux classiques due probablement à des défauts de gènes impliqués dans la réparation de l'ADN, et (vi) une analyse tumorale suggérant une altération constitutionnelle <sup>19</sup>.

## Critères de Chompret

---

1. Le cas index atteint d'une tumeur appartenant au spectre tumoral du LFS (sarcome des tissus mous, ostéosarcome, tumeur cérébrale, cancer du sein pré-ménopausique, corticosurréalome, leucémie ou cancer du poumon broncho-alvéolaire) avant l'âge de 46 ans ET au moins un apparenté du premier ou du deuxième degré atteint d'une tumeur LFS (sauf cancer du sein si le proposant est atteint d'un cancer du sein) avant l'âge de 56 ans ou présentant des tumeurs multiples
2. Un cas index présentant des tumeurs multiples (sauf tumeurs mammaires multiples), dont deux appartenant au spectre tumoral du LFS et dont la première est survenue avant l'âge de 46 ans
3. Le cas index atteint d'un corticosurréalome ou d'une tumeur des plexus choroïdes, quels que soient les antécédents familiaux
4. Patiente ayant développé un cancer du sein d'apparition précoce avant l'âge de 31 ans et chez qui la recherche de mutations des gènes *BRCA1/BRCA2* a abouti à un résultat négatif.

**Tableau 2** : Présentation des quatre portes d'entrée du syndrome de Li-Fraumeni d'après les critères de Chompret version réactualisée en 2015.

LFS = *Li-Fraumeni Syndrome* ou Syndrome de Li-Fraumeni



#### **d. Méthodes d'identification des CPG**

Depuis la découverte de l'implication de *RB1* dans le rétinoblastome à la fin des années 90<sup>12</sup>, l'identification de nouveaux CPG n'a cessé d'augmenter : en moyenne, un CPG est identifié chaque année, avec un pic en 1996 où il a été mis en évidence 10 nouveaux CPG<sup>2</sup>. Le nombre actuel de CPG est difficilement estimable, probablement compris entre 120 et 200<sup>2,20</sup>. Il existe, trois grandes méthodes d'identification de ces CPG qui se sont succédées en fonction des évolutions technologiques : l'analyse de liaison génétique, les approches par gène candidat et enfin les approches d'analyse mutationnelle à l'échelle du génome (*genome wide mutational analysis*).

#### **Les approches par analyses de liaison génétique**

Il s'agit d'une approche sans *a priori*, basée sur la co-ségrégation de marqueurs génétiques et de la maladie. Cette stratégie a connu son apogée dans les années 1990 et a permis la découverte d'environ la moitié des gènes de prédisposition au cancer. L'exemple emblématique du succès de ce type d'analyses est l'identification des gènes de prédisposition aux cancers du sein et de l'ovaire, *BRCA1* et *BRCA2*. L'analyse de liaison génétique chez une trentaine de familles présentant des cancers du sein *a priori* héréditaire a permis à Mary Claire King et son équipe en 1990 d'incriminer la région 17q21<sup>21</sup>, région chromosomique porteuse du gène *BRCA1*.

Les principales limites de ce type d'analyse sont la possibilité d'une phénocopie et d'une pénétrance incomplète faussant l'hypothèse de base. La taille de la famille est également un critère important dans ce type d'analyse : une famille de petite taille ne permettra pas d'obtenir un nombre suffisant de méioses et donc de recombinaisons (*crossing-overs*) déterminantes. Aujourd'hui, ce type d'analyse est généralement employé en complément du séquençage à haut-débit. Les principales applications sont d'évaluer la véracité des informations communiquées sur le *pedigree* de la famille et de limiter le nombre de régions d'intérêt à analyser et donc de restreindre le nombre de variants à interpréter.

#### **Les approches de gène candidat**

Il s'agit sûrement de l'approche la plus compliquée car elle nécessite une bonne compréhension de la physiopathologie tumorale. Ainsi, beaucoup de gènes candidats ont été proposés et étudiés avec un succès cependant limité. Malgré tout, un succès élégant de cette approche a été par exemple l'identification des gènes *MLH1*, *MSH2*, *MSH6* comme gènes de prédisposition au HNPCC. Alors que les recherches du ou des gènes

responsables du syndrome de Lynch étaient basées sur une approche par analyse de liaison impliquant les chromosomes 2p et 3p, il a été démontré que les tumeurs se produisant chez les patients HNPCC possédaient un phénotype particulier (phénotype MSI+, *MicroSatellite Instability*), conséquence d'une réparation défectueuse post-réplication. Cette observation a permis l'identification des deux premiers gènes sur le 2p et le 3p à savoir *MSH2* et *MLH1*. En effet, les fonctions de ces gènes étaient semblables aux homologues bactériens MutS et MutL, accélérant le processus d'identification des gènes impliqués dans le HNPCC <sup>21-23</sup>.

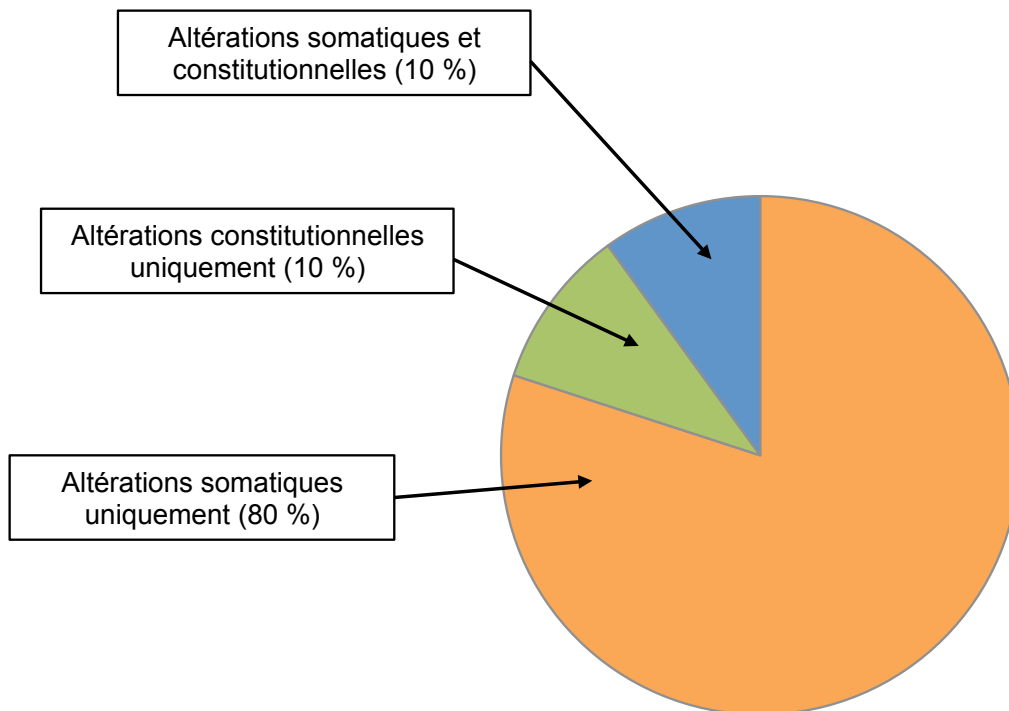
### **Les approches mutationnelles à l'échelle du génome "genome-wide mutational analysis"**

Ces approches sont basées sur la recherche d'altérations pouvant expliciter la pathologie dans l'ensemble du génome. Conceptuellement, il s'agit donc de l'approche inverse de la méthode des gènes candidats : cette démarche part de l'altération pour identifier le gène impliqué. Ce type d'analyse sans *a priori*, connaît un réel succès depuis l'émergence de nouvelles technologies pangénomiques comme le séquençage à haut-débit. Etant donné le grand nombre de variants identifiés par ces approches, plusieurs méthodes et stratégies d'analyses ont été développées pour faciliter l'analyse (cf. Chapitre III). Aujourd'hui, ces approches ne sont plus limitées techniquement et représentent une opportunité exceptionnelle d'identifier de nouveaux gènes à l'origine de prédispositions au cancer et sont à l'origine d'un réel changement de paradigme. La limite ne réside plus dans l'identification de variations mais dans leur interprétation et dans la difficulté de démontrer un enrichissement en variants rares chez des patients par rapport à des individus témoins.

### III. L'émergence de nouveaux concepts

#### a. Le nombre de gènes prédisposant au cancer serait limité

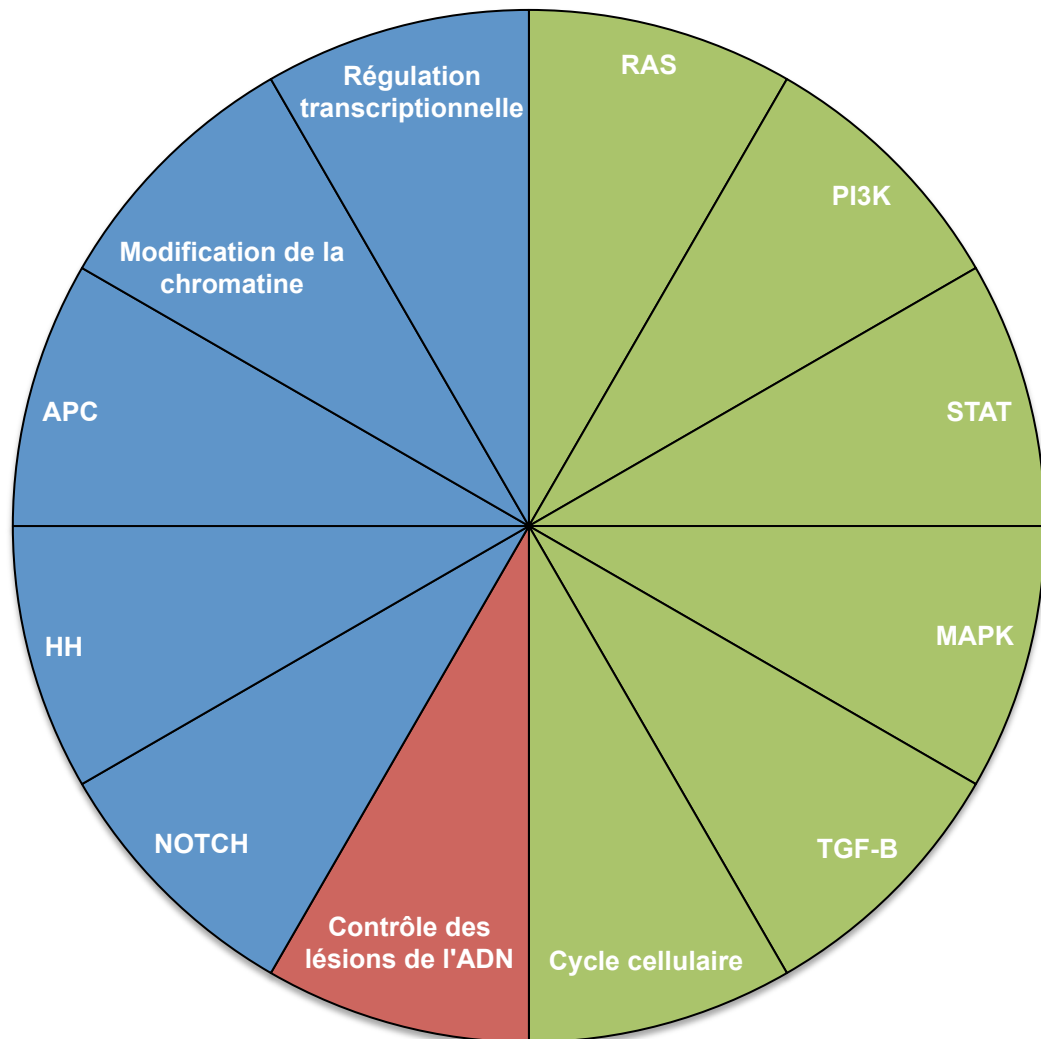
En 2004, avant l'émergence du séquençage à haut-débit, le groupe de M.Stratton et N.Rahman a émis l'hypothèse que le nombre de gènes impliqués dans la cancérogenèse serait en réalité limité. Les auteurs ont recensé, sur la base de données bibliographiques, 291 gènes dont l'implication dans le cancer était clairement démontrée, tant au niveau somatique qu'au niveau constitutionnel. Environ 90 % des 291 gènes identifiés présentaient des altérations somatiques, 20 % des altérations constitutionnelles. Ainsi, 10 % des gènes incriminés étaient à la fois altérés somatiquement et constitutionnellement (Figure 2). Les auteurs ont estimé qu'environ 1 % des 25 000 gènes participeraient aux processus tumoraux dont une soixantaine de gènes au niveau constitutionnel <sup>20</sup>. Dix ans plus tard, Nazneen Rahman a effectué une analyse bibliographique similaire conduisant à l'identification de 114 gènes CPG dont l'altération constitutionnelle est à l'origine d'une prédisposition héréditaire aux cancers <sup>2</sup>. Donc, malgré l'émergence du séquençage à haut-débit, l'identification des CPG ne semble donc pas être exponentielle mais plutôt limitée.



**Figure 2** : Altérations des gènes impliqués dans le Cancer

Pourcentage de gènes altérés dans les cancers humains par des mutations somatiques (en vert), par des mutations constitutionnelles (en bleu) et par les deux types d'altérations (en orange). Figure adaptée de Futréal et al., 2004 <sup>20</sup>.

De plus, en 2013, Vogelstein et ses collaborateurs ont décrit 12 voies potentielles de l'oncogenèse<sup>24</sup>. Ainsi, il est très probable qu'en plus d'être limitées, ces altérations soient en plus, confinées dans un nombre limité de voies de signalisation et de régulation (Figure 3).



**Figure 3** : Voies biologiques impliquées dans la cancérogenèse selon Vogelstein *et al*<sup>24</sup>. Vogelstein et ses collaborateurs ont décrit 12 voies de l'oncogenèse à partir de données de séquençage à haut-débit de tumeurs. Ces voies sont en plus confinées autour de trois grands processus cellulaires : la survie cellulaire (vert), le maintien de l'intégrité du génome (rouge) et dans la croissance cellulaire (bleu). Figure adaptée de Vogelstein *et al*<sup>24</sup> HH = voie *HedgeHog*

Comme décrit ci-dessus, il existe un recouvrement entre les gènes altérés dans la tumeur et les gènes altérés constitutionnellement. Environ 10 % des gènes mutés dans les tumeurs sont connus pour être des gènes de prédisposition au cancer lorsqu'ils sont mutés à

l'état constitutionnel. Paradoxalement, 40 % des gènes de prédisposition au cancer peuvent également contribuer à l'oncogenèse quand ils sont altérés de façon somatique dans la tumeur. Cette contradiction apparente est explicable, au moins en partie, par un artefact dans l'approche d'identification. En effet, il est plus fréquent d'explorer les altérations d'un CPG au niveau somatique que d'explorer la contribution constitutionnelle des gènes mutés au niveau somatique.

Un élément supplémentaire venant corroborer l'hypothèse d'un nombre limité de gènes est que certains gènes peuvent être altérés à l'état hétérozygote ou à l'état bi-allélique, avec généralement un phénotype aggravé dans ces cas <sup>25</sup>. L'exemple emblématique est le gène *BRCA2* qui lorsque altéré mono-alléliquement est un gène de prédisposition au cancer du sein (OMIM: #612555) <sup>26</sup> alors que les altérations bi-alléliques de ce gène sont à l'origine d'une des formes d'anémie de Fanconi (OMIM: #605724) <sup>27</sup>. Rahman a dénombré seize gènes suivant ce modèle (ex: *PALB2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*) <sup>2</sup>.

#### **b. Parallélisme entre les gènes du développement et les gènes de l'oncogenèse**

Il est aujourd'hui bien admis que plusieurs gènes mis en cause dans l'oncogenèse ont des rôles majeurs lors de la différenciation cellulaire et tissulaire se produisant durant l'embryogenèse. Ceux-ci sont généralement impliqués dans les voies de signalisation telles que les voies TGF $\beta$ /BMP, Wnt, Notch, EGF, FGF et Hedgehog. Cette hypothèse est soutenue par (i) l'identification d'altérations constitutionnelles d'un gène du développement prédisposant à un risque élevé de développer un cancer, (ii) l'identification d'anomalies du développement chez les porteurs d'altérations constitutionnelles d'un CPG, (iii) des phénotypes distincts développementaux ou oncogéniques en fonction de l'altération (gain ou perte de fonction ou altération mono-allélique ou bi-allélique) (iv) et également l'identification d'altérations somatiques des gènes du développement dans la cellule tumorale. Plusieurs exemples de mutations constitutionnelles activatrices ou inactivatrices permettent d'illustrer ce paradigme :

- **Les altérations de la voie Hedgehog.** Cette voie est impliquée dans la détermination des patrons de segments, des yeux et des membres de l'embryon. Les protéines intervenant dans cette voie sont généralement des morphogènes c'est-à-dire qu'elles sont distribuées en gradient selon les régions de l'embryon. Les altérations de cette voie sont responsables de deux phénotypes distincts. Les altérations inactivatrices sont associées à une holoprosencéphalie, une anomalie du développement du prosencéphale <sup>28</sup>.

A l'inverse, les altérations constitutionnelles activatrices conduisent au syndrome de Gorlin dû au rôle de la voie Hedgehog dans la prolifération cellulaire<sup>29</sup>.

- **les altérations du gène *RET***. Ce gène code pour un récepteur transmembranaire à activité tyrosine kinase qui, en cas de mutations activatrices, va être activé de façon permanente et conduire à une stimulation de la prolifération cellulaire. Cliniquement, les mutations activatrices constitutionnelles de ce gène conduisent à un cancer médullaire de la thyroïde<sup>30</sup>. A l'inverse, l'inactivation constitutionnelle de ce gène conduit à la maladie de Hirschprung<sup>31</sup> qui est un trouble congénital de la motilité intestinale caractérisé par des signes d'occlusion intestinale dus à un défaut de développement du système nerveux entérique dans le colon terminal.

- les altérations de la **voie *RAS-RAF-MEK***. L'activation somatique de certains gènes impliqués dans ces voies conduit au développement de cancers coliques<sup>32</sup>, de mélanomes<sup>33</sup> ou de cancers des poumons<sup>34</sup>. L'activation constitutionnelle de gènes impliqués dans cette voie de signalisation conduit à des maladies du développement définies sous le terme de « *RASopathies* ». Ainsi, les altérations constitutionnelles activatrices du gène *PTPN11*, *SOS1*, *K-RAS* sont impliquées dans le syndrome de Noonan<sup>35-37</sup>. Ce syndrome est une maladie du développement caractérisée par une petite taille, une dysmorphie faciale caractéristique et des anomalies cardiaques congénitales associées avec un risque important de développer des tumeurs telles que les rhabdomyosarcomes.

D'autre part, il est également intéressant de noter que la majorité des CPG ne sont pas retrouvés altérés constitutionnellement à l'état bi-allélique, ce qui souligne leur importance tôt dans le développement. Il est fortement probable que la double inactivation constitutionnelle de certains CPG entraîne une mort *in utero*. Cette hypothèse est corroborée par la forte létalité embryonnaire des modèles murins *knock-out* de CPG<sup>30</sup> suggérant que les voies du développement et de l'oncologie se recouvrent ou au moins en partie.

Enfin, un dernier argument permettant d'étoffer ce parallélisme est que le fait que les cellules tumorales utilisent des processus développementaux pour la dissémination tumorale : la transition épithélium-mésenchyme (*Epithelial to Mesenchymal Transition*; EMT) et la transition mésenchyme-épithélium (*Mesenchymal to Epithelial Transition*; MET). La EMT permet à une cellule épithéliale de se transformer en cellule mésenchymateuse, elle est à l'origine de la migration et de la dissémination cellulaire. La MET est le processus inverse, qui permet à la cellule mésenchymateuse, dans la circulation sanguine, de coloniser un organe et d'entraîner des métastases. Ainsi, il existe un parallélisme mécanistique entre le développement et la cancérogenèse par l'utilisation de voies biologiques similaires.

### c. Oligogénisme/Polygénisme

Un des apports importants du NGS (*Next Generation Sequencing; NGS*) est la mise en évidence du déterminisme oligogénique de certaines maladies. En effet, en dehors des maladies connues de déterminisme multifactoriel impliquant des variants fréquents (fréquence allélique > 1 %) révélés par étude de GWAS (*Genome-Wide Association Study; GWAS*), les études d'exome et de panel de gènes par NGS ont révélé qu'un certain nombre de maladies impliqueraient une contribution de quelques variations de fréquence allélique rare (fréquence allélique < 1 % )

L'oligogénisme regroupe en réalité deux modèles : (i) le modèle synergique où les variants contribuent ensemble de façon équivalente au phénotype (les porteurs d'un seul variant sont donc asymptomatiques) et (ii) le modèle modificateur où un variant contribue majoritairement au phénotype et où les autres variants influent sur la gravité ou sur l'âge de survenue par exemple. Les porteurs du variant contributeur majoritaire sont donc affectés de façon modérée voire avec des tableaux phénotypiques incomplets, alors que les porteurs des variants modificateurs sont porteurs sains en règle générale.

Cette notion commence à fortement émerger dans le contexte des prédispositions héréditaires au cancer. Par exemple, l'équipe de Ballinger a très récemment démontré que les sarcomes, longtemps considérés comme majoritairement non héréditaires en dehors du contexte du syndrome de Li-Fraumeni, des formes familiales de rétinoblastomes et de neurofibromatoses pouvaient résulter de variants oligogéniques dans des gènes de prédisposition héréditaire. Pour cela, les auteurs ont séquencé 72 gènes grâce à un panel de gènes et ont démontré qu'environ la moitié des patients présentant un sarcome possèdent soit une variation constitutionnelle prédite délétère monogénique, soit plusieurs variations constitutionnelles dans des gènes de prédisposition héréditaire au cancer connus ou non encore impliqués<sup>39</sup>.

## IV. Du diagnostic moléculaire au développement de nouvelles thérapies

### a. Déroulement des consultations d'oncogénétique

Comme pour toute maladie génétique, la première consultation d'oncogénétique doit s'effectuer si possible chez le cas index, adressé par le médecin traitant, gastro-entérologue, gynécologue, pédiatre, oncologue, chirurgien... afin de pouvoir proposer une analyse génétique à partir d'une prise de sang si, à l'issue de la consultation, le diagnostic de forme Mendélienne de cancer est envisagé. La première consultation a pour objectif d'établir les

antécédents personnels et familiaux de cancers en validant les données médicales par l'obtention en particulier des comptes-rendus anatomopathologiques, éléments essentiels au dossier. Comme dans toute consultation de génétique la règle essentielle est, avant de réaliser l'analyse génétique, d'anticiper les conséquences d'une analyse génétique positive en termes de prise en charge médicale, à la fois pour le cas index et ses apparentés. Il est important de souligner que les patient(e)s vu(e)s en consultation d'oncogénétique ont souvent une histoire médicale lourde personnelle et familiale et sont parfois en cours de traitement.

L'identification chez le cas index de la mutation constitutionnelle causale permettra dans un second temps de proposer aux apparentés asymptomatiques à risque génétique une **analyse pré-symptomatique** basée sur la recherche ciblée de l'altération identifiée. Il sera alors possible, chez les apparentés non porteurs de la mutation familiale, de lever une inquiétude illégitime et une surveillance médicale inappropriée, leur risque de développer un cancer étant celui de la population générale et aux apparentés porteurs de leur proposer une prise en charge personnalisée permettant de détecter de façon précoce des tumeurs voire de supprimer le risque tumoral par chirurgie prophylactique. Même si le bénéfice médical de cette prise en charge pré-symptomatique est le plus souvent majeur, il est important de proposer un **accompagnement psychologique** afin que ce bénéfice médical ne soit pas contrebalancé par un impact psychologique négatif.

#### **b. Prise en charge médicale : exemple des familles avec mutation des gènes *BRCA***

Comme le stipulent les recommandations de l'INCa et le présente le tableau 3, le suivi d'une femme porteuse d'une mutation d'un gène *BRCA* doit débuter dès 20 ans et est basé sur la palpation des seins et des creux axillaires tous les 6 mois. L'examen d'imagerie de référence chez une patiente présentant une mutation d'un gène *BRCA* est **l'IRM mammaire**. En effet, si la mammographie est un excellent examen de dépistage du cancer du sein en population générale chez les femmes de plus de 50 ans, cet examen est inadapté à l'imagerie mammaire de la femme jeune dont les seins sont plus denses et la sensibilité de la mammographie chez les porteuses d'une mutation *BRCA* est mauvaise, ne dépassant pas 50 %. De plus, chez les femmes jeunes, cet examen répété de façon annuelle en réalisant plusieurs incidences est dangereux car les niveaux de radiations ionisantes exposent à un risque mutagène. Chez les porteuses d'une mutation d'un gène *BRCA*, la surveillance par IRM doit commencer à partir de 30 ans ou 5 ans avant le cas de cancer du sein le plus précoce dans la famille s'il existe un cas avant 30 ans. L'IRM doit être réalisée à un rythme



annuel et être complétée, sur un délai d'un maximum de 2 mois, d'une incidence de mammographie permettant de détecter les micro-calcifications. Il est à noter que la surveillance par IRM génère dans 15 % des cas, dès le premier examen, des faux positifs conduisant à des biopsies. Si l'IRM a représenté une avancée majeure dans la prise en charge des familles avec mutation des gènes *BRCA*, le suivi par IRM de femmes porteuses d'une mutation a malheureusement révélé des cancers du sein de l'intervalle survenant entre deux IRM réalisées à un an d'intervalle. Ces cancers de l'intervalle, observés en particulier chez les porteuses d'une mutation *BRCA1* et la rapidité évolutive des cancers du sein résultant de mutations *BRCA1*, constituent les principaux arguments justifiant de proposer, en particulier, aux femmes porteuses d'une mutation du gène *BRCA1* une **mastectomie prophylactique** avec conservation ou non de la plaque aréolo-mammelonnaire et reconstruction simultanée. Afin que la femme puisse murir sa réflexion et prendre sa décision en étant parfaitement informée, il est crucial de proposer plusieurs séances d'information auprès d'un chirurgien spécialisé et un accompagnement psychologique et l'indication chirurgicale doit être posée lors d'une **réunion de concertation pluri-disciplinaire** (RCP) spécialisée. Il est recommandé de réaliser la mastectomie avant 40 ans, période pendant laquelle le risque de développer un cancer du sein est élevé et l'imagerie mammaire peut être prise en défaut. Aucune étude ne permet actuellement de considérer qu'il existe un âge limite pour cette intervention prophylactique. Chez les hommes porteurs d'une mutation *BRCA*, le risque de développer un cancer du sein reste suffisamment faible pour qu'il soit justifié de ne pas leur proposer un dépistage systématique par imagerie.

Concernant le risque de cancers de l'ovaire chez les porteuses d'une mutation *BRCA*, il n'existe à ce jour aucune technique d'imagerie permettant une détection précoce des tumeurs. L'efficacité de l'échographie trans-vaginale ou trans-pariétale chez ces patientes, est nulle. Par conséquent, la seule mesure capable d'éviter le risque de cancers de l'ovaire, dont le diagnostic clinique est en règle générale tardif, est l'**annexectomie complète prophylactique** à réaliser après 40 ans et après tout projet parental. Ce geste réalisé par coelioscopie est beaucoup plus simple que la mastectomie prophylactique et la majorité des porteuses de mutation s'orientent facilement vers ce choix prophylactique.

## Prise en charge des femmes porteuses d'une mutation d'un gène *BRCA*

---

- À partir de 20 ans, palpation des seins et des creux axillaires tous les 6 mois
- À partir de 30 ans, ou 5 ans avant le cas de cancer du sein s'il existe un cas avant 30 ans, IRM mammaire annuelle, complétée d'une seule incidence de mammographie
- Avant 40 ans, en particulier chez les porteuses d'une mutation de *BRCA1*, après information, temps de réflexion et accompagnement psychologique, mastectomie prophylactique
- Après 40 ans et tout projet parental, annexectomie prophylactique

**Tableau 3 :** Prise en charge des femmes porteuses d'une mutation d'un gène *BRCA*

La récurrence et le risque de tumeurs multiples étant plus élevés chez les porteurs d'altérations de CPG, il est important de prendre en considération les autres risques tumoraux associés aux mutations des gènes *BRCA*. Les hommes présentant une mutation de *BRCA1* et surtout de *BRCA2* ont une augmentation de risque de développer des **adénocarcinomes prostatiques**. Le risque cumulé a été respectivement estimé à 10 et 20 %. Ces cancers de la prostate sont en règle générale agressifs, d'évolution métastatique et surviennent avant 65 ans, soit 10 ans avant la population générale. Les porteurs et porteuses de mutations du gène *BRCA2* ont une augmentation du risque de développer des **cancers du pancréas** et des **mélanomes** mais, dans les 2 cas, le risque cumulé à 70 ans est inférieur à 10 %.

### c. Les nouveaux traitements basés sur les inhibiteurs de PARP

L'avancée la plus importante au cours des dernières années dans les cancers du sein et de l'ovaire héréditaires est le développement de thérapies ciblées exploitant l'altération du système de réparation des cassures double brin d'ADN dans ces cancers. Ces thérapies ciblées sont basées sur les inhibiteurs de PARP (*Poly(ADP)-ribose polymerase*). L'enzyme PARP-1 joue un rôle clef dans la réparation des cassures d'ADN simple brin par le

système BER (*Base Excision Repair*). L'inhibition de PARP empêche la réparation de ces cassures d'ADN simple brin, ce qui entraîne des cassures d'ADN double brin normalement réparées par recombinaison homologue. Si dans la cellule la protéine BRCA1 ou BRCA2 est inactivée, cette réparation ne peut avoir lieu et cela active le système NHEJ, ce qui entraîne une instabilité génétique<sup>34,35</sup>. Les cellules présentant une inactivation complète de la protéine BRCA1 ou BRCA2 sont 1000 fois plus sensibles que les autres cellules aux inhibiteurs de PARP. Les inhibiteurs de PARP provoquent donc une létalité sélective des cellules cancéreuses avec inactivation des gènes *BRCA*, et ont donc une toxicité réduite vis à vis des cellules non cancéreuses et sont donc bien tolérés chez les patientes. Plusieurs essais thérapeutiques ont démontré l'efficacité des inhibiteurs de PARP dans le traitement des cancers du sein et de l'ovaire et ces médicaments ont reçu l'autorisation de mise sur le marché à partir de 2015. Chez une patiente atteinte d'un cancer du sein ou de l'ovaire, l'identification d'une mutation délétère des gènes *BRCA* a donc un double impact : suivi personnalisé de la patiente et de la famille d'une part, et thérapie ciblée d'autre part. Ainsi, il est donc essentiel que la prescription d'une analyse des gènes *BRCA* se fasse en étroite interaction entre médecins oncologues et généticiens.

## **CHAPITRE II :**

### **La révolution du séquençage de nouvelle génération**

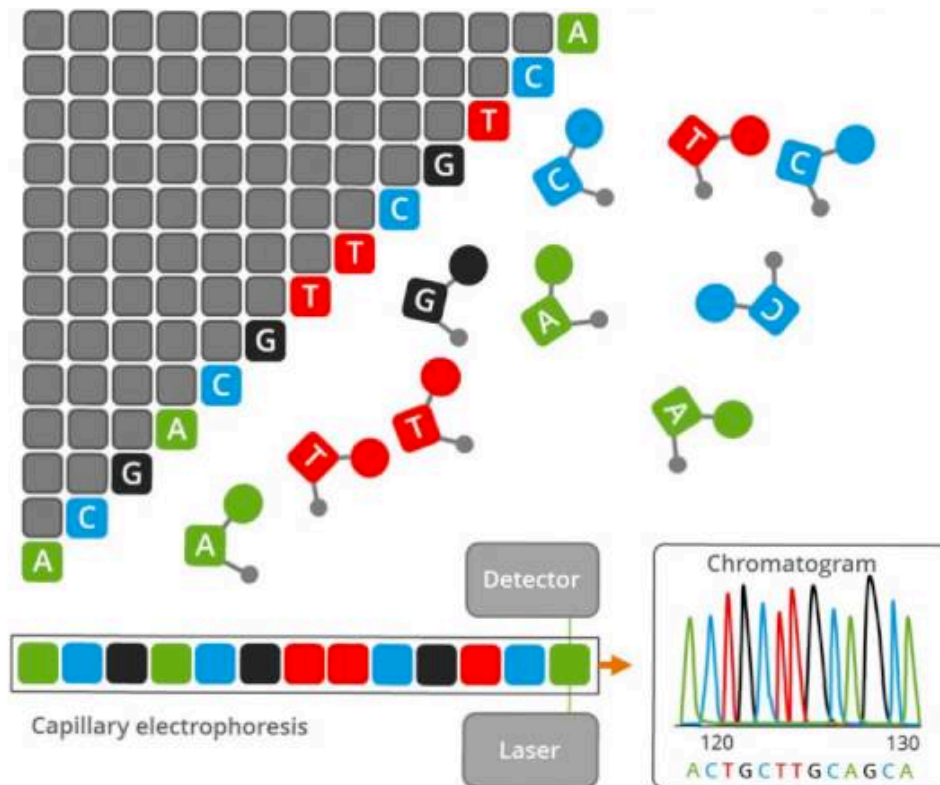
## I. Du séquençage Sanger au séquençage du génome humain

Depuis la découverte de l'ADN comme support de l'hérédité, le défi de sa lecture et de son décryptage est au cœur de la Génétique du XXI<sup>ème</sup> siècle. La lecture moléculaire de l'ADN ou séquençage représente donc une nouvelle porte ouverte vers une meilleure connaissance de la physiologie humaine et des maladies génétiques. En permettant l'identification des bases moléculaires des maladies génétiques, le séquençage de l'ADN favorise le diagnostic de ces pathologies et ouvre la voie vers de nouveaux développements thérapeutiques.

Historiquement, le séquençage de l'ADN prend naissance à la fin des années 1960 et n'a cessé d'évoluer conceptuellement et techniquement depuis. Les améliorations de ce champ d'application ont suivi les progrès réalisés dans de nombreux domaines parallèles comme celui de la chimie, de l'optique ou de l'informatique notamment.

La première véritable révolution dans le domaine du séquençage de l'ADN a été apportée par l'équipe de Frederick Sanger en 1977<sup>42</sup>. L'originalité de la technique Sanger est d'employer des nucléotides terminateurs de séquence : les didésoxyribonucléotides (ddNTP). Ces nucléotides ont la particularité de ne pas posséder d'extrémité 3' hydroxyl (OH) libre, empêchant de poursuivre l'élongation. Depuis cette découverte majeure dans l'histoire du séquençage, un grand nombre de perfectionnements de cette technique ont eu lieu comme l'amélioration des techniques de PCR (*Polymerase Chain Reaction*), l'utilisation de la fluorescence à la place de la radioactivité et l'arrivée des électrophorèses par capillaire (Figure 4). Ces développements permettent de séquencer aujourd'hui à l'aide de cette technique jusqu'au kilobase en une seule réaction. Le séquençage Sanger est une technologie robuste et fait encore aujourd'hui référence pour séquencer un nombre limité de gènes, ou rechercher des mutations récurrentes par exemple.

La réalisation la plus emblématique employant cette technologie est sans aucun doute le séquençage du génome humain (*Human Genome Project*, HGP), projet phare de la génétique à la fin des années 1990. Ce faramineux projet de 3 milliards de dollars, débuté en 1990 par un *consortium* international incluant 20 centres de 6 pays, a duré 13 ans avant d'être en mesure de révéler la séquence des quelques trois milliards de paires de bases qui constituent notre génome<sup>43</sup>. Le séquençage et la cartographie du génome humain avaient alors été réalisés pas à pas par une stratégie BAC à BAC c'est à dire un séquençage Sanger de courts fragments sous-clonés dans des BAC (*Bacterial Artificial Chromosome*). Cette étape clef a permis d'obtenir les premières informations sur la globalité du génome et fait donc date.



**Figure 4 :** Le séquençage par la méthode de Sanger

Le séquençage Sanger est basé sur l'utilisation de nucléotides particuliers : les didésoxyribonucléotides (ddNTP). Ces derniers ne possèdent pas d'extrémité 3'OH libre empêchant ainsi de poursuivre l'élongation. Grâce à l'utilisation d'un capillaire et d'un marquage de chaque ddNTP avec un fluorophore différent, il est possible de réaliser une électrophorèse séparant les fragments selon leurs tailles. Puis un système de laser et de détecteur sont employés pour la détection de la fluorescence et la quantification de l'intensité du signal, permettant de déduire la base incorporée.

Il est important de noter que parallèlement, en 1998, Craig Venter a fondé une compagnie privée, *Celera Genomics*, ayant comme objectif d'être en mesure de séquencer un génome humain en trois ans par une approche hautement robotisée. L'objectif de *Celera* était de devancer la communauté scientifique afin de breveter la séquence du génome humain pour la monétiser par la suite. Cette rivalité a officiellement pris fin le 26 Juin 2000 lorsque le président Bill Clinton, Francis Collins représentant le *consortium* international et Craig Venter ont annoncé conjointement la fin du séquençage brut du génome humain <sup>44,45</sup>.

Bien que s'agissant d'un mélange de génomes de plusieurs individus, ces succès ont signé le début d'une course encore aujourd'hui effrénée du séquençage de génomes individuels. En effet, depuis 2004, des machines dotées de débits supérieurs à ceux des

séquenceurs traditionnels Sanger sont apparus sur le marché. Ces séquenceurs nouvelle génération permettent de séquencer en parallèle plusieurs millions de fragments différents et *in fine*, permettent l'analyse du patrimoine génétique dans sa globalité dans un laps de temps jamais égalé.

## II. Le Séquençage de Nouvelle Génération

La seconde révolution dans le domaine du séquençage de l'ADN a eu lieu en 2004 avec l'apparition du premier séquenceur à haut-débit commercial de la société Roche (le séquenceur 454). Le séquençage à haut-débit, ou NGS (*Next* ou *New Generation Sequencing*) également appelé HTS (*High-Throughput Sequencing*) permet de réaliser le séquençage de millions de fragments en parallèle. Cette technologie fine permet par ailleurs, de séquencer ces fragments de façon clonale, conservant l'information portée par chaque fragment et d'identifier également des événements moléculaires présents à bas bruit dans un échantillon.

L'apparition des séquenceurs de deuxième génération a permis une réelle rupture en terme de débit, avec un passage du kilobase (kb) au gigabase (Gb) et donc l'accès au génome en un court délai. De plus, cette avancée technologique s'accompagne d'une chute drastique des coûts : le séquençage d'un génome est 50 000 fois moins cher en NGS qu'en Sanger<sup>38</sup>. A titre d'illustration, soixante ans seraient nécessaires pour séquencer un génome humain si un seul séquenceur capillaire de 1ère génération était utilisé.

Les deux principaux challenges technologiques du séquençage de deuxième génération sont d'arriver à séquencer en parallèle plusieurs millions de fragments et de le faire de façon clonale. Pour y parvenir, le déroulement d'un séquençage de deuxième génération suit généralement le même schéma. Premièrement, les molécules d'ADN sont clivées en courts fragments. Puis, des adaptateurs universels sont ajoutés aux extrémités de ces fragments. À l'aide de ces adaptateurs, l'ensemble des fragments est amplifié en parallèle. Cette étape de PCR est dite clonale car chaque fragment est amplifié de façon isolée. Après cette étape de préparation de la librairie et de PCR clonale, les fragments sont séquencés à proprement dit. Plusieurs solutions techniques ont été développées pour conduire l'ensemble de ces étapes.

### **a. L'amplification clonale : une étape nécessaire au séquençage de nouvelle génération**

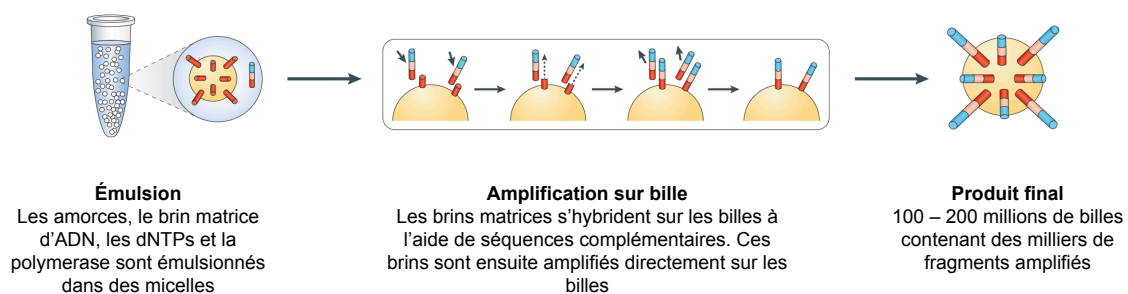
La création de milliers de copies identiques d'un fragment d'ADN est une étape nécessaire et préalable au séquençage. Cela permet l'obtention d'un signal différenciable du bruit de fond. A ce jour, trois méthodes principales d'amplification ont été développées :

- la PCR en émulsion (emPCR),
- l'amplification sur support solide



- et l'amplification en phase liquide

La **PCR émulsion** est une PCR réalisée dans une phase aqueuse dispersée en solution huileuse. Généralement, cette PCR se fait sur des billes émulsionnées contenant des séquences complémentaires aux adaptateurs universels. Pour réaliser cette PCR clonale, il y a création d'une micro-gouttelette de milieu aqueux contenant une bille, un fragment unique d'ADN ainsi que les réactifs nécessaires à l'amplification. Cette gouttelette représente finalement un micro-réacteur permettant la réalisation d'une amplification clonale de la molécule d'ADN contenue dans cette dernière. Après la PCR, un million de copies du fragment original se retrouvent attachés à la bille dans une gouttelette d'environ 1  $\mu\text{M}$  (Figure 5) <sup>47</sup>. L'émulsion est ensuite rompue et les billes sont distribuées sur un support solide en vue du séquençage. Ce type d'amplification est notamment employé par les sociétés Life Technologies et Roche.

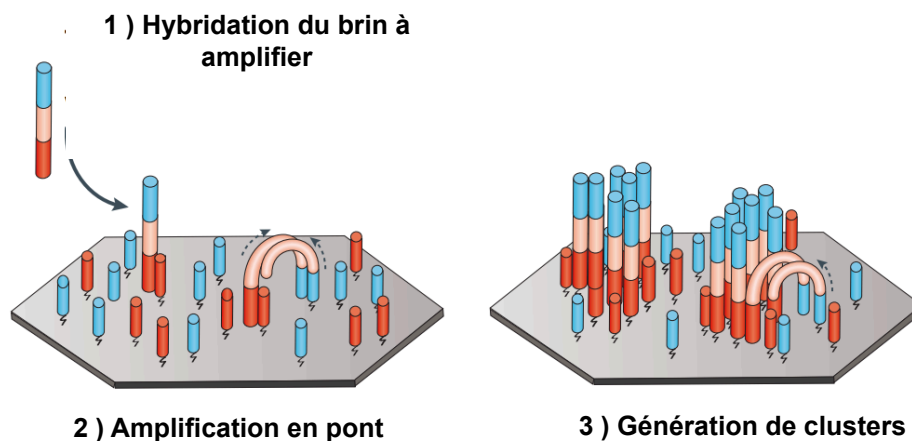


**Figure 5** : L'amplification clonale à l'aide de la PCR émulsion

La PCR en émulsion est une méthode permettant une amplification monoclonale d'un fragment d'ADN au sein d'une émulsion de phase aqueuse en solution huileuse. Les micelles formées contiennent l'ensemble des réactifs d'amplification (la polymérase, les dNTPs, des amorces, le fragment d'ADN d'intérêt flanqué d'adaptateurs universels) ainsi qu'une bille contenant des séquences complémentaires aux adaptateurs universels. L'amplification réalisée dans ce micro-compartiment permet de fixer de façon covalente les produits d'amplification à la bille. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

**L'amplification sur surface solide** permet, quant à elle, l'obtention de fragments amplifiés clonalement sectorisés (*cluster*) sur une surface solide, généralement en verre. Historiquement, cette étape a été réalisée par l'établissement de "colonies d'ADN" appelées polonies, résultat de l'amplification d'une seule molécule d'ADN par une polymérase au sein d'une matrice de gel<sup>48</sup>. Actuellement, la méthode principalement employée est l'amplification en pont. Les ADN fragmentés dotés de leurs adaptateurs sont distribués sur la surface de verre appelée *flowcell*. La *flowcell* possède à sa surface des séquences complémentaires aux adaptateurs nécessaires à la fixation des fragments d'ADN et permet à la fois le passage de fluide nécessaire à l'amplification et à la réaction de séquençage. La distribution des fragments sur la *flowcell* doit être précise de telle sorte que chaque fragment soit séparé

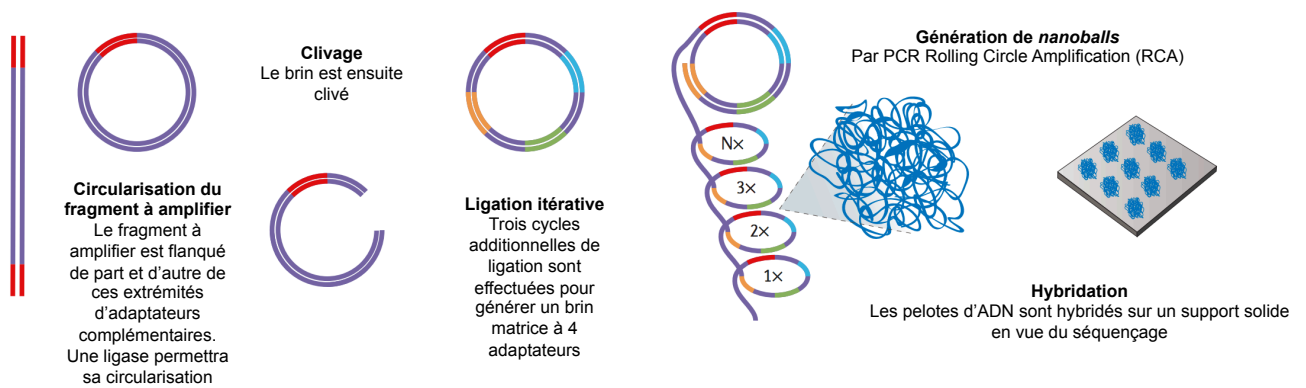
des autres fragments. Ensuite, l'ensemble des réactifs de PCR sont distribués au travers de la *flowcell*. L'amplification sur support solide peut ainsi produire 100 à 200 millions de *clusters* séparés spatialement où chacun des *clusters* représente une amplification clonale d'un fragment unique (Figure 6). Cette technologie est notamment employée par la société Illumina.



**Figure 6 :** L'amplification clonale en pont sur support solide

L'amplification sur support solide permet d'obtenir une mono-clonalité sous forme de *clusters* de molécules identiques sur un support solide appelé *flowcell*. Les fragments d'ADN flanqués de leurs adaptateurs universels sont préalablement distribués sur la *flowcell*. Cette dernière contient des séquences complémentaires aux adaptateurs permettant leurs attachements. Les éléments nécessaires à l'amplification sont ensuite distribués séquentiellement dans le système microfluidique de la *flowcell*. *In fine*, l'amplification sur support solide peut ainsi produire 100 à 200 millions de *clusters* séparés spatialement. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

**L'amplification en phase liquide** est une technologie récente, principalement utilisée par la compagnie *Complete Genomics*. Dans cette technologie, l'ADN est fragmenté puis des adaptateurs sont ajoutés aux extrémités permettant une circularisation par une ligase. Une fois circularisé, le fragment est amplifié sous forme de boucle par une PCR RCA (*rolling-circle amplification*). Le fragment amplifié forme une pelote d'ADN dénommée *DNA nanoball* (Figure 7). Vingt milliards de pelotes d'ADN sont ainsi générées au sein du même tube. Les *nanoballs* sont ensuite distribuées sur une surface solide de façon à ce que chaque balle soit unique et distinctes des autres.



**Figure 7** : Amplification clonale en phase liquide à l'aide d'une *Rolling Circle Amplification PCR* (RCA-PCR)

L'amplification en phase liquide conduit à la formation de pelote d'ADN où chaque pelote correspond à l'amplification d'un fragment unique d'ADN. Le fragment d'ADN à amplifier est préalablement flanqué d'adaptateurs pour permettre sa circularisation. L'ensemble subit ensuite une série de digestions enzymatiques et ligations pour préparer l'amplification à proprement dit. L'amplification est réalisée grâce à une PCR RCA (*rolling circle amplification*), permettant de former les pelotes d'ADN monoclonales. Figure adaptée de Goodwin et al. 2016<sup>46</sup>

## b. Le Séquençage Nouvelle Génération (NGS)

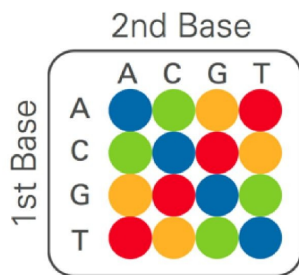
Après l'étape d'amplification clonale des fragments d'ADN, les échantillons sont à proprement séquencés. Les séquenceurs de deuxième génération permettent le séquençage en parallèle de courts fragments amplifiés clonalement. Il existe aujourd'hui deux approches distinctes de séquençage dans cette génération : le séquençage par ligation (*Sequencing By Ligation, SBL*) et le séquençage par synthèse (*Sequencing By Synthesis, SBS*).

### Séquençage par ligation

L'approche de séquençage par ligation (SBL) est une technologie sans polymérase basée sur deux éléments importants : l'hybridation de sondes fluorescentes et une ligation. La sonde fluorescente possède deux domaines notables : un premier domaine dégénéré tolérant à une complémentarité approximative et un second domaine porteur de la fluorescence et strictement complémentaire à une base ou à un dibase donné. L'émission de la fluorescence indique l'identité de la base ou des bases complémentaires incorporées à la position de la sonde. Pour réaliser la lecture moléculaire des fragments d'ADN préalablement amplifiés, le séquençage SBL débute par une étape l'hybridation d'une sonde appelée ancre, généralement complémentaire à l'adaptateur préalablement ajouté lors de l'amplification clonale. Au cycle 1 du séquençage, la sonde fluorescente complémentaire est liée par une ligation à l'ancre. Puis, la sonde est ensuite clivée libérant ainsi la fluorescence,

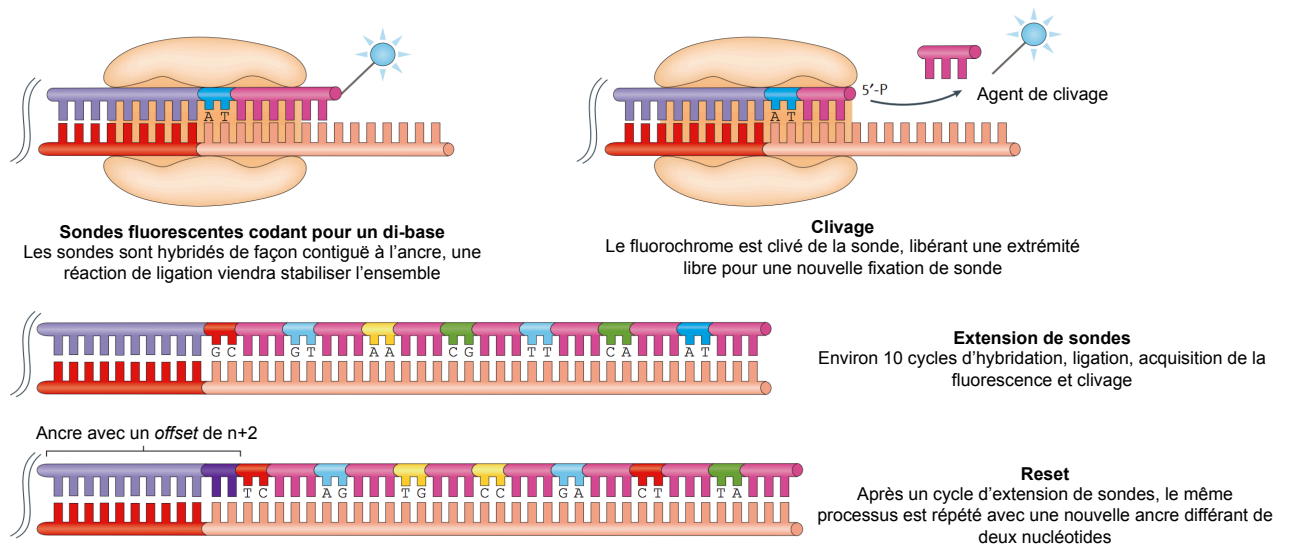
ce qui permet la détermination du nucléotide présent à cette position et de poursuivre le séquençage. Le séquençage se poursuit de façon itérative pendant une dizaine de cycles d'ajout d'une nouvelle sonde, du clivage du fluorophore et une ligation à une nouvelle sonde. *In fine*, ces étapes permettent d'identifier une base ou deux bases selon le système employé toutes les 5 à 8 bases . Pour obtenir la séquence complète du fragment, le séquençage est repris à partir de l'adaptateur universel en employant un système d'ancres dégénérées créant le décalage de lecture.

La technologie SoLiD (*Sequencing by Oligonucleotide Ligation and Detection; SoLiD*) emploie cette approche SBL. Ce système est basé sur une amplification par PCR émulsion suivie de l'hybridation/ligation d'octanucléotides fluorescents codant pour un di-base. Cette technologie utilise quatre fluorochromes pour coder seize combinaisons possibles (Figure 8). Cette technologie permet ainsi que chaque base soit donc lue deux fois par deux réactions de ligations indépendantes (Figure 9)



**Figure 8** : Système de codage du dibase employé dans la technologie SoLiD

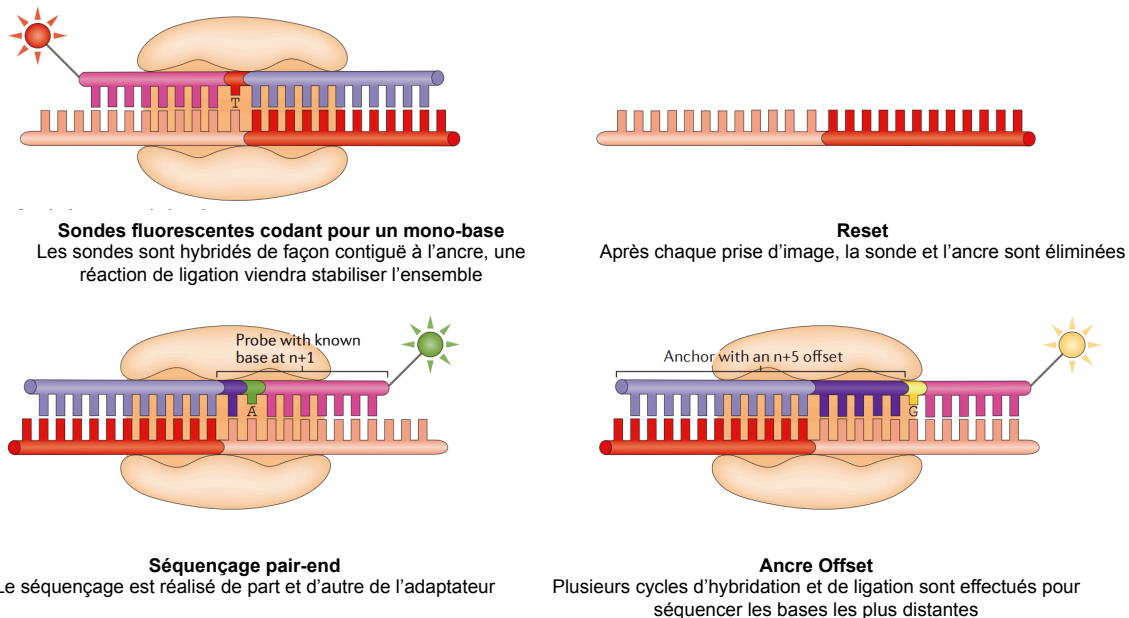
Quatre fluorochromes permettent l'encodage de seize combinaisons de dibases possibles.



**Figure 9** : Le séquençage SBL de la société SoLiD (Thermo Fisher)

La technologie de séquençage par ligation de SoLiD est basée sur l'utilisation de quatre sondes fluorescentes codant chacune pour un dibase. Les fragments d'ADN à séquencer sont flanqués à leurs extrémités d'un adaptateur universel. Pour réaliser le séquençage, une amorce complémentaire de l'adaptateur est utilisée pour permettre la fixation des sondes fluorescentes. Le clivage de la sonde libère le fluorochrome permettant à la fois de déduire le dibase à cette position et également de poursuivre le séquençage par la fixation d'une autre sonde fluorescente complémentaire. Figure adaptée de Goodwin et al. 2016<sup>46</sup>

Une autre approche SBL est celle employée par *Complete Genomics* : la technologie cPAL (*combinatorial probe-anchor ligation*). L'approche est assez proche de la technologie SOLiD étant donné qu'elle a été développée dans le même laboratoire (*Church Laboratory of Harvard Medical School*). L'objectif était de produire un système abordable et libre de droit. La différence principale avec la technologie SOLiD réside dans les sondes : des nonamères fluorescents codant pour une seule base. Après une amplification en phase liquide, le cycle de séquençage débute par l'hybridation de l'ancre à l'un des adaptateurs. Quand la sonde s'apparie avec la bonne séquence et de façon adjacente à l'ancre, il y a ligation stabilisant l'ensemble. La *flowcell* est ensuite interrogée par une détection de 4 couleurs (une couleur par base) (Figure 10).



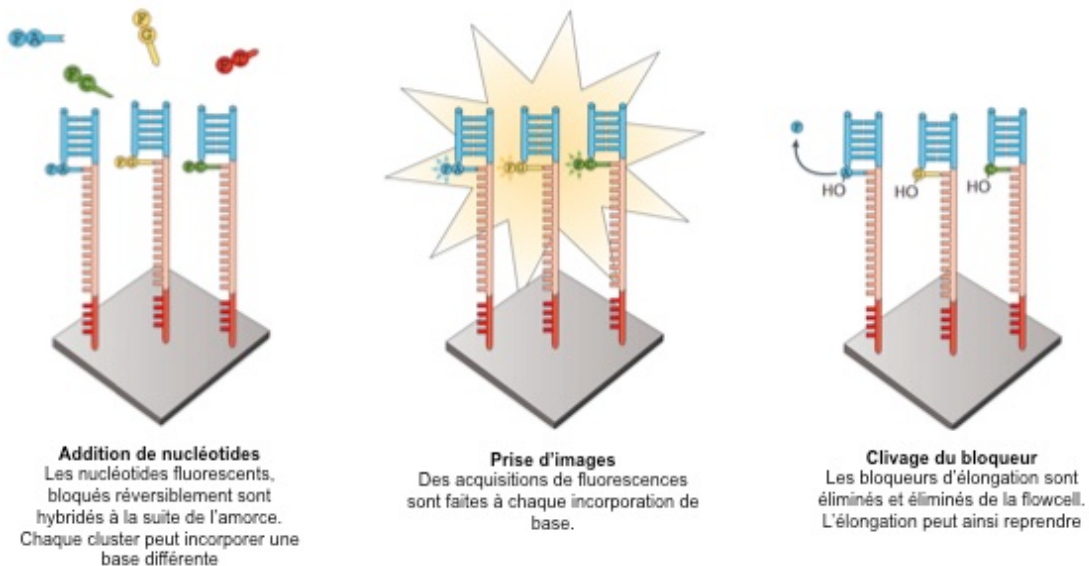
**Figure 10 :** Le séquençage développé par Complete Genomics basé sur la technologie cPAL

Le système de séquençage développé par *Complete Genomics* est basé sur l'utilisation d'une sonde fluorescente codant pour une base. Les fragments d'ADN à séquencer sont flanqués d'un adaptateur universel. Pour réaliser le séquençage, une amorce complémentaire de l'adaptateur est utilisée pour permettre la fixation des sondes fluorescentes. Le clivage de la sonde libère le fluorochrome, ce qui permet à la fois de déduire la base à cette position et également de poursuivre le séquençage par la fixation d'une autre sonde fluorescente complémentaire. Figure adaptée de Goodwin et al. 2016<sup>46</sup>

### Séquençage par synthèse

Une seconde approche de séquençage de deuxième génération est le séquençage par synthèse (SBS). A l'instar du séquençage Sanger et à l'inverse du séquençage SBL précédemment décrit, le séquençage SBS repose sur l'utilisation d'une polymérase. Il est possible de catégoriser le séquençage par synthèse en deux sous-ensembles : la technologie CRT (*Cyclic Reversible Termination*) basée sur l'utilisation de nucléotides fluorescents et la technologie SNA (*Single Nucleotide Addition*) basée sur la quantification d'un sous-produit de la réaction d'élongation. L'approche CRT (*Cyclic Reversible Termination*) se définit par l'utilisation de molécules terminatrices de séquence réversibles. L'ADN amplifié est amorcé par une séquence complémentaire de l'adaptateur, ce qui permettra à la polymérase de débiter l'élongation de façon universelle sur toutes les régions à séquencer en parallèle. Pendant chaque cycle, un mélange des quatre bases marquées et bloquées transitoirement en 3' est introduit dans le milieu. Une base unique est incorporée

à chaque cycle puisque chaque incorporation bloque temporairement l'élongation. L'acquisition de la fluorescence est réalisée à ce moment permettant de déterminer la base incluse. La fluorescence et le groupement bloqueur sont éliminés pour permettre un nouveau cycle de lecture. Ces cycles se poursuivent de façon itérative généralement de 75 à 300 fois (Figure 11).



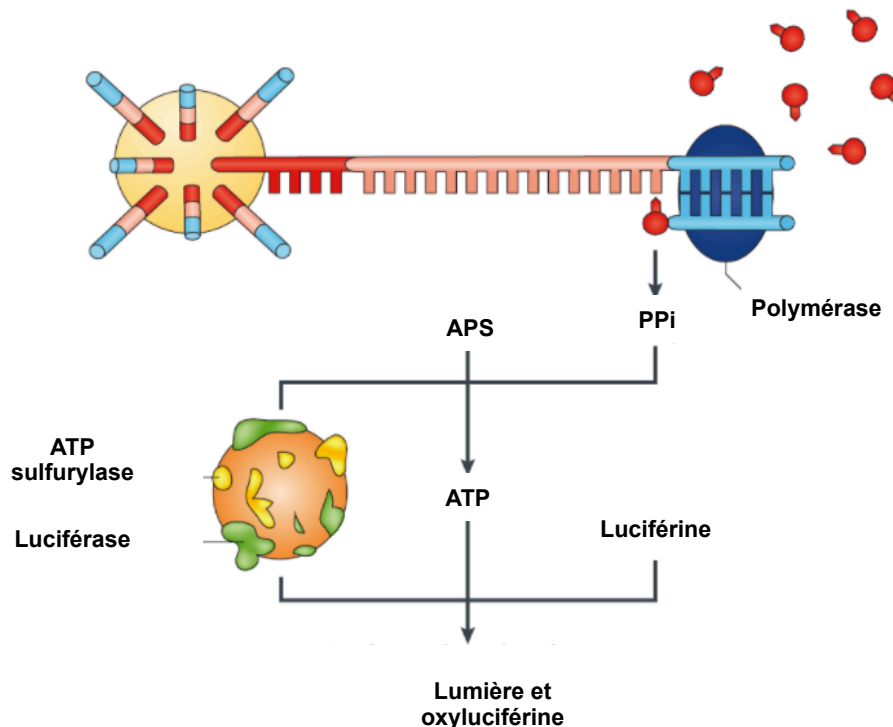
**Figure 11 :** Le séquençage basé sur l'utilisation de nucléotides terminateurs de séquences réversibles

Le séquençage CRT (*Cyclic Reversible Termination*) est basé sur l'utilisation de dNTP terminateur de séquences bloqué réversiblement. Les fragments monoclonalement amplifiés étant fixés de façon covalente à la flowcell sous forme de *clusters*, le mélange réactionnel de séquençage est introduit dans le milieu. La mesure de fluorescence émise lors du clivage du bloqueur du dNTP terminateur permet de déterminer la base. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

Actuellement, la technologie CRT d'Illumina est la technologie la plus répandue. La société Illumina propose des instruments allant des petits séquenceurs de paillasse aux séquenceurs de plus grande taille permettant le séquençage de plusieurs génomes simultanément. Dans le système d'Illumina, l'identification des bases incorporées est permise par un système de microscopie 'TIRF' (*total internal reflection fluorescence*) <sup>49</sup>. Dans la plupart des instruments Illumina, quatre canaux lasers sont employés, un pour chaque base. A noter que dans leurs derniers appareils (Next-Seq et Mini-Seq), Illumina utilise seulement deux canaux lasers pour la détection des quatre bases.

A l'inverse des approches précédentes, les technologies *Single Nucleotide Addition* (SNA) sont quant à elles basées sur la quantification d'un produit de la réaction d'élongation. Il peut s'agir du pyrophosphate (PPi) ou des protons H<sup>+</sup> libérés par la polymérase lors de l'ajout d'une base. Ce système impose donc de ne pas injecter les quatre bases ensemble

mais séquentiellement. Au début des années 2000, une compagnie américaine *Life Sciences*, rachetée plus tard par Roche, a eu l'idée d'employer le pyroséquençage pour le séquençage à haut débit. Techniquement, le pyroséquençage est basé sur la conversion lumineuse du pyrophosphate libéré pendant le processus d'élongation de l'ADN et la quantification de ce dernier pour déterminer la base incorporée<sup>50,51</sup>. Le pyrophosphate libéré lors de l'ajout d'une base est converti en ATP par une enzyme appelée sulfurylase. L'ATP est utilisé ensuite comme substrat par la luciférase, produisant l'émission d'un photon en quantité proportionnelle à la quantité de pyrophosphate<sup>52</sup> (12). Cette technologie utilise une amplification clonale par émulsion. Une fois l'émulsion rompue, les billes contenant les fragments amplifiés clonalement sont réparties sur une plaque contenant de multiples puits, des "PicoTiterPlate". La fluorescence émise à l'introduction d'une des quatre bases dans chaque puits est mesurée par une caméra CCD (*charge-coupled device*) pour déterminer la séquence. En commercialisant leur appareil, le 454, Life Science a été le premier à mettre en place l'automatisation et la parallélisation d'un grand nombre de réactions de pyroséquence sur le marché, produisant une grande quantité d'ADN séquencé en un seul run<sup>53</sup>.



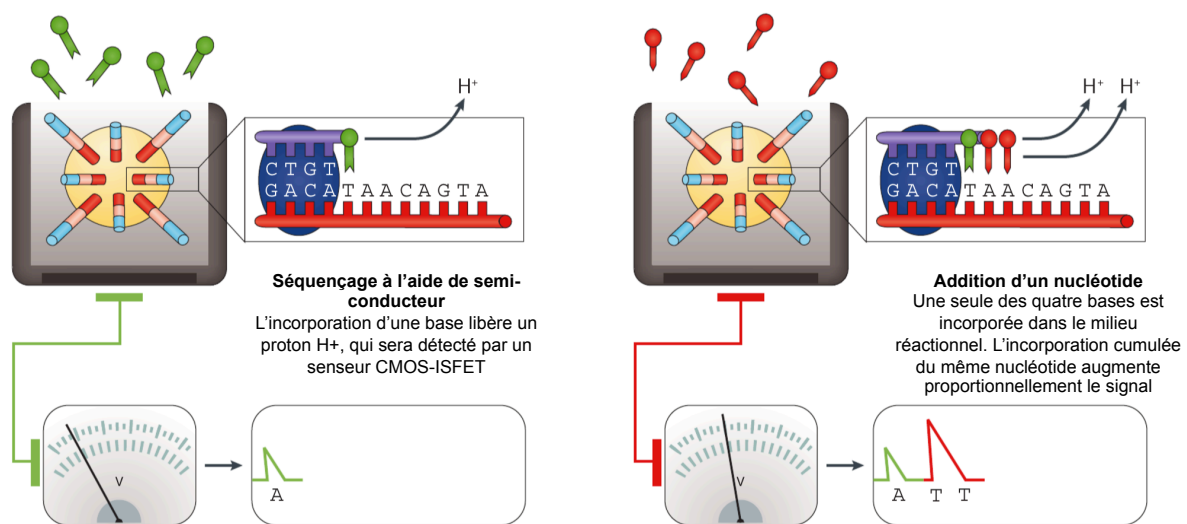
**Figure 12** : Le séquençage à haut-débit par pyroséquençage

Après un enrichissement de la matrice des fragments d'ADN à séquencer grâce à l'utilisation d'une PCR en émulsion, les billes sont disposées sur une plaque de microtitrage avec des amorces et un cocktail d'enzymes. Au cours du premier cycle, les bases sont introduites séquentiellement dans le milieu. Chaque base complémentaire est incorporée par une ADN polymérase dans un brin nouvellement synthétisé. L'un des sous-



produits alors générés par cette réaction est une molécule de pyrophosphate (PPi). Le PPi généré est ensuite transformé en molécule d'ATP à partir d'adénosine 5'phosphosulfate (APS). L'ATP est ensuite utilisé comme un cofacteur pour la conversion de la luciférine en oxyluciférine par la luciférase. L'émission de lumière accompagnant cette réaction est ensuite détectée par une caméra pour déterminer la base ainsi incorporée. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

Une seconde approche de séquençage SNA a été développée par Thermo Fisher Scientific. Leur appareil, l'Ion Torrent, est capable de séquencer grâce à la mesure de la variation de pH du milieu réactionnel. En effet, chaque incorporation de base pendant la phase d'élongation, libère un proton H<sup>+</sup>. L'amplification clonale préalable permet d'obtenir une modification significative de pH. Ainsi, la variation de pH est proportionnelle aux nombres de bases incorporées permettant la détection à la fois de la base incorporée et d'homopolymères (Figure 13).



**Figure 13** : Le séquençage à haut-débit basé sur la variation de pH : l'Ion Torrent

Après l'enrichissement par PCR émulsion, les billes sont réparties dans une plaque de microtitrage où au maximum une bille unique occupera un seul puits. Lors de la réaction d'élongation, à chaque base incorporée, un ion H<sup>+</sup> est généré. La libération de H<sup>+</sup> entraînera un changement de pH, qui sera détecté dans le puits par un semi-conducteur d'oxyde métallique complémentaire intégré (CMOS) et un dispositif à transistor à effet de champ sensible aux ions (ISFET). Ainsi, les nucléotides sont introduits un à un, il y a quantification de la variation de pH dans chaque puits, puis les bases non incorporées sont éliminées par lavage et la base suivante est ajoutée. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

### Comparaison des technologies de deuxième génération

La technologie de séquençage par ligation (SBL) utilisée par SOLiD et Complete Genomics offre une précision de séquençage élevée (99,99 %), puisque chaque base est lue plusieurs fois par des réactions indépendantes d'hybridation et de ligation. Cependant, certaines publications rapportent un nombre élevé de faux positifs et de nombreux vrais variants sont parfois manqués <sup>54,55</sup>. Principalement, ces artefacts sont dus à des sous

représentations des régions riches en AT. Les limites d'utilisation de séquençage SBL sont liées à la taille de leurs reads (inférieur à 75 pb pour SOLiD et à 100 pb pour le Complete Genomics). Cela ne permet donc pas d'identifier des variants structuraux ni de réaliser de l'assemblage *de novo* de génomes. Malgré cela, Complete Genomics à l'aide de sa technologie cPAL, a annoncé son intention de concurrencer Illumina en terme de coût et de débit. Il est encore un peu tôt pour en juger, étant donné que cette technologie est commercialisée depuis 2016 et est quasi-exclusivement implantée en Chine.

A l'heure actuelle, le marché des séquenceurs de deuxième génération est dominé par Illumina grâce à sa compatibilité d'activité avec un grand nombre de laboratoires de recherche et de diagnostic. En effet, la suite commerciale des instruments Illumina permet de faire du séquençage de faible débit avec des appareils tels que le MiniSeq jusqu'à l'ultra débit avec le HiSeq X capable de séquencer 1,800 génomes humains par an. Etant donné que les plateformes Illumina utilisent une approche *Cycle Reversible Termination* (CRT), elles sont moins susceptibles de faire des erreurs liées aux homopolymères comparé aux plateformes *Single Nucleotide Addition* (SNA). Bien que la technologie Illumina possède une excellente qualité de séquençage, leur chimie rencontre encore quelques différences de représentation des régions riches en AT par rapport aux régions riches en GC.

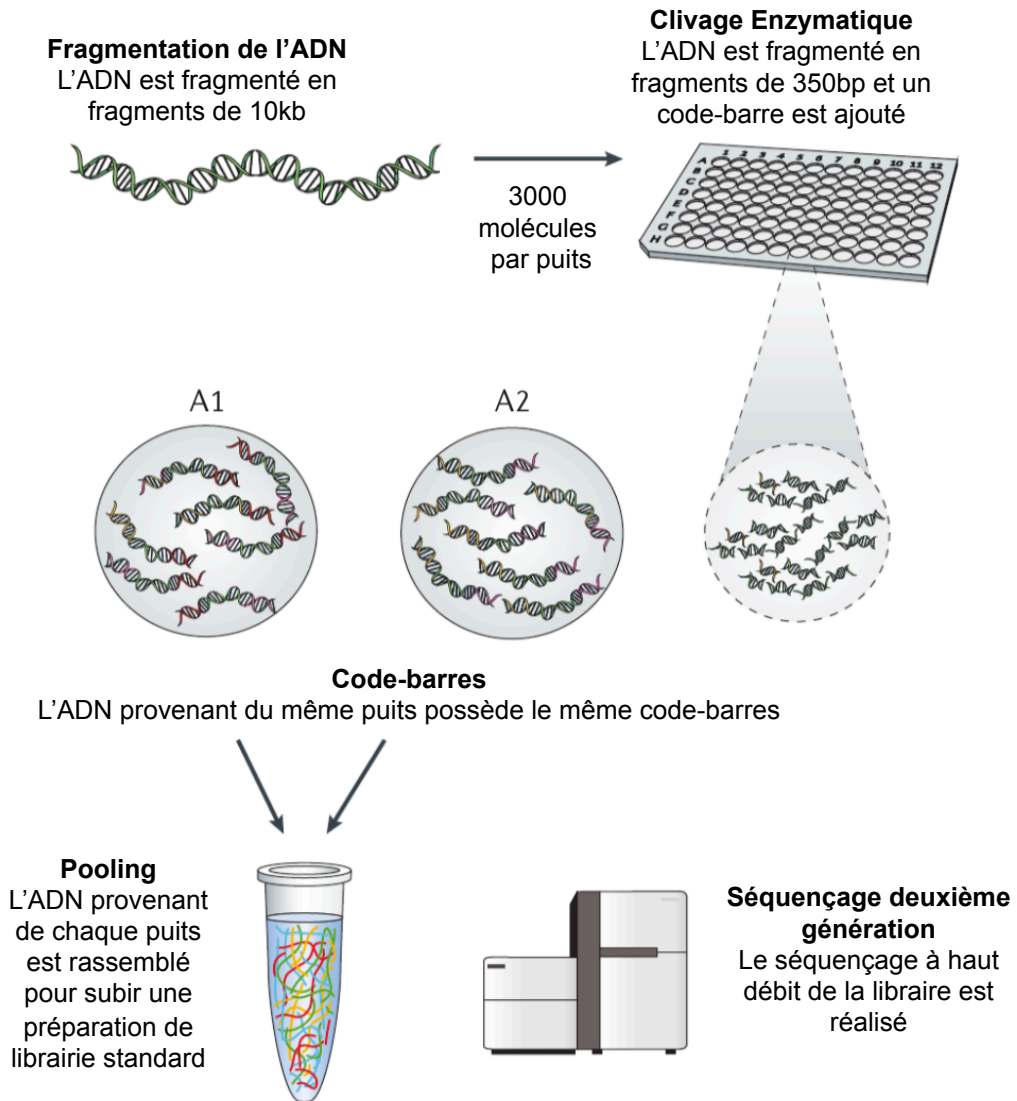
Les technologies SNA (454 et Ion Torrent) offrent en revanche des lectures plus longues, 700 et 400 pb respectivement, fournissant un avantage dans la lecture de régions génomiques plus complexes comme les remaniements. Ces deux technologies SNA partagent néanmoins les mêmes inconvénients, principalement une mauvaise lecture des régions répétées comme les homopolymères (supérieur à 6-8 pb) et la mauvaise détection des indels <sup>46</sup>. Il est intéressant de noter que l'Ion Torrent possède aujourd'hui les durées de lecture les plus courtes du marché (de 2 à 7h) grâce à sa chimie SNA, le rendant très compétitif sur le marché des séquenceurs deuxième génération.

### **Vers une troisième génération de séquençage : les reads longs synthétiques**

Comme exposé précédemment, l'une des limites du séquençage à haut débit deuxième génération est l'impossibilité de séquencer des longues molécules d'ADN. Pour palier à ce défaut, plusieurs sociétés ont développé des approches artificielles qualifiées de longues lectures synthétiques (*synthetic long read*). La technologie des longs *reads* synthétiques est basée sur un système de code-barres nucléotidiques pour identifier de courts fragments comme provenant de la même grande molécule. Ces courts fragments sont ensuite séquencés sur des séquenceurs de 2<sup>ème</sup> génération.

Cette approche débute par une répartition de larges fragments dans de petits compartiments. Chaque molécule est clivée en molécules plus petites et marquée d'un code

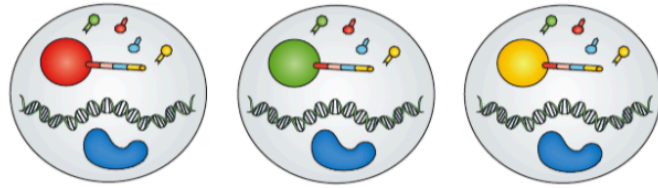
barre nucléotidique propre à chaque puits. Puis, l'ensemble des fragments marqués est rassemblé pour être séquencé. Dans l'analyse informatique découlant, les fragments partageant le même code barre seront rassemblés comme appartenant à la même grande molécule. Il s'agit finalement d'un démultiplexage basé sur l'origine du fragment. Cette approche, conceptuellement proche du séquençage BAC à BAC, permet de résoudre des problèmes d'assemblage. Actuellement, deux systèmes sont disponibles pour générer des longs reads synthétiques : le système Moleculo d'Illumina et le système 10X Genomics. Le système Moleculo ne requière pas d'instrument particulier hormis des consommables (Figure 14) alors que le système 10X Genomics (composé du GemCode et du Chromium) utilise une étape d'émulsion et donc nécessite des instruments de micro fluidique (Figure 15). Avec 1 ng d'ADN comme matériel de départ, la technologie 10X Genomics permet d'obtenir des fragments de 100 kb dans des micelles appelées GEMS. Chaque GEMS contient au maximum 1 fragment avec un code barre unique par micelle (Figure 15).



**Figure 14 :** Séquençage de longs reads synthétiques d'Illumina : Moleculo

Les matrices d'ADN génomique sont fragmentées en fragments de 8-10 kb. Ils sont ensuite divisés en une plaque de micro titrage de telle sorte qu'il y ait environ 3 000 modèles dans un seul puits. Dans la plaque, chaque fragment est à nouveau fragmenté en plus petits fragments de 350 pb. Sur chacun des sous fragments est ajouté un code à barres unique à chaque puits. L'ensemble des ADN fragmentés est ensuite rassemblé pour être séquençé sur des séquenceurs deuxième génération. Cette étape permet donc de conserver l'origine fragmentaire de sous-fragments générés. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

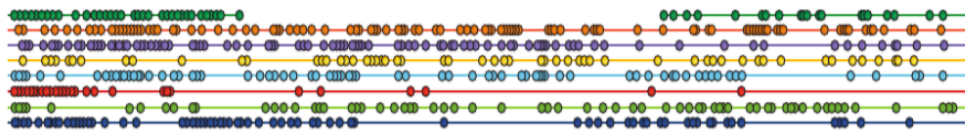
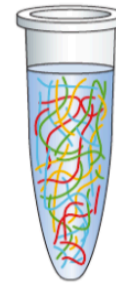
**PCR Emulsion**  
Arbitrairement, de longues molécules d'ADN sont émulsionnées avec des amorces possédant un code-barres, des dNTPs et des enzymes



**GEMs**  
Chaque micelle contient un code-barres uniques parmi les 750 000

**Amplification**  
Les longs fragments sont amplifiés et produisent un fragment marqué d'environ de 350 bp.

**Pooling**  
L'émulsion est rompue et les ADN sont rassemblés. Puis, ces derniers subissent une préparation classique de librairie.



**Figure 15** : Le séquençage synthétique développé par 10X Genomics

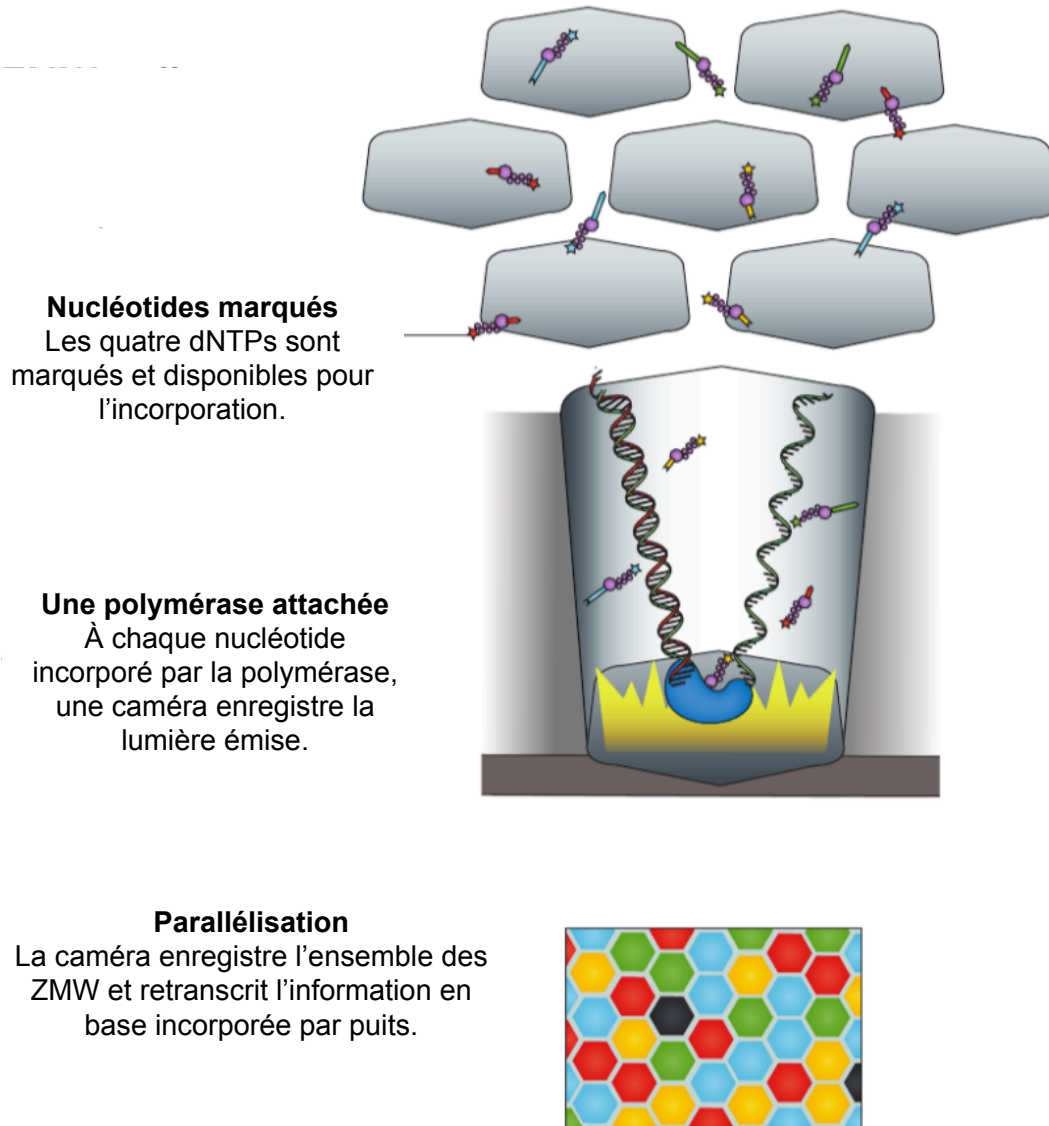
Avec seulement 1 ng de matériel de départ, le GemCode peut répartir arbitrairement de gros fragments d'ADN, jusqu'à ~ 100 kb, dans des micelles (GEM) contenant des séquences d'adaptateur et des code-barres. Les GEM contiennent au maximum une molécule du génome et un code-barres propre à chaque GEM. Les fragments d'ADN sont clivés en sous fragments. Une amplification est réalisée dans la GEM permettant à chaque sous-fragment d'être flanqué du code-barres de la GEM source. Après séquençage, les lectures sont alignées et liées ensemble pour retrouver leurs origines fragmentaires. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

### c. La révolution du séquençage de molécule unique : *The next-next generation sequencing*

Le principal inconvénient des séquenceurs de deuxième génération est la nécessité d'amplifier préalablement les fragments à séquencer. En effet, cette étape est potentiellement génératrice d'artefacts notamment liés aux erreurs de la polymérase. C'est dans ce contexte que de nouvelles technologies qualifiées de 3<sup>ème</sup> génération ont fait leur apparition. Ces séquenceurs s'affranchissent de l'amplification clonale préalable et sont capables de lire directement un fragment d'ADN non amplifié. Cette technologie est appelée SMS (*Single Molecule Sequencing; SMS*). Il ne s'agit donc plus d'un système artificiel mais réellement de longues lectures allant jusqu'à 200 Kb. Ce type de séquençage facilite ainsi la réalisation d'assemblage *de novo* de génomes, la détection d'éléments structuraux de grandes tailles comme les variations du nombre de copies (CNV), les inversions, et les translocations.

La première technologie SMS a été développée dans le laboratoire de Stephen Quake et commercialisée plus tard par Helicos BioSciences. Dans ce système, l'ADN est dilué de telle sorte qu'une molécule unique d'ADN soit partitionnée dans une chambre réactionnelle. Chacune de ces chambres possède une polymérase immobilisée au fond du puits. Le séquençage est réalisé grâce à l'emploi de dNTP terminateurs réversibles, appelés "*virtual terminator*". La fluorescence émise pendant l'incorporation dans le milieu réactionnel est ensuite mesurée pour déterminer la base incorporée.

A l'heure actuelle, la technologie de troisième génération de séquençage la plus avancée est probablement celle des séquenceurs PacBio développés par Pacific Biosciences. Il s'agit d'une technologie SMS où la lecture est réalisée en temps réel (*Single Molecule Real Time, SMRT*). Le PacBio emploie des nanostructures appelés *zero-mode waveguides* ou ZMWs. Ces véritables petites chambres de l'ordre du zeptolitre ( $10^3 \text{ nm}^3$  soit  $10^{-21} \text{ L}$ ) sont capables de détecter les ondes nanophotoniques, c'est à dire de la fluorescence dans un très faible volume et à très bas bruit. A chaque nucléotide fluorescent incorporé par l'ADN polymérase, la fluorescence de la ZMW est mesurée et retranscrite en base (Figure 16). L'avantage de cette technologie est sa rapidité : le séquençage est aussi rapide que l'ADN polymérase puisqu'il n'y a pas de clivage des nucléotides fluorescents réversibles par exemple comme dans la chimie Illumina. Enfin, le PacBio permet de réaliser des lectures de 10 kb, arguments importants pour les problématiques d'assemblage *de novo*.

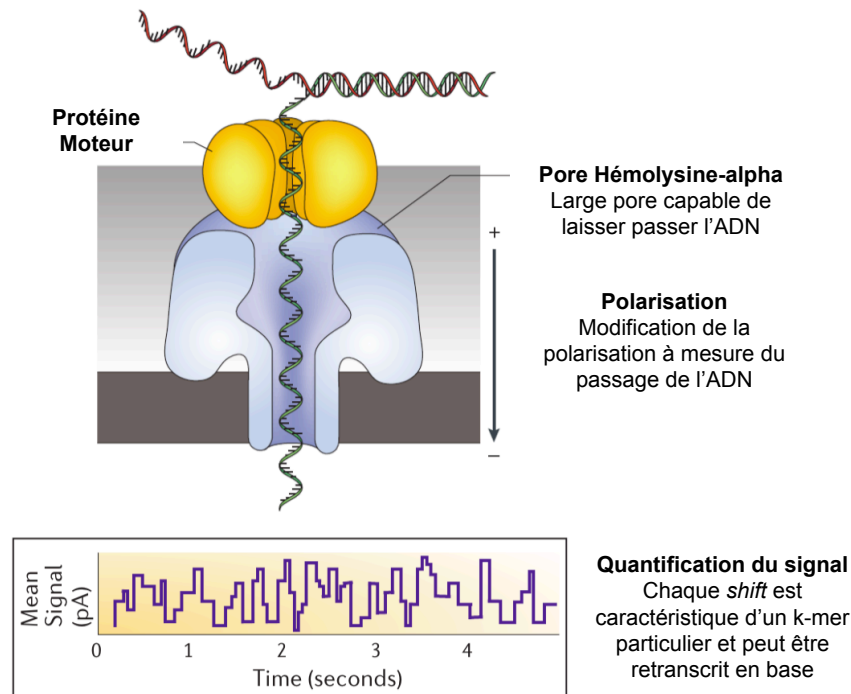


**Figure 16** : Séquençage d'une seule molécule en temps réel (SMRT) de Pacific Biosciences (PacBio)

Ce séquençage se déroule dans une nanostructure appelée *Zero-Mode Waveguide* (ZMW) possédant une polymérase fixée de façon covalente au fond du ZMW. La molécule d'ADN est insérée dans la ZMW puis est dénaturée pour permettre son passage à travers la polymérase fixée. Le séquençage est permis par l'utilisation d'un mélange de nucléotides fluorescents. Lors de l'incorporation d'un nucléotide, de la fluorescence est émise. La fluorescence de chaque ZMW est enregistrée par une caméra permettant de déduire les bases incorporées en temps réel. Figure adaptée de Goodwin et al. 2016 <sup>46</sup>

Une seconde approche de type SMS émergente est le séquençage par nanopores. Ce séquençage met en jeu deux éléments : un nanopore d'origine protéique formant un tube de quelques nanomètres et une membrane lipidique synthétique polarisée. Chaque base transitant au travers du pore crée une modification caractéristique de la polarité de la membrane. Le suivi dans le temps des modifications de la membrane permet de déterminer la séquence nucléotidique <sup>56</sup> (Figure 17). L'une des entreprises les plus avancées sur l'usage

des nanopores pour le séquençage est *Oxford Nanopore Technologies* (ONT) avec son séquenceur appelé "minION". Ce séquenceur a la particularité d'être de la taille d'une clef USB (3 cm×10cm) et de coûter 1000 dollars (Figure 18).



**Figure 17** : Séquençage par nanopore d'Oxford Nanopore Technologies (ONT)

Ce séquençage repose sur une membrane lipidique polarisée et un nanopore protéique. Le séquençage débute par une fragmentation de l'ADN en large fragments de 8-10 kb et sur la fixation d'un adaptateur à une extrémité de la molécule à séquencer. Cette séquence est fortement affine pour le pore et permet donc de diriger le fragment d'ADN vers le nanopore. Lorsque l'ADN passe à travers le pore, cela modifie ces caractéristiques. Il est alors possible d'observer des changements des caractéristiques du pore et de la membrane. Chaque base possède des caractéristiques propres (magnitude et la durée du changement). Figure adaptée de Goodwin et al. 2016<sup>46</sup>

L'objectif de Oxford Nanopore Technologies est de développer en 2017, le PromethION, un instrument à 48 *flowcell*, avec chacune 3 000 pores séquençant à 500 pb par seconde. Cet appareil pourrait donc générer 2 à 4 Tb en deux jours, ce qui le placerait en compétition avec le HiSeq X d'Illumina dans la course à l'ultra haut débit. Néanmoins, il est important de pondérer ces propos en précisant que cette technologie produit un grand nombre d'erreurs de lecture<sup>57</sup>.

L'utilisation de l'approche des longs reads a permis, récemment, à l'équipe de Chaisson et ses collaborateurs, d'ajouter plus d'un mégabase à la nouvelle séquence du génome de référence humain GRCh37. Ils ont identifié plus de 26 000 insertions/délétions (indel) supérieures à 50 bp, fournissant la référence la plus complète du génome humain disponible à l'heure actuelle<sup>58</sup>. Un dernier exemple démontrant la puissance de cette



génération de séquenceurs est l'utilisation du MinION pendant l'épidémie d'Ebola, virus touchant l'Afrique de l'Ouest, infectant 28 599 de cas et causant la mort de 11 299 personnes. Quick et ces collaborateurs ont génotypé 142 échantillons du virus Ebola entre mars et octobre 2015 avec une analyse en moins de 24h, étant donné que le séquençage prenait entre 15 à 60 minutes, permettant de suivre l'évolution du virus <sup>59</sup>.



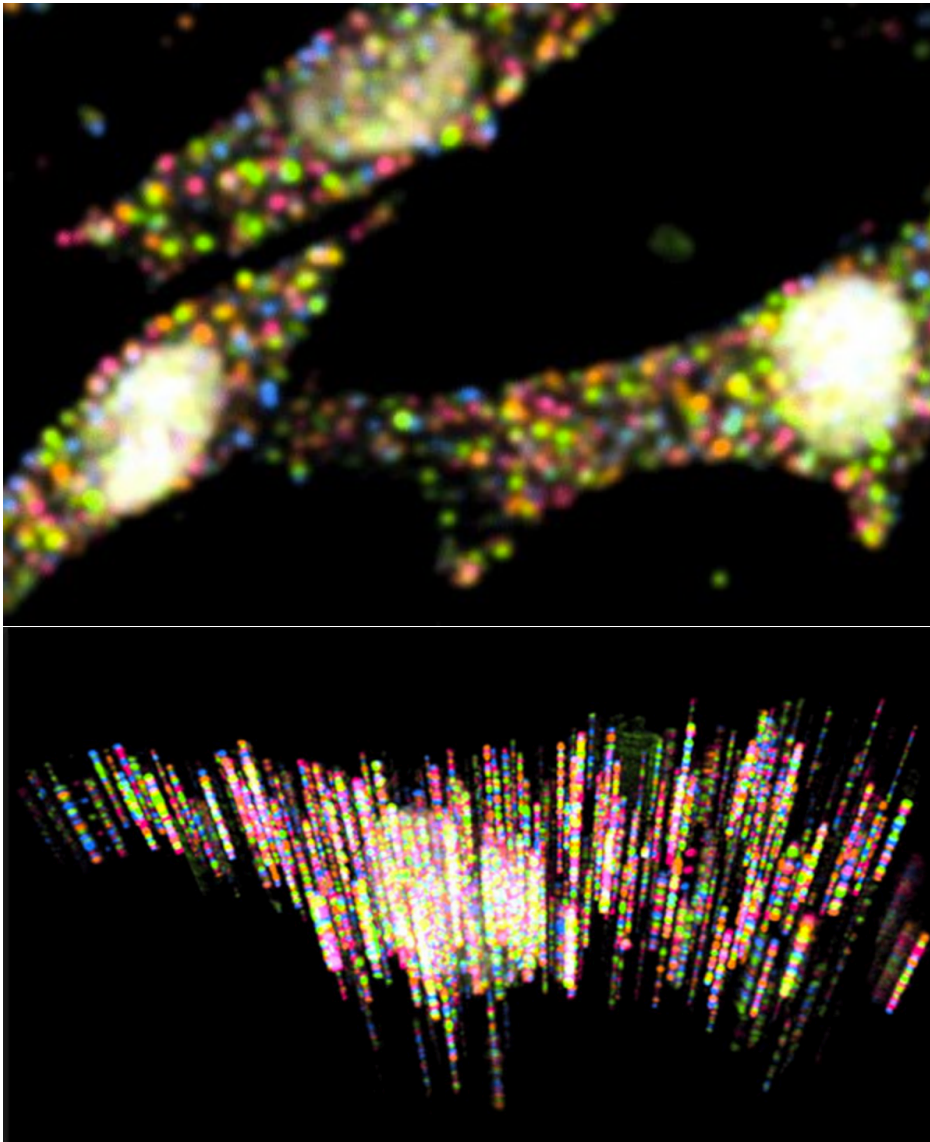
**Figure 18** : Photographie du MinION d'Oxford Nanopore Technologies

#### **d. Evolution rapide des technologies et des coûts de séquençage**

Le séquençage n'a cessé d'évoluer rapidement ces dernières années, qu'il s'agisse de débit, de chimie ou également d'approche conceptuelle.

Très récemment, une nouvelle approche a fait son apparition : le séquençage de cellule unique *one cell* et le séquençage directement au sein de la cellule *in situ* <sup>60</sup>. L'intérêt de ces approches est de mieux appréhender la physio-pathologie à l'échelle de la cellule sans être exposé à l'hétérogénéité moléculaire et tissulaire, mais également d'être en mesure de déterminer spatialement la localisation des ARN. Le pionnier de cette génération est George Church du *Wyss Institute* d'Harvard avec une technologie appelée FISSEQ pour *Fluorescent In Situ SEQuencing* <sup>61</sup>. L'idée est de réaliser une coupe fine d'un tissu, de la fixer

sur un support immobile, de réaliser une *reverse transcription* (RT-PCR) de l'ensemble des ARN présents par une amplification particulière appelée *rolling circle amplification* (RCA), de fixer les ADN complémentaires générés, pour enfin, séquencer par la technologie SoLiD, *in situ* les ADNc fixés. L'acquisition des images permet de déterminer la position spatiale d'un ARN donné (Figure 19). Bien que ces notions soient très récentes, des compagnies de biotechnologie telles que Fluidigm développent des stratégies de séquençage "*one cell*" ou "*in cell*".



**Figure 19** : Images illustrant un séquençage FISSEQ

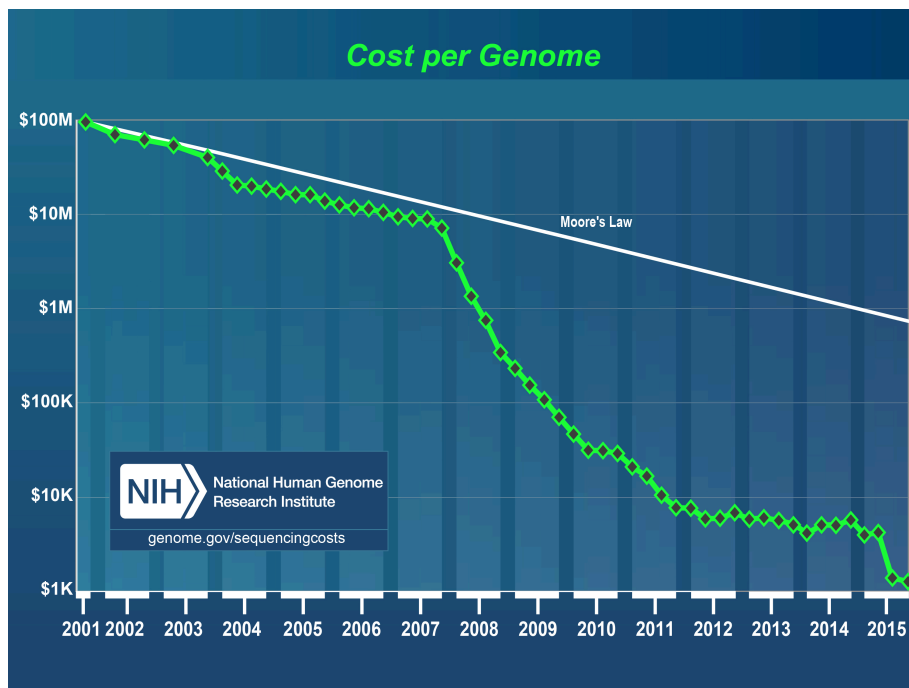
Le séquençage FISSEQ (*Fluorescence In Situ SEQuencing*) permet d'obtenir la position spatiale d'un ARNm au sein d'une cellule. Pour cela, la cellule a préalablement été fixée et les ARNm retrotranscrits en ADNc. Ensuite, un séquençage SoLiD est réalisé. Ainsi, il est possible de déterminer la position spatiale de chaque ARNm. En haut, une cellule en cours de séquençage, chaque point lumineux représentant un ADNc en cours de séquençage. En bas, une reconstruction informatique de la séquence en 3D où chaque colonne lumineuse représente une molécule d'ADNc séquencée avec sa position spatiale.

Dans la course effrénée au séquençage à haut débit, et à très haut débit, de nombreuses compagnies développent des nouveaux séquenceurs et de nouvelles approches de séquençages. Parmi ces développements, on peut citer les sociétés privées suivantes :

- la start-up Genia, soutenue par Roche, a annoncé développer une nouvelle approche de *Single Molecule Sequencing* par nanopore.
- Illumina a très récemment annoncé l'arrivée de Firefly, un séquenceur avec des puces semi conductrices (CMOS) dans des nanocompartiments avec une chimie de séquençage par synthèse à une seule couleur (*one-channel sequencing-by-synthesis chemistry*).
- GnuBio (Bio-Rad) est en train de développer quant à lui, une approche de séquençage par transfert d'énergie entre molécule fluorescente (*Fluorescence Resonance Energy Transfer, FRET*)<sup>52</sup>.
- La société Electron Optica annonce quant à elle une technologie basée sur de la microscopie électronique<sup>63</sup>.

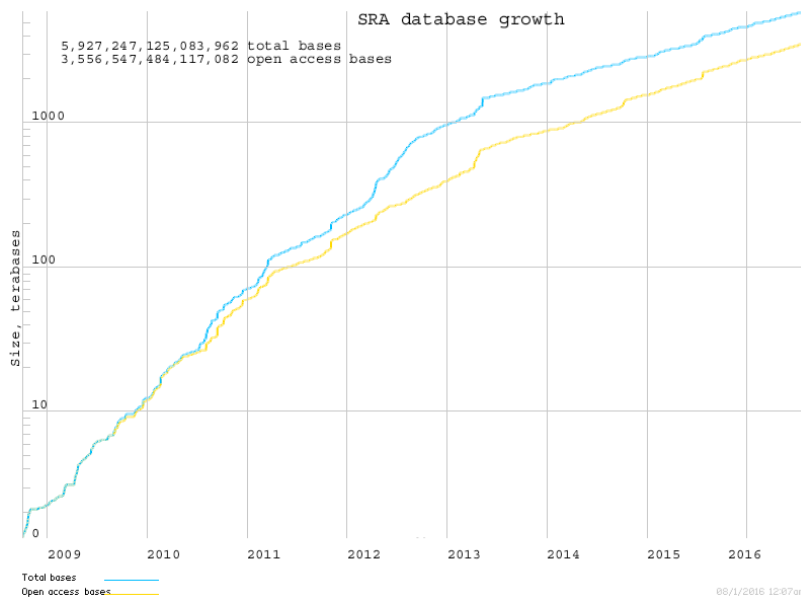
La première conclusion à souligner de ces évolutions technologiques est l'incroyable changement d'échelle entre la première génération de séquenceur Sanger et la deuxième génération de séquenceur à haut débit avec un séquençage de quelques kilobases (à l'échelle de l'exon) au milliard de bases (à l'échelle du génome entier). Ce bond technologique a ouvert la voie aux analyses globales sans *a priori* ainsi qu'à l'assemblage *de novo* de génomes. Il a également rendu possible l'analyse des variants structuraux sur la base des données de séquençage et a permis l'analyse à haut débit du transcriptome.

Ce bond technologie s'est accompagné d'une chute drastique des coûts de séquençage. De manière notable, les coûts de séquençage sur les quinze dernières années ont drastiquement chutés passant d'un coût de 3 milliards de dollars (Human Genome Project) à 1000 dollars pour un génome (*National Human Genome Research Institute*). Cette chute vertigineuse dépasse même les conjectures de Moore. Ces conjectures sont un ensemble de lois empiriques prévoyant que les composants électroniques doivent doubler leurs complexités chaque année pour rester compétitif en terme d'évolution technologique. Le dépassement de ces conjectures prouve l'énorme avancée réalisée dans le domaine du séquençage<sup>64</sup> (Figure 20).



**Figure 20** : Graphique illustrant le coût d'un génome humain en fonction du temps en comparaison avec la conjecture de Moore (<http://genome.gov/sequencingcosts>)

Ainsi, depuis 2008, les coûts de séquençage ne sont plus aussi prohibitifs et les séquenceurs nouvelle génération ont fait leur entrée dans un grand nombre de laboratoires et notamment dans les laboratoires hospitaliers. En 2015, le consortium ExAC (*Exome Aggregation Consortium*) estime ainsi que plus d'un demi-million de personnes dans le monde auraient bénéficié d'une analyse à haut débit <sup>65</sup>. De même, Schatz et Langmead rapporte, en 2013, que la capacité mondiale en terme de séquençage est de l'ordre du petabase ( $10^{15}$ ) par an <sup>66</sup>. Ces chiffres sont difficilement évaluables en 2017 étant donné que le nombre et le débit des séquenceurs n'a cessé d'augmenter (Figure 21).



**Figure 21** : Graphique illustrant la capacité de séquençage mondiale en térabases en fonction du temps

Données du SRA (*Sequence Read Archive*) à la date du 1 Aout 2016 (<http://www.ncbi.nlm.nih.gov/Traces/sra/>)

### III. Développement de technologies d'enrichissement

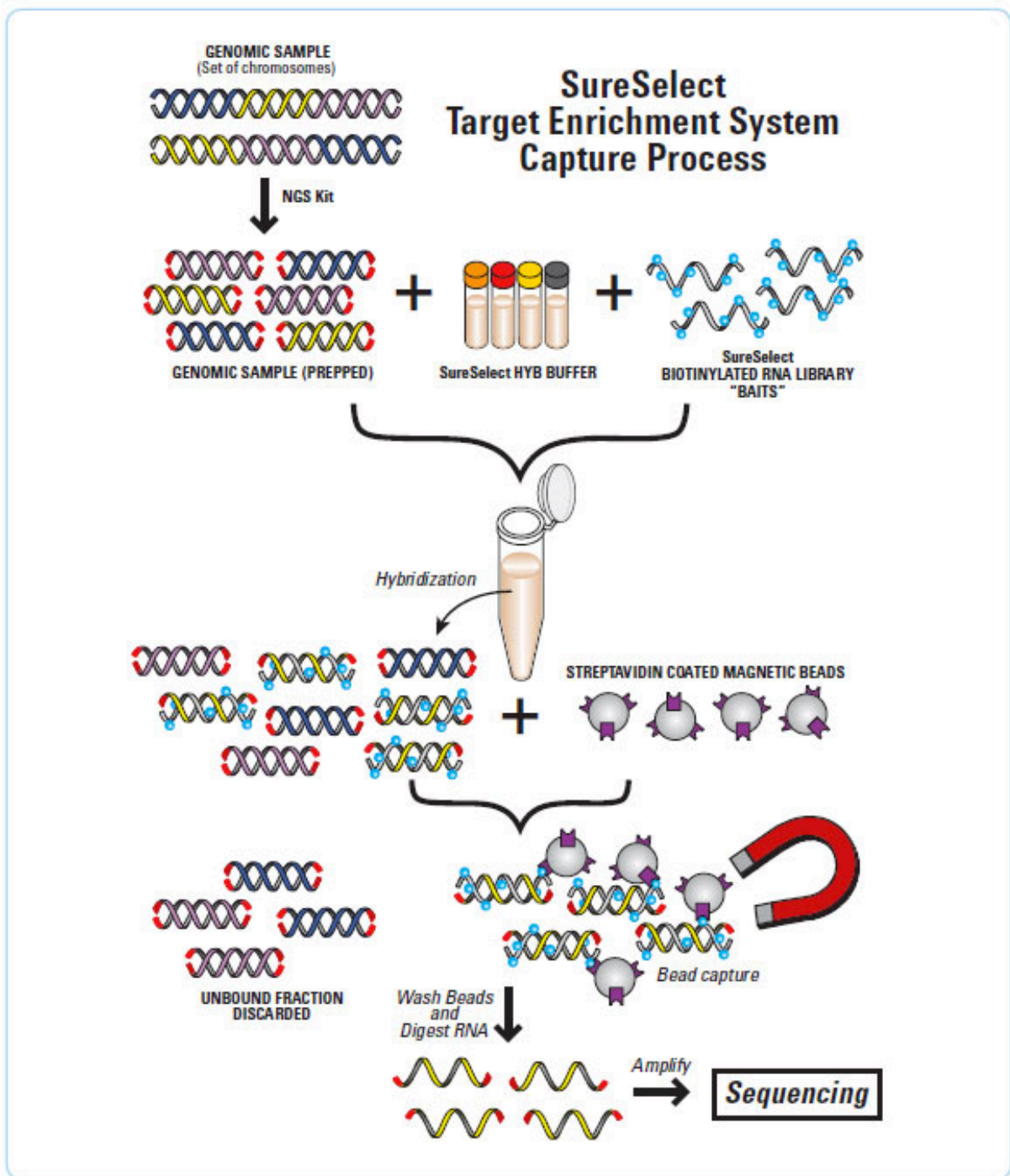
Aujourd'hui, il est donc techniquement possible de séquencer des génomes entiers. Pourtant, cela soulève des difficultés et notamment en Génétique Médicale : des coûts de séquençage encore élevés, des questions de profondeurs de lecture et donc de la détection des variants à bas bruit, l'acquisition de serveurs informatiques volumineux et puissants, et des problématiques de stockage. Outre, ces défis technologiques, **le principal défi demeure celui de l'interprétation des variations identifiées** (voir chapitre suivant).

Dans ce contexte, des nouvelles techniques d'enrichissement en régions d'intérêt ont fait leur apparition. Ainsi, étant donné que les séquenceurs ont des capacités fixes, en séquençant un nombre réduit de régions génomiques, il est possible d'obtenir une plus forte profondeur de lecture sur ces régions, ou de séquencer plus de patients au sein du même run de séquençage. Par exemple, il est possible de restreindre le séquençage à l'ensemble des régions codantes du génome (exome), à un panel de gènes d'intérêt, ou encore à des *hotspots* de mutations.

En fonction de la taille de la région à séquencer, il est possible d'employer des approches différentes. Il existe actuellement trois grandes approches : la capture par hybridation et la technique de PCR Amplicon majoritairement employées et la technique de sondes moléculaires inversées (MIP).

#### a. Enrichissement par hybridation

L'enrichissement par capture repose sur une hybridation des fragments d'intérêt à des sondes complémentaires en phase aqueuse<sup>67</sup>. L'ADN génomique est d'abord fragmenté de façon stochastique. Les fragments sont ensuite flanqués de part et d'autre de petites séquences nucléotidiques appelées adaptateurs, nécessaires à l'amplification clonale et au séquençage. Ensuite, l'ADN fragmenté est enrichi en séquences d'intérêt en utilisant un système de *pulldown* : les fragments d'ADN sont mis en contact avec une banque de sondes oligonucléotidiques (120-mers généralement) complémentaires des régions d'intérêt. Après une période d'hybridation, les complexes sondes / régions d'intérêt sont récupérés par l'utilisation de billes magnétiques recouvertes de streptavidine, molécule fortement affine à la biotine qui est complexée aux sondes. Les fragments d'intérêt sont ensuite purifiés et souvent pré-amplifiés pour former une librairie qui sera soumise à amplification clonale et séquençage (Figure 22).



**Figure 22** : Enrichissement en régions d'intérêt par capture.

Pour conduire à l'enrichissement en régions d'intérêt, l'ADN est d'abord fragmenté en courts fragments. Ensuite, un système de sondes ARN biotinylés complémentaires aux régions d'intérêts sont insérées dans le milieu. Après une incubation, les hybrides sondes ARN/fragments d'intérêt sont capturés à l'aide de billes magnétiques cotées à la streptavidine, fortement affine pour la biotine. Grâce à un aimant, seuls ces complexes ARN/fragment d'intérêt sont retenus lors de lavages. Les complexes sont ensuite détruits pour ne conserver que les fragments ADN d'intérêt en vue du séquençage.

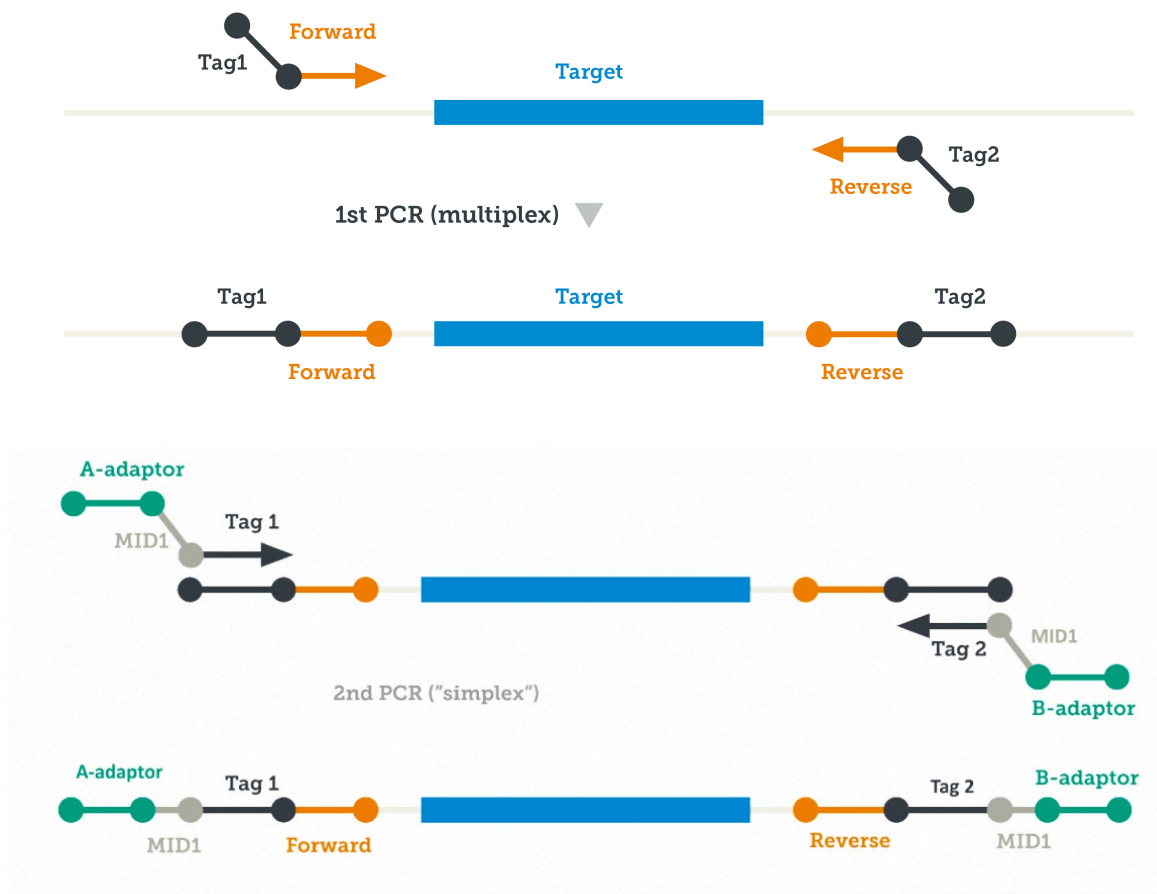
Au moins trois sociétés commercialisent actuellement ces kits de capture: Agilent, Illumina, Nimblegen. Bien qu'il y ait des différences techniques entre eux, Agilent utilise des sondes ARN et Illumina et Nimblegen des sondes ADN par exemple, les kits commercialisés varient peu en terme de performance et sont assez équivalents. La différence majeure réside dans le contenu des kits dédiés à la capture d'exome qui varient d'une société à l'autre et d'une version de kit à l'autre en fonction notamment de l'inclusion ou non des régions transcrites non traduites (UTR) ou des gènes des ARN non codants. A l'heure actuelle, l'enrichissement par capture reste la technique de choix pour réaliser du séquençage d'exome ou de panels de gènes.

### **b. Enrichissement par PCR amplicon**

Cette technologie repose sur l'amplification en parallèle par PCR multiplexe de nombreuses régions d'intérêt. Cette solution est principalement utilisée pour la sélection d'un nombre relativement faible de cibles. Basiquement, l'amplification débute par la fixation d'amorces encadrant les régions d'intérêt. Ces amorces contiennent également des séquences servant d'adaptateurs universels pour une amplification en parallèle de l'ensemble des fragments ainsi que d'adaptateurs de séquence (Figure 23).

Aujourd'hui, des sociétés comme ThermoFisher ont étendu cette approche à l'exome (ex : AmpliSeq Exome Kit). Damiani et ces collaborateurs ont rapporté en 2016 un bon fonctionnement de cette technologie pour le séquençage de l'exome avec 94 % des séquences codantes enrichies avec la technologie AmpliSeq.





**Figure 23 :** Technique d'enrichissement par PCR amplicon

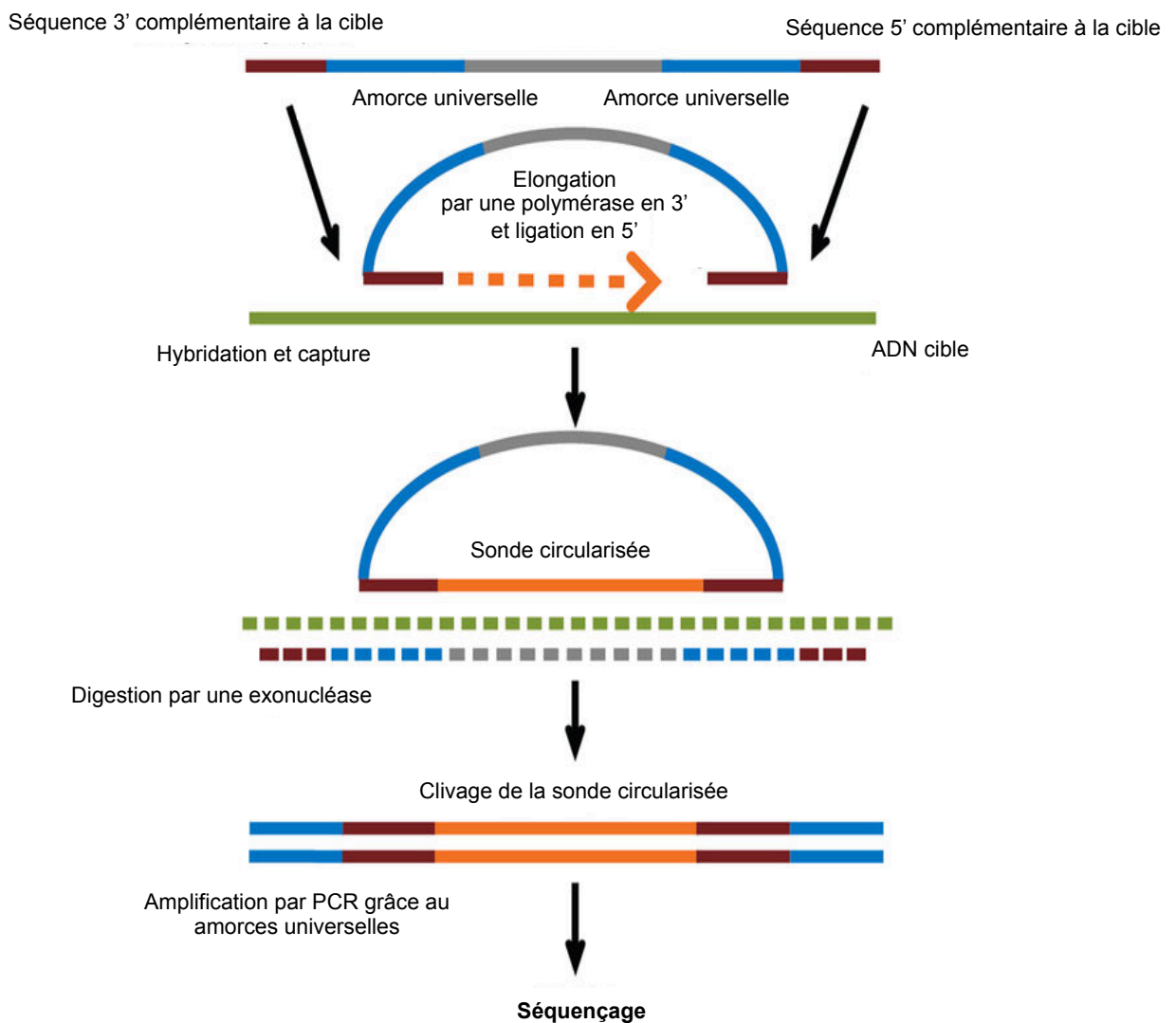
L'enrichissement par PCR amplicon est basé sur l'utilisation de couples d'amorces encadrant l'ensemble des régions d'intérêt. Grâce à une première étape de PCR multiplexe, les régions d'intérêt sont amplifiées. Pour finir, l'ensemble des amplicons sont amplifiés et flanqués d'adaptateurs, en parallèle grâce à la reconnaissance de tags présents sur les couples d'amorces initiaux. Source : <http://www.multiplicom.com/>

### c. Enrichissement par Molecular Inversion Probe (MIP)

La technique de "*molecular inversion probe*" (MIP) est basée sur l'utilisation de sondes ADN simples brins. Ces sondes possèdent plusieurs régions notables : (i) à chaque extrémités de la sonde présente une séquence complémentaire des extrémités 5' et 3' du fragment d'intérêt, (ii) une région centrale possédant un site de restriction commun à toutes les sondes et enfin (iii) deux régions adjacentes aux régions complémentaires pour réaliser une PCR universelle.

Grâce aux séquences complémentaires, la sonde vient s'hybrider aux régions d'intérêts. Le segment entre les deux sondes hybridées est ensuite synthétisé par une polymérase. La sonde ainsi que son néofragment est circularisée grâce à une ligase. Puis, le fragment circularisé est clivé par une enzyme de restriction au niveau du site central de

restriction permettant d'obtenir un fragment linéarisé possédant aux extrémités 5' et 3' des sites pour la réalisation d'une PCR universelle de l'ensemble des fragments d'intérêts <sup>68,69</sup> (Figure 24). Les avantages de cette technologie sont notamment (i) la spécificité de sélection de la région d'intérêt grâce aux deux sondes, (ii) la facilité de préparation de la librairie étant donné qu'il n'y a pas d'étape de clivage mécanique ou enzymatique de l'ADN, (iii) la possibilité d'ajouter des codes-barres, (iv) l'absence de matériel spécifique autre que du matériel classique de laboratoire.



**Figure 24** : Technique d'enrichissement par sondes moléculaires inversées (*Molecular Probes Inversion*)

L'ADN génomique est fragmenté (de façon enzymatique ou mécanique) puis enrichi grâce à une sonde constituée d'une séquence universelle flanquée aux extrémités de séquences spécifiques de la région d'intérêt. Après ligation et circularisation des fragments, l'ADN capturé est amplifié en utilisant la séquence universelle.

#### d. Comparaison de l'enrichissement par amplicon et par hybridation

En 2015, Samorodnitsky et ses collaborateurs ont comparé ces deux approches d'enrichissement à la fois sur séquenceur Illumina et sur séquenceur Ion Torrent. Les deux systèmes d'enrichissement par hybridation et par PCR amplicon ont montré un degré important d'enrichissement des régions d'intérêt permettant de focaliser le séquençage sur les régions souhaitées en augmentant donc leurs profondeurs de lecture à coût constant. Le point fort de la technologie amplicon est l'obtention de meilleurs *on-target* (taux de régions d'intérêt sur total des régions séquencées) alors que la capture par hybridation permet par contre une meilleure uniformité dans l'enrichissement. En terme de détection de variants, l'enrichissement par capture semble néanmoins plus optimal qu'il s'agisse de variants ponctuels ou de variations du nombre de copies <sup>70</sup>.

## **CHAPITRE III :**

### **Les enseignements et applications du Séquençage Nouvelle Génération**

## I. Les enseignements du NGS : les changements de paradigmes

Le séquençage à haut débit, dans un laps de temps très court, a amené la communauté à repenser de nombreux concepts. Tout ceci a été permis par la démocratisation des séquenceurs haut débit dans les laboratoires mais également grâce à la compilation massive des données issues de consortiums de séquençage. Bien que le consortium ExAC annonce qu'un demi-million de personnes dans le monde aurait bénéficié d'une analyse NGS <sup>65</sup>, ces derniers ne sont pas tous compilés ensemble sûrement pour des raisons éthiques, de logistique et de pratiques bio-informatiques non uniformes. Cependant, un effort international est constamment effectué pour collecter un grand nombre de génomes ou d'exomes, avec comme objectif de les rendre accessibles sous forme de bases de données. On peut noter l'existence aujourd'hui de trois grands projets accessibles publiquement :

- Le projet *NHLBI Grand Opportunity Exome Sequencing Project* (ESP) collectant 6 515 exomes incluant 4 298 exomes américains d'origine européenne et 2 217 exomes américains d'origine africaine.

- Le *1000 Genome Project* est un projet européen issu de l'EMBL (*European Molecular Biology Laboratory*) réunissant 2,504 génomes de 26 nationalités.

- Le projet ExAC (*Exome Aggregation Consortium*) est un consortium international recueillant les données de 60,706 exomes.

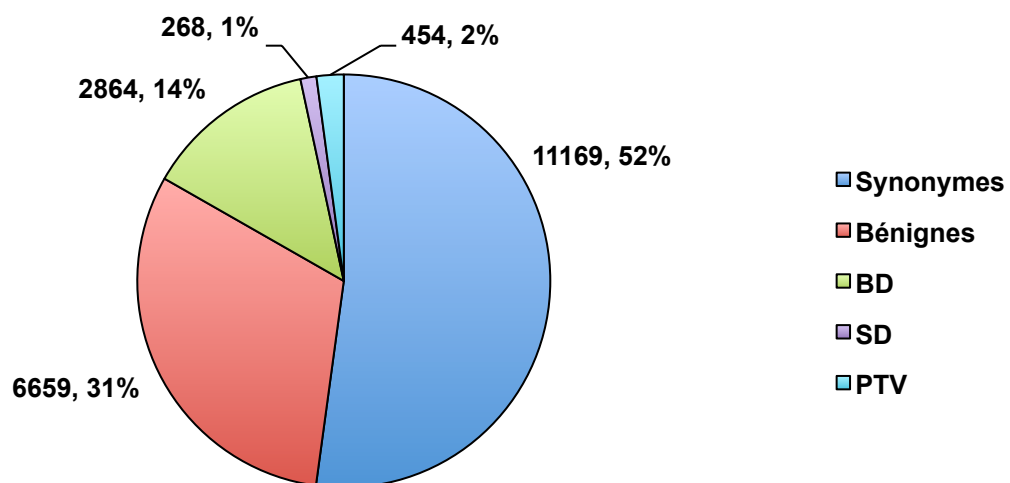
Il est également important de souligner l'existence de deux projets européens. Le premier dénommé *100K Genomes Project* vise à élucider les bases moléculaires du cancer et des maladies rares. Pour conduire cet objectif, 100 000 génomes d'origine anglaise vont être séquencés représentant 40 000 patients : un patient ayant développé un cancer aura deux génomes séquencés (sang + tumeur) et un patient ayant développé une maladie rare aura trois génomes séquencés (sang + le sang de deux personnes apparentées). Le second projet européen dénommé FREX (*FRench EXome project*) est un projet français visant à obtenir les exomes d'individus contrôles français (environ 600 exomes disponibles à ce jour).

Il est important de noter que l'ensemble de ces bases n'incluent pas uniquement des individus témoins (en bonne santé). Par exemple, l'objectif d'ESP est d'identifier de nouveaux mécanismes contribuant aux maladies cardiovasculaires, pulmonaires et hématologiques. Ainsi, les données ESP disponibles sont donc les individus contrôles pour cette cohorte et ne signifie pas qu'il n'y a pas d'individu ayant développé un cancer par

exemple. C'est pourquoi, il est nécessaire d'avoir une bonne connaissance de la constitution de ces bases de données lorsqu'elles sont utilisées comme filtre notamment.

#### a. La variabilité insoupçonnée du génome humain

Le premier enseignement du séquençage à haut débit est **l'importante variabilité insoupçonnée du génome humain**. En effet, il était bien admis qu'il existe des variations inter-individuelles comme les variations nucléotidiques polymorphique (*Single Nucleotide Polymorphism*; SNP) ou des variations du nombre de copies (CNV), mais il était totalement insoupçonné que chacun de nos génomes renferme plusieurs millions de variations nucléotidiques. Une fois éliminées les variations polymorphiques déjà répertoriées dans les bases de données, il demeure environ 500 variants rares dans un exome dont la fréquence allélique est inférieure à 0.1 %<sup>71,72</sup>. Certaines de ces variations sont privées et spécifiques à l'exome d'un individu. Chaque exome contient donc environ 500 variations susceptibles d'avoir un impact biologique<sup>73,74</sup>. En 2016, ces chiffres ont été affinés grâce au consortium ExAC et à la compilation de plus de 60 000 exomes. Selon ExAC, un individu porte en moyenne 21 451 variations exomiques (ce résultat peut varier en fonction du kit de capture utilisé) dont 1 500 avec une fréquence allélique inférieure à 1 %, environ la moitié étant des variations synonymes et l'autre moitié des variations non synonymes regroupant les variants tronquant la protéine et les variants faux-sens<sup>65</sup> (Figure 25).



**Figure 25** : Nombre moyen de variations génétiques par exome

Nombre de variants synonymes, faux-sens, bénins, *Broadly Damaging (BD)*, *Strictly Damaging (SD)*, et *Protein Truncating Variants (PTV)* retrouvées en moyenne par individu. Les variants SD sont les variations prédits pathogènes par les trois logiciels de prédiction de pathogénicité SIFT, Polyphen-2 et Mutation Taster. Les variants BD sont prédits pathogènes par au moins l'un des trois logiciels précédemment cités et les variants bénins sont prédits non pathogènes par les trois logiciels.

## b. Taux de variations *de novo*

L'analyse de l'exome a également permis la mise en évidence d'un **taux élevé de variations *de novo***. Ces derniers correspondent à des accidents génétiques survenus sur l'un des gamètes des parents de l'individu ou au niveau post-zygotique. Comme évoqué précédemment, le taux de variation *de novo* est estimé à  $1,8 \times 10^{-8}$  par base ce qui représente entre 0 à 2 variations *de novo* par exome. Ce taux reste similaire lorsqu'on s'intéresse au génome entier et représente alors 60 à 80 mutations *de novo* <sup>75,76</sup>

Il est important de noter que le taux de variations *de novo* chez un individu est fonction de l'âge de ses parents et particulièrement de l'âge paternel : chaque année ajouterait une à deux variations *de novo* génomique supplémentaire <sup>77-79</sup>. Ceci s'expliquerait par la différence physiologique homme/femme dans la production de gamètes et la défaillance croissante des systèmes de réparation avec l'âge. Aujourd'hui, il est donc impossible de ne pas tenir compte de ce taux de variabilité trans-générationnelle dans le diagnostic et la prise en charge des familles atteintes de maladies mendéliennes quelles qu'elles soient.

## c. Notions de singletons et de variants rares

Les notions de singletons et de variants rares sont des notions assez récentes, qui ont émergé notamment avec les travaux du consortium ExAC. Cette collection importante d'exomes a permis de découvrir que les variants rares et peu fréquents sont majoritaires dans l'exome, et qu'une grande majorité d'entre eux auraient un impact fonctionnel. Ainsi, il est estimé qu'environ 50 % des allèles identifiés dans ExAC sont des *singletons* c'est à dire reportés chez un seul individu et qu'en plus, 99 % des allèles présents dans ExAC ont une fréquence inférieure à 1 %. De plus, 72 % des variants identifiés par ExAC étaient inconnus des les projets 1000 Genomes et ESP.

Cette notion avait été déjà suggérée par les données issues des premières versions du projet ESP rapportant le séquençage exomique de 2439 individus <sup>73</sup>. En effet, parmi les 503 481 variants identifiés, une très grande majorité (82 %) étaient préalablement inconnus, et 98 % d'entre eux étaient rares (définis ici avec une fréquence allélique inférieure à 0,5 %). Le nombre moyen de variations nucléotidiques par exome étaient alors estimé à environ 13 000 et chaque individu possédait environ 580 variations prédites délétères.

Traditionnellement, un variant était considéré comme polymorphique lorsque sa fréquence allélique dépassait 1 %. Ainsi, ces nouvelles informations amènent donc à concevoir une notion différente du polymorphisme et des variants rares; et donc à repenser directement l'interprétation clinique de ces variants rares observés chez les patients. On comprend bien que la rareté d'un variant n'est pas synonyme de causalité. Le risque d'identifier un variant rare, voire privé dans les études de séquençage à haut débit est une notion considérable et doit donc soulever une certaine prudence dans les analyses. La filtration des variants candidats par la fréquence et leur effet prédit est sûrement une nouvelle étape clef dans la découverte des variants causaux à l'origine de ces maladies.

#### **d. La tolérance aux variants perte de fonction**

Un autre enseignement important du séquençage à haut-débit est le fait que certains gènes sont tolérants à la perte de fonction. Il est d'ailleurs possible de les regrouper sous l'appellation "gènes tolérants à la perte de fonction" ou "non tolérants à l'haploinsuffisance". Un gène tolérant à la perte de fonction est un gène possédant un nombre important de variants perte de fonction. Les variants perte de fonction (*Protein Truncating Variant*; PTV) peuvent se définir comme un ensemble regroupant notamment les variants nucléotidiques non-sens, les variants détruisant un site canonique d'épissage, les courtes insertions et délétions décalant le cadre de lecture. Dans les données ESP, chaque individu possédait environ 35 variants non-sens et était homozygote pour au moins l'un d'eux<sup>73</sup>. De même, sur les données de 185 génomes issues du *1000 Genome Project*, une centaine de variants perte de fonction et vingt gènes inactivés de façon biallélique ont été identifiés par individu<sup>80</sup>. Dans les données ExAC, parmi les 7 milliards de variants répertoriés, presque 200 000 correspondent à des variants PTV (dont 120 000 sont des singletons). Les variants PTV sont les variants les plus rares (2,4 % de l'ensemble de variants détectés dans la cohorte ExAC). En moyenne, chaque individu possède dans la séquence codante de son génome 85 PTV à l'état hétérozygotes et 35 à l'état homozygotes. Cependant, la grande majorité des PTV présents chez un individu sont communs (MAF > 1 %) réduisant ainsi le nombre de PTV rares à 18 (MAF < 1 %) <sup>65</sup>. Le consortium a recensé ces gènes et a défini une liste de 3230 gènes non tolérants à l'haploinsuffisance, dont 72 % n'ont pas de lien (encore) établi avec un maladie Mendélienne (OMIM, ClinVar).

Ces données suggèrent donc que pour un nombre important de gènes, la perte de fonction n'entraîne pas de pression de sélection. Il existe sûrement une compensation par des gènes possédant une fonction analogue comme des gènes de la même famille. Cette redondance biologique supposée est appuyée par le fait que les gènes inactivés



bialléliquement appartiennent à des familles de gènes plus grandes<sup>80</sup>. Le grand nombre de variants perte de fonction identifiables dans un exome amène donc à une grande prudence vis à vis de l'interprétation des données pour un patient. Il convient probablement de considérer les gènes soumis à l'haploinsuffisance à part dans un premier temps, étant donné leurs faibles contributions *a priori* au phénotype du patient. De même, si un allèle perte de fonction est rare, il est important d'évaluer la charge globale en variants perte de fonction du gène entier chez des individus contrôles.

#### **e. La filtration des variants**

Compte-tenu des éléments décrits ci-dessus, la stratégie de filtration des variants détectés est désormais cruciale et doit intégrer les nouvelles connaissances concernant la grande variabilité du génome humain. Afin de déterminer le variant causal à l'origine de la pathologie observé chez un patient donné à partir des données du séquençage de son exome, il est nécessaire de filtrer progressivement ses 20 000 variants exoniques en tenant compte du mode de transmission supposé. Généralement, les premiers filtres employés concernent la qualité du variant, sa nature, sa fréquence en population générale, la prédiction *in silico* de son impact sur la protéine ou sur l'épissage. Ceci est généralement complété par l'interrogation de bases de données spécifiques de gènes ou de pathologies.

#### **Qualité des variants :**

La première étape de filtration est généralement basée sur la qualité des variants et permet notamment de s'affranchir des erreurs de séquençage. Cette étape prend généralement en compte le score de qualité du variant, le nombre de lectures portant le variant, la répartition égale de ces lectures en sens et antisens et la proportion de reads mutants par rapport au nombre total de lectures (balance allélique).

#### **Filtration selon la fréquence allélique :**

La fréquence allélique d'un variant en population générale (*Major Allele Frequency*; MAF) est un critère déterminant dans le cadre des maladies Mendéliennes. Le seuil à utiliser pour la filtration des variants va dépendre de la fréquence de la pathologie considérée et du mode de transmission envisagé. A la lumière des connaissances apportées par le séquençage de masse de populations générales, il est essentiel de comparer la fréquence des variants détectés chez un individu à la fréquence de ces variants chez une population contrôle de même origine ethnique et géographique. De plus, la technologie utilisée pour l'enrichissement et le séquençage peut également introduire des biais (génomomes vs exomes vs panels). Les projets ne disposant pas en général de contrôles appariés, il est fréquent

d'utiliser les bases de données publiques rassemblant les variants détectés dans 2500 génomes séquencés (1000 GENOMES), dans 6 500 exomes séquencés (*Exome Sequencing Project*, ESP) et plus récemment dans 60 000 exomes séquencés (Projet ExAC). A l'intérieur de ses bases de données, il est possible de ne considérer que la sous-population la plus proche géographiquement des individus étudiés. Il est également possible de ne considérer pour la filtration que la fréquence maximale observée pour ces sous-populations.

#### **Effet prédit des variants faux-sens :**

Devant le grand nombre de variants faux-sens détectés par exome, de nombreux logiciels de prédiction de pathogénicité ont été développés afin de faciliter leur interprétation. Ces outils sont en général basés sur des algorithmes prenant en compte la conservation au cours de l'évolution des acides aminés et la distance physico-chimique entre les deux acides aminés impliqués. A l'issue de cette analyse, les variants sont classés selon plusieurs classes allant de *damaging/deleterious/disease causing* à *benign/neutral*. À noter qu'il est sûrement préférable d'utiliser une terminologie basée sur l'impact fonctionnel possible sur la protéine et non pas sur l'interprétation de la causalité du variant (*disease causing*). Les logiciels les plus couramment employés sont Polyphen2, SIFT, et Mutation Taster<sup>81</sup>. Aucun de ces logiciels n'étant parfait, il est commun de combiner plusieurs de ces logiciels et de ne considérer que les variants prédits délétères par plus d'un logiciel.

#### **Bases de données spécifiques de gènes ou de pathologies :**

Une étape importante dans la filtration/interprétation des variants est la recherche de leur présence dans des bases de données. Il existe deux types de bases à savoir celles non spécifiques d'un gène à l'instar de ClinVar, OMIM, HGMD et celles spécifiques d'un gène comme LOVD ou les bases UMD. Le recours à de telles bases peut dans certains cas faciliter l'interprétation, mais nécessite parfois une certaine prudence car certaines de ces bases sont régulièrement employées pour recenser les variants détectés sans que le biologiste ait approfondi son interprétation.

La filtration des variants est une étape essentielle avant l'interprétation des variations candidates mais est surtout une étape discriminante dans un projet, véritable balance entre un grand nombre de variants à interpréter et le risque d'éliminer la variation causale. Il convient donc d'adapter ces étapes en fonction du projet, du mode de transmission supposé et de la fréquence de la maladie. Cette étape permet donc d'obtenir une liste réduite de variants candidats. Cependant, dans le cas d'une analyse d'exome, ce nombre devra encore être réduit par l'utilisation de stratégies d'analyse décrites ci-après (partie III de ce chapitre).

Après identification de variants possiblement causaux, il conviendra de vérifier leur pathogénicité par les moyens classiquement utilisés : études de co-ségrégation, études de récurrence, tests fonctionnels ou modèles animaux.

## II. Application du NGS : le séquençage de panels de gènes

Aujourd'hui, la promesse qu'un génome humain peut être séquençé pour 1000 dollars et en quelques jours est bien réelle. Cependant, malgré tout, la compréhension du génome humain n'est pas totale. Cette limitation a directement conduit à l'essor de nouvelles techniques d'enrichissement (Chapitre II-III) permettant de recentrer l'analyse sur des régions d'intérêt et de limiter les coûts de séquençage en augmentant le nombre de patients par *run*.

Ainsi, au cours des dernières années, le paysage du diagnostic des maladies Mendéliennes et notamment des formes héréditaires aux cancers, s'est largement métamorphosé avec une introduction massive des séquenceurs nouvelle génération de plus en plus miniaturisés dans les laboratoires, et l'emploi des nouvelles techniques d'enrichissement en régions d'intérêt.

Grâce aux techniques d'enrichissement en régions génomiques d'intérêt préalablement décrites, il est maintenant envisageable de séquençer un grand nombre de gènes en parallèle pour faciliter le diagnostic des maladies Mendéliennes. Ces tests génétiques, appelés panel de gènes, ciblent généralement un nombre limité de gènes (de 2 à 200) et sont souvent employés dans des stratégies pour lesquelles les résultats d'un test préliminaire aurait conditionné le suivant (séquençage du gène A puis si retour négatif, séquençage du gène B).

### a. Intérêts du séquençage de panel de gènes

Le premier intérêt est bien sûr la possibilité de cribler plusieurs régions en une seule réaction. Cette approche offre donc directement la possibilité de gagner **du temps en terme de rendu diagnostic** dans les pathologies impliquant plusieurs gènes tout en étant moins **coûteuse en matériel génétique** bien souvent précieux. Dans le contexte des prédispositions héréditaires aux cancers, cela est particulièrement intéressant étant donné les problématiques évoquées précédemment comme les chevauchements phénotypiques.

Le deuxième intérêt est l'opportunité de séquençer **plus de régions génomiques** que par des techniques conventionnelles précédentes (séquençage Sanger) en adjoignant par exemple des régions introniques et régulatrices. Ceci offre donc une possibilité supplémentaire d'identifier un événement moléculaire à l'origine de la pathologie observée.

Etant donné qu'il est possible de restreindre le séquençage non plus sur le génome entier mais sur un nombre limité de régions, il est possible d'analyser en parallèle un

**nombre plus important de patients au sein d'un même run de séquençage.** Cette propriété permet donc une augmentation du débit à coût de séquençage fixe, permettent de ce fait, le raccourcissement des délais des rendus diagnostiques.

Le séquençage de panels de gènes offre également l'avantage **d'être flexible dans son design.** Bien que certaines sociétés comme GeneDX, Myriad ou AmbryGenetics proposent des panels dédiés, il est possible de réaliser **un panel de gènes à façon**, et donc d'inclure rapidement de nouveaux gènes. Cet intérêt est particulièrement relevant dans le contexte des prédispositions héréditaires aux cancers de par l'évolution rapide des connaissances. Cependant, il est intéressant de noter que devant l'absence de référence internationale, la sélection des gènes à inclure dans un panel est à la charge de chaque laboratoire. Généralement, le socle minimal commun aux panels de gènes est composé de gènes dont les variations sont fortement pénétrantes pour la pathologie d'intérêt. De plus, grâce à cette souplesse de conception, **des gènes à visée recherche** peuvent être également inclus en parallèle dans un panel diagnostique afin d'évaluer leurs possibles contributions.

En conclusion, le séquençage de panels de gènes représente une alternative au séquençage Sanger, majoritairement par leur efficacité en terme de temps et de coûts <sup>82</sup>. Ainsi, pour l'ensemble des raisons évoquées, le séquençage de panels de gènes occupe aujourd'hui une place de plus en plus prépondérante dans le diagnostic des maladies héréditaires et notamment dans les prédispositions héréditaires aux cancers.

#### **b. L'utilisation des panels de gènes dans le contexte du diagnostic des prédispositions héréditaires aux cancers**

Historiquement, les premiers panels de gènes disponibles étaient des panels consacrés à l'étude des cardiomyopathies <sup>83</sup>. Ce n'est que deux ans après (2010) que les premières études précliniques ont exploré la performance et la robustesse de ces approches dans le contexte des prédispositions héréditaires aux cancers comparé au séquençage Sanger. Trois études pionnières font date.

La première conduite par Hoppman-Chaney et ses collègues, était focalisée sur cinq patients Lynch présentant une mutation MMR connue. Grâce à une capture de 22 gènes (incluant *MLH1*, *MSH2*, *MSH6* et *APC*) et en utilisant deux plate-formes de séquençage (Roche 454 et Illumina) en parallèle, les auteurs avaient alors mis en évidence qu'un certain nombre de variants (non pathogènes) précédemment identifiés avec des méthodes standard

avaient été manqués par ces nouvelles méthodes. De plus, selon la plate-forme utilisée, la sensibilité variait entre 90 % et 93 %, principalement à cause de problématiques d'enrichissement sous-optimaux des régions riches en GC <sup>84</sup>.

Parallèlement, dans la même année, une seconde étude a été conduite par Walsh et ses collègues de l'Université de Washington. Leur panel dénommé BROCA, en l'honneur du docteur français Paul Broca, l'un des premiers pathologistes à décrire une agrégation familiale de cancers du sein et de l'ovaire, est justement destiné à l'étude des prédispositions héréditaires aux cancers du sein et de l'ovaire. Ce panel était composé originellement de 21 gènes et initialement testé sur 20 femmes avec des mutations constitutionnelles connues. Le protocole employé s'était révélé particulièrement sensible à la détection de tous les types d'altérations (substitutions d'une seule base, petites insertions et délétions, et réarrangements de grandes tailles) sans aucun faux positif. Ce même protocole avait été ensuite appliqué à 360 femmes ayant développé un carcinome ovarien indépendamment de leurs histoires familiales. Le panel BROCA avait permis alors la mise en évidence de la présence d'une mutation perte de fonction dans l'un des gènes du panel chez 23 % de ces femmes (dont 11 % dans le gène *BRCA1* et 6 % dans le gène *BRCA2*). Une perte d'hétérozygotie dans la tumeur avait été ensuite documentée pour 80 % de ces patientes <sup>85</sup>.

Enfin, une approche similaire a été utilisée dans le contexte du syndrome de Lynch avec un panel de gène appelé ColoSeq. Le séquençage du panel ColoSeq avait été originellement réalisé chez 23 patients atteints d'un syndrome de Lynch, 31 patients évocateurs d'un syndrome de Lynch et 19 échantillons de contrôles indemnes de cancer. Ce panel avait alors permis la détection d'altérations préalablement connues chez les patients Lynch ainsi que six nouvelles mutations et trois variants de signification inconnue dans la cohorte prospective de 31 patients évocateurs d'un syndrome de Lynch. Au delà de la preuve de concept, l'originalité de cette étude pré-clinique avait été de séquencer deux fois 75 échantillons et de démontrer ainsi la reproductibilité technique du séquençage du panel ColoSeq <sup>86</sup>.

Ces premières études princeps ont donc apporté à la fois la preuve technique et conceptuelle de l'intérêt du séquençage à haut-débit de panels de gènes dans le contexte des prédispositions héréditaires aux cancers. Ces études démontrent une robustesse, une reproductibilité du séquençage de panels de gènes ainsi qu'une capacité à mettre en évidence des variants que les méthodes traditionnelles n'avaient pas pu identifier, confirmant une utilité clinique importante.

Depuis, plusieurs comparaisons techniques entre les panels de gènes et le séquençage Sanger ont été entreprises dans de nombreuses études et démontrent une concordance souvent parfaite <sup>75-78</sup> y compris sur d'importantes cohortes (n>1000 individus).

La comparaison de kits d'enrichissements différents ciblant les mêmes gènes a également été évaluée par deux équipes allemandes grâce un échange mutuel de 12 échantillons. Les deux laboratoires ont obtenu une excellente concordance (99,5 %) : seul un variant non pathogène a été manqué dû à une région mal capturée <sup>91</sup>. Cette étude a permis de démontrer qu'à la fois la préparation de l'échantillon et les kits employés influençaient peu les résultats obtenus par séquençage par panel de gènes.

Ainsi depuis 2012, ces méthodes ont fait leurs entrées massives dans les laboratoires publics et privés. En effet, les tests génétiques par panel de gènes en diagnostic offrent une approche peu coûteuse, fiable et rapide. Une fois développée, ces approches appliquées en routine diagnostique permettent un temps de rendu diagnostique plus court comparé aux approches gène à gène traditionnelles. La grande capacité à séquencer beaucoup d'individus offerte par ces techniques est sans précédent. La société GeneDX a publié par exemple les résultats de l'analyse moléculaire de plus de 10 000 patients atteints de cancers héréditaires du sein, de l'ovaire, du pancréas, du colon, du rectum et de l'endomètre <sup>92</sup>. Cette étude a permis de mettre en évidence une altération constitutionnelle à l'origine de la pathologie dans environ 9 % des patients (9,7 % pour le cancer du sein, 13,4 % dans le cancer de l'ovaire, 14,8 % pour les cancers colorectaux). Très récemment, une autre étude a criblé 25 gènes chez plus de 35 000 femmes ayant développé un cancer du sein avant l'âge de 45 ans ou possédant une histoire familiale de cancer du sein et/ou de l'ovaire. Les auteurs de cette étude ont mis en évidence l'identification d'une mutation constitutionnelle dans l'un des 25 gènes dans 9,3 % des femmes criblées <sup>93</sup>.

Les processus techniques continuent sans cesse de s'améliorer qu'il s'agisse de nombre de gènes testés, de technologies d'enrichissement employées, de séquenceurs et des *pipelines* bioinformatiques utilisés. Tout cela contribue à faire de ces nouvelles méthodes des outils puissants dans le diagnostic des pathologies héréditaires et notamment dans les prédispositions héréditaires aux cancers.

### **c. Applications du séquençage de panels de gènes : paradigme des prédispositions héréditaires aux cancers du sein et de l'ovaire**

Il serait impossible de récapituler ici l'ensemble des panels existants, ainsi la suite de cet exposé sera focalisée sur les panels relatifs aux prédispositions héréditaires aux cancers du sein et de l'ovaire (HBOC). Dans cet ensemble, il existe une multitude de panels consacrés à l'étude de ces prédispositions, principalement dû au nombre important de gènes impliqués dans cette pathologie et aux pénétrances des variations associées à ces gènes.

En règle générale, les panels consacrés au HBOC comportent de 12 à 45 gènes avec l'existence d'un socle commun de gènes. Ces gènes sont les gènes dont les variations sont considérées comme fortement pénétrantes : *BRCA1*, *BRCA2*, *PALB2*, *TP53*, *PTEN*, *CDH1*, *STK11* (Tableau 4)<sup>94</sup>. Les autres gènes inclus dans les panels de gènes consacrés au HBOC sont généralement des gènes dont les variations confèrent un risque modéré de développer un cancer du sein et/ou de l'ovaire (les paralogues *RAD51*, *CHEK2*, *ATM*, *MEN1*, *BLM*, *BRIP1*, *NBN*, *BARD1*, *MRE11A*, *MUTYH*, *XRCC2 et 3*, *RINT1*, *RECQL*,...). Etant donné que leurs contributions relatives respectives ne sont pas certaines, chaque laboratoire de diagnostic est libre de les inclure ou non dans son panel HBOC.

Ainsi, la problématique d'un panel de gènes résulte donc dans le fait de trouver un socle minimal de gènes permettant d'identifier un maximum d'altérations, tout en criblant un maximum de patients. Plus le nombre de gènes sera important, plus le nombre de patients pouvant être analysés au sein d'un même run de séquençage sera réduit.

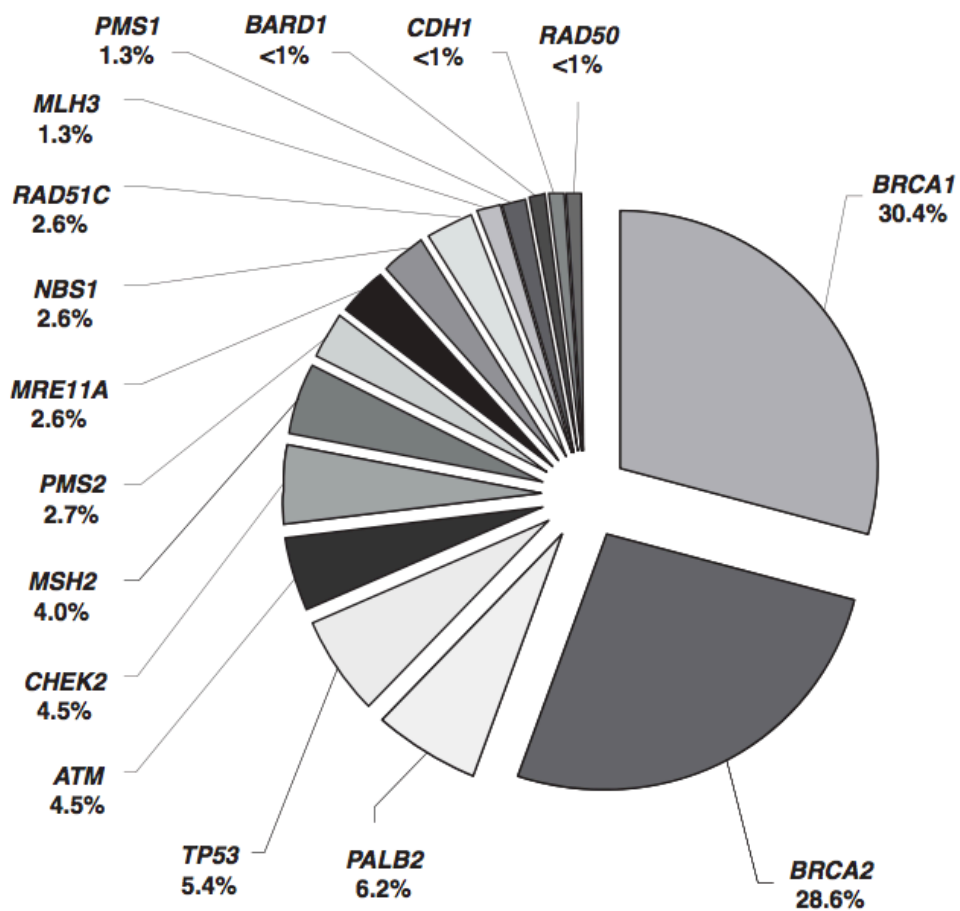
#### Panel de gènes consacré à l'étude des gènes de prédisposition héréditaire aux cancers du sein et de l'ovaire : l'exemple caennais.

A l'instar de l'équipe de MC King et du panel BROCA<sup>85</sup>, l'équipe caennaise de notre unité pilotée par D.Vaur et L.Castéra a développé, en 2014, un panel de gènes diagnostique pour l'analyse des gènes *BRCA1* et *BRCA2* chez des patientes évocatrices d'une prédisposition héréditaire au cancer du sein et/ou de l'ovaire avec l'objectif de remplacer les techniques conventionnelles et *in fine* réduire le délai du rendu diagnostique<sup>87</sup>. Dans ce contexte, l'équipe a développé un panel incluant les gènes *BRCA*. Ce panel a été utilisé de façon prospective chez 708 patientes évocatrices permettant l'identification de 69 variations dans les gènes *BRCA1/BRCA2*, quatre dans le gène *TP53* (environ 1 % des patientes criblées) et 36 variants dans les autres gènes du panel (Figure 26). Cette étude a donc permis de démontrer le transfert du diagnostic conventionnel vers des techniques haut-débit avec une sensibilité optimale répondant aux exigences diagnostiques existantes.



Company	Test	Website	Genes Included†
Ambry Genetics	BreastNext	<a href="http://www.ambrygen.com/tests/breastnext">www.ambrygen.com/tests/breastnext</a>	ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, MRE11A, MUTYH, NBN, NF1, PALB2, PTEN, RAD50, RAD51C, RAD51D, TP53
BreastHealth UK	BreastGene	<a href="http://www.breasthealthuk.com/screening-services/genetic-testing/breastgene">www.breasthealthuk.com/screening-services/genetic-testing/breastgene</a>	ATM, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, PALB2, PTEN, STK11, TP53
Centogene	Breast Ovarian Cancer Panel	<a href="http://www.centogene.com/centogene/centogene-test-catalogue.php">www.centogene.com/centogene/centogene-test-catalogue.php</a>	ATM, BARD1, BRIP1, CDH1, CHEK2, MEN1, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PMS1, PMS2, RAD50, RAD51C, RAD51D, XRCC2
Emory Genetics Laboratory	High Risk Breast Cancer Panel	<a href="http://geneticslab.emory.edu/tests/MM201">http://geneticslab.emory.edu/tests/MM201</a>	PTEN, STK11, TP53
Fulgent Diagnostics	Breast Ovarian Cancer NGS Panel	<a href="http://fulgentdiagnostics.com/test/breast-ovarian-cancer-ngs-panel/">http://fulgentdiagnostics.com/test/breast-ovarian-cancer-ngs-panel/</a>	APC, ATM, ATR, AXIN2, BAP1, BARD1, BLM, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, CTNNB1, EPCAM, FANCC, HOXB13, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PALLD, PMS2, PTEN, RAD50, RAD51, RAD51C, RAD51D, SMAD4, STK11, TP53, VHL, XRCC2, XRCC3
GeneDx	OncoGeneDx	<a href="http://www.genedx.com/test-catalog/available-tests/breastovarian-cancer-panel">www.genedx.com/test-catalog/available-tests/breastovarian-cancer-panel</a>	ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, EPCAM, FANCC, MLH1, MSH2, MSH6, NBN, PALB2, PMS2, PTEN, RAD51C, RAD51D, STK11, TP53, XRCC2
Illumina	TruSight Cancer	<a href="http://www.illumina.com/clinical/translational_genomics/panels/kits.html">www.illumina.com/clinical/translational_genomics/panels/kits.html</a>	94 Genes plus 287 SNPs reported to be associated with risk of breast cancer
Invitae	Hereditary Breast Cancer, High-Risk Panel	<a href="http://www.invitae.com/en/physician/panel-detail/PNL0009/">www.invitae.com/en/physician/panel-detail/PNL0009/</a>	BRCA1, BRCA2, CDH1, PALB2, PTEN, STK11, TP53
Myriad Genetics†	myRisk	<a href="http://www.myriad.com/products-services/hereditary-cancers/myrisk-hereditary-cancer/">www.myriad.com/products-services/hereditary-cancers/myrisk-hereditary-cancer/</a>	ATM, BARD1, BRCA2, BRIP1, CDH1, CHEK2, NBN, PALB2, PTEN, RAD51C, STK11, TP53
CD Genomics	Genetic Testing for the Cancer Susceptibility	<a href="http://www.cd-genomics.com/Genetic-Testing-for-the-Cancer-Susceptibility.html">www.cd-genomics.com/Genetic-Testing-for-the-Cancer-Susceptibility.html</a>	Not specified
University of Washington†	BROCA – Cancer Risk Panel	<a href="http://web.labmed.washington.edu/tests/genetics/BROCA">http://web.labmed.washington.edu/tests/genetics/BROCA</a>	AKT1, ATM, BARD1, BRCA1, BRCA2, BRIP1, CDH1, CHEK2, EPCAM, FAM175A, GEN1, MRE11A, MUTYH, NBN, PALB2, PIK3CA, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53, XRCC2

**Tableau 4** : Les différents panels de gènes disponibles consacrés à l'étude des prédispositions héréditaires au cancer du sein et de l'ovaire. Issu de Easton et al.,<sup>94</sup>



**Figure 26** : Distribution relative des variants détectés par séquençage de panel de gènes chez 708 patients présentant un cancer du sein et/ou de l'ovaire dans la cohorte Caennaise<sup>87</sup>.

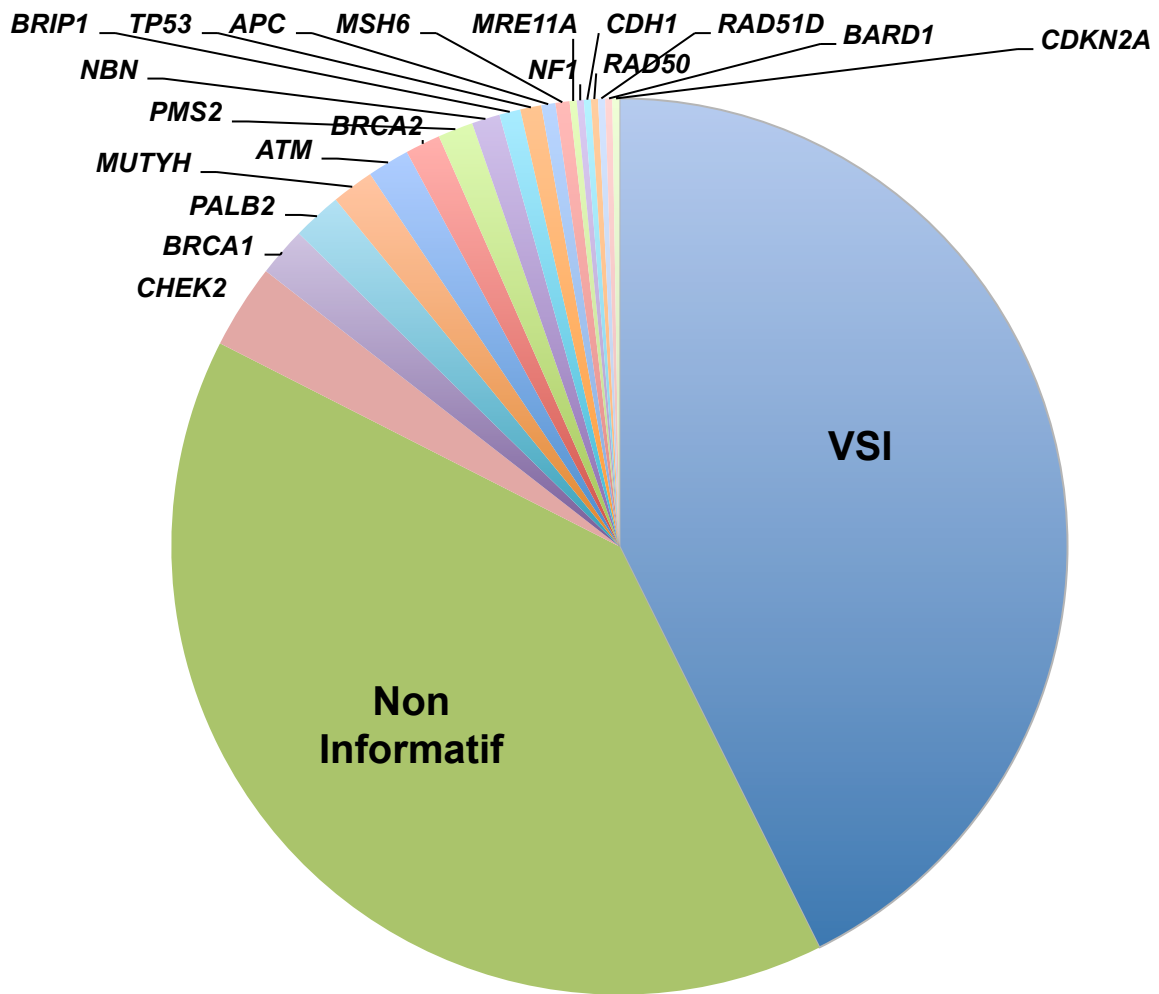
#### d. Les variations de signification biologique inconnue

En règle générale, les panels de gènes permettent l'identification d'une altération pathogène chez 8 à 15 % des patients analysés <sup>95</sup>. Malgré cela, de nombreuses variations dont la pathogénicité n'est pas encore déterminée, sont identifiées (Tableau 5). En 2015, Slavin et ses collaborateurs estiment à 42 % le nombre de variants de signification biologique inconnue (VSI) dans leurs études <sup>96</sup> (Figure 27). Ce taux de VSI a encore été confirmé récemment <sup>97</sup>. La présence de VSI soulève de nombreux problèmes et pour le conseil génétique des patients et de leurs familles. La relation causale entre la maladie et le variant identifié n'étant pas claire, il devient donc impossible de proposer des traitements préventifs comme la mastectomie ou l'annexectomie prophylactique, comme cela peut être proposé aux porteuses d'altérations pathogènes des gènes *BRCA*.

Référence	Patients inclus	Nombre de Gènes	Mutation pathogénique hors BRCA (%)	VSI (%)
Kurian et al.	198	42	11	88
Maxwell et al.	278	22	11	19
Lincoln et al.	1062	29	4,5	41
Tung et al.	488	25	5	33

**Tableau 5** : Liste non exhaustive des études rapportant le pourcentage de mutations pathogènes et de variants de signification biologique inconnue identifiés par séquençage à haut-débit de panels de gènes. <sup>82,98-100</sup>

En conclusion, la possibilité de séquencer à haut-débit un grand nombre de régions d'intérêt grâce à l'enrichissement est à l'origine d'un saut qualitatif et quantitatif dans le domaine du diagnostic des maladies Mendéliennes et également dans le domaine de la recherche fondamentale grâce notamment à l'inclusion de gènes à visée recherche dans les panels diagnostiques mais il demeure une difficulté d'interprétation avec la présence d'un grand nombre de VSI.



**Figure 27** : Représentation des variants identifiés par l'équipe de Salvin par séquençage à haut-débit d'un panel de gènes consacré à l'étude des prédispositions héréditaires au cancer du sein et de l'ovaire. Non informatif : Absence de variant pathogène ou de signification biologique inconnue. VSI : Variation de signification biologique inconnue

### III. Le séquençage exomique et ses applications

Le perfectionnement des techniques d'enrichissement permet aujourd'hui d'enrichir un grand nombre de régions d'intérêt; à tel point qu'il est aujourd'hui possible de cibler l'exome entier, c'est à dire l'ensemble des 180 000 régions codantes du génome humain. Généralement, la technique de choix pour réaliser l'enrichissement de l'exome est l'hybridation par capture. Le séquençage à haut débit de ces régions est dénommé WES (*Whole Exome Sequencing*).

#### a. Intérêts de l'exome

De façon analogue aux séquençages de panels de gènes, l'avantage premier de l'exome est sa taille réduite comparée à la taille d'un génome entier (34 Mb, 1,5 % du génome). Cette commodité facilite l'augmentation du nombre d'individus criblés par *run* de séquençage tout en maintenant des profondeurs de lecture élevée. Ainsi, en terme de matériel génétique utilisé, de coût et de stockage, le séquençage de l'exome rejoint les avantages du panel de gènes.

Le réel avantage compétitif de l'exome est son approche non biaisée. Étant donné que l'ensemble des régions exomiques est enrichi, l'analyse sera donc portée sur l'ensemble de ces régions sans aucun *a priori*. Ainsi, cet outil permet donc d'explorer des régions qui n'auraient pas été explorées en première intention. De plus, comme le nombre de cibles n'est plus une limitation, il est possible de ne plus se restreindre aux gènes connus et répertoriés, mais d'adjoindre des régions possiblement impliquées dans les maladies Mendéliennes comme les gènes codant pour des protéines hypothétiques, des régions régulatrices (UTR/promoteurs) et également des gènes de microARN par exemple. Ainsi, l'exome aurait déjà permis d'incriminer 130 gènes dans autant de pathologies jusqu'alors orphelines.

L'avantage découlant directement de la taille réduite de l'exome est son plus faible nombre de variants comparé au génome. En effet, un génome contient en moyenne 3,5 millions de variants nucléotidiques alors que l'exome en comporte 20 000 (dont 10 000 non-synonymes) <sup>101</sup>. Ainsi, l'exome favorise une certaine rapidité d'analyse bio-informatique grâce à ce nombre réduit de variants et deuxièmement, recentre l'analyse sur des régions interprétables. L'impact des variants situés en dehors des régions codantes reste en effet difficilement appréhendable <sup>102</sup>. De plus, 85 % des maladies Mendéliennes s'expliqueraient par des altérations présentes dans les régions codantes du génome. Pour cette raison,

séquencer l'exome est un outil intéressant pour identifier les bases moléculaires de maladies Mendéliennes telles que les cancers héréditaires<sup>103-105</sup>.

## **b. Stratégies d'analyses employant l'exome**

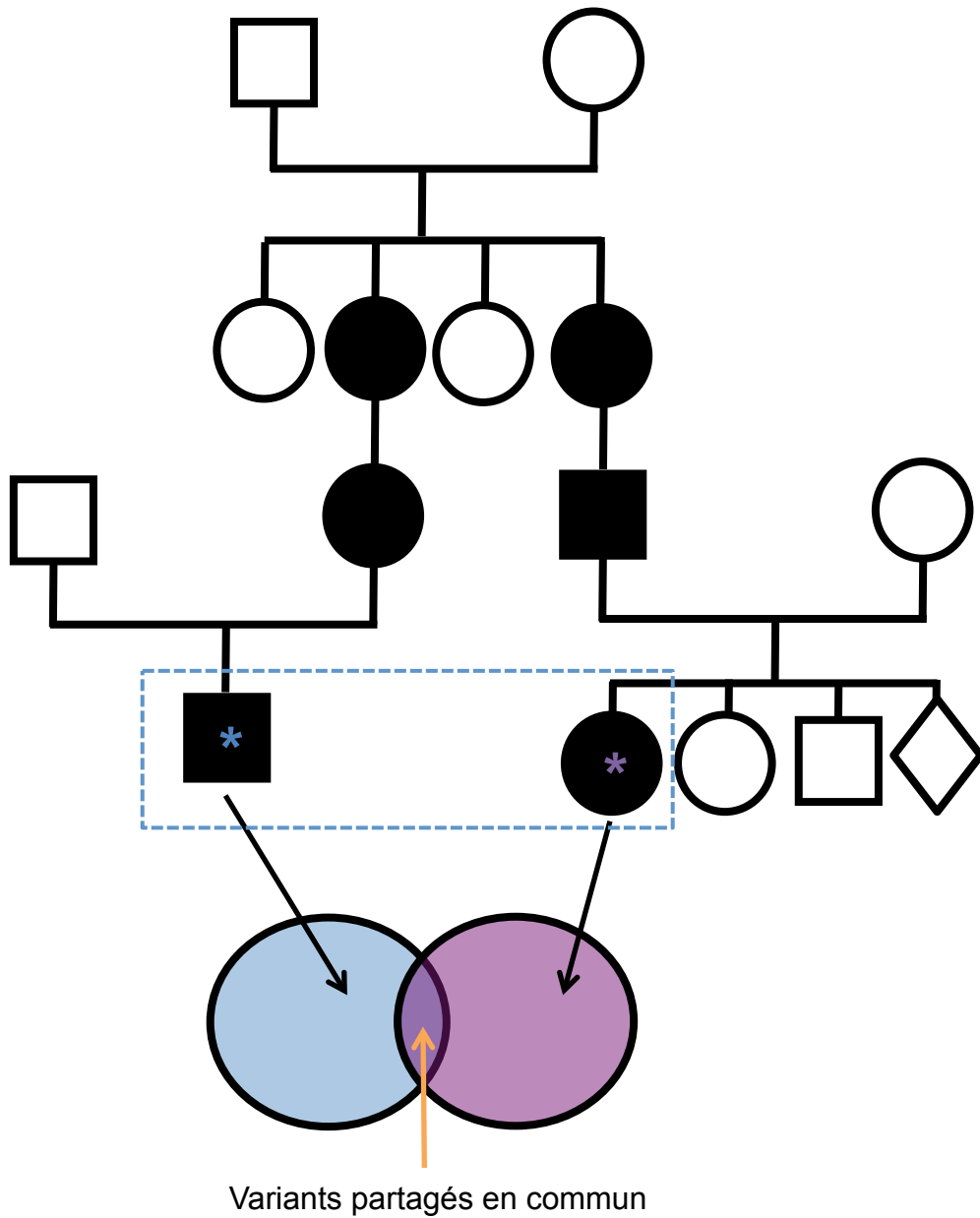
Depuis les articles princeps publiés par Sarah Ng et collaborateurs en 2009<sup>71,106,107</sup>, plusieurs dizaines de publications rapportent l'utilisation de l'exome comme outil dans l'identification de nouvelles bases moléculaires des prédispositions héréditaires aux cancers<sup>108</sup>.

Identifier la ou les variation(s) à l'origine du phénotype observé parmi les 20 000 d'un exome n'est pas chose aisée. De plus, 99 % des variations identifiées en exome ont une fréquence allélique inférieure à 1 % et sont pour plus de moitié des *singletons* (propre à un seul individu). Ainsi, en parallèle du développement technologique des techniques d'enrichissement, de nouvelles stratégies d'analyse bioinformatique ont été élaborées<sup>79,109</sup>. Il s'agit principalement de quatre grandes stratégies développées à partir d'un nombre restreint de patients :

- (i) la stratégie d'analyse exomique intrafamiliale
- (ii) la stratégie d'analyse exomique interfamiliale
- (iii) la stratégie d'analyse exomique soustractive, également dénommée analyse de trios enfants malade/parents sain.
- (iv) la stratégie exomique de comparaison tissu tumoral / tissu sain

### (i) La stratégie d'analyse exomique intrafamiliale

La stratégie exomique intrafamiliale est basée sur la comparaison d'exomes d'individus au sein de la même famille avec l'objectif d'identifier une altération commune à l'origine de la pathologie observée dans la famille (Figure 28). Généralement, cette stratégie est conduite dans une famille de grande taille en réalisant les séquençages de l'exome de plusieurs individus atteints et si possible, distants sur le plan génétique. L'avantage découlant est la restriction de l'analyse aux variants présentés en commun par ces individus; sans *a priori* sur le mode de transmission. A noter, il est possible de réaliser également le séquençage exomique d'un individu apparenté non atteint afin de faciliter l'analyse bioinformatique.



**Figure 28** : Stratégie d'analyse exomique intrafamiliale

L'objectif de cette stratégie est d'identifier la (ou les) altération(s) commune(s) aux patients de la famille pouvant expliquer leurs phénotypes. Pour cela, les exomes de deux patients éloignés sur le plan génétique sont séquencés. Puis, bioinformatiquement, les variants non partagés en commun sont éliminés de l'analyse afin de ne conserver que les variants communs pouvant expliquer la pathologie observée dans la famille.

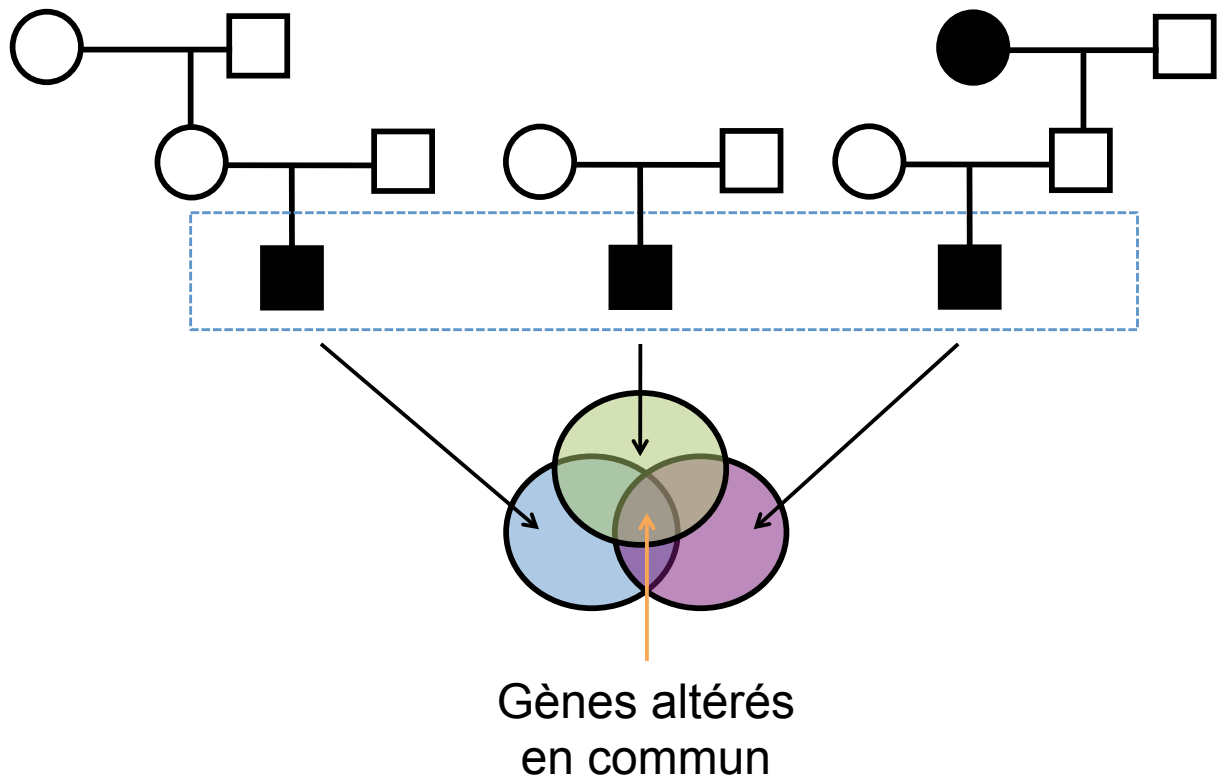
A titre d'illustration, ce type d'analyse a permis à Zhang et ses collaborateurs d'élucider les bases moléculaires d'une prédisposition héréditaire aux schwannomes multiples observée dans une grande famille alors qu'aucune base moléculaire n'avait pu être mise en évidence dans cette famille (*SMARCB1wt*)<sup>110</sup>. Les auteurs ont réalisé le séquençage exomique de trois individus présentant cette pathologie commune. L'analyse comparative a révélée 12 variations partagées en commun dont une variation dans le gène *COQ6*. Etant donné le rôle biologique de ce gène, les auteurs ont entrepris une démonstration fonctionnelle de la variation identifiée permettant par la suite d'imputer ce gène dans les prédispositions héréditaires aux schwannomes multiples.

Le succès de cette stratégie repose essentiellement sur le choix des individus. Le pourcentage de matériel génétique partagé par deux apparentés est inversement proportionnel aux nombres de méioses séparant ces individus. La disponibilité de deux individus apparentés séparés par un nombre suffisant de méioses réduit le nombre de variants en commun et garantit l'efficacité de la filtration. Pour exemple, le pourcentage de similitude entre deux frères est d'environ 50 %, alors qu'entre deux cousins germains éloignés au deuxième degré, ce taux descend à 12,5 %.

#### (ii) La stratégie d'analyse exomique interfamiliale

Ce type d'analyse est basé sur le séquençage de l'exome de plusieurs cas index de familles non apparentés à la recherche du gène altéré en commun chez ces familles. Le postulat de cette stratégie est donc de présupposer que le même gène est imputé chez plusieurs individus non apparentés présentant le même phénotype (Figure 29). L'efficacité de cette stratégie repose principalement sur un grand nombre de cas index séquencés pour obtenir une puissance statistique suffisante et sur une sélection phénotypique rigoureuse afin de s'affranchir d'éventuelles phénocopies. A titre d'illustration, cette stratégie a permis d'identifier un nouveau gène impliqué dans le mélanome. Pour cela, les auteurs ont sélectionné cinq cas index non apparentés provenant de Romagne (Italie) et présentant des mélanomes malins, sans base moléculaire imputable. Grâce au séquençage exomique des cas index et à l'analyse comparative interfamiliale, les auteurs ont pu mettre en évidence une altération fondatrice dans le gène *POT1* (*Protection Of Telomeres 1*) présente chez l'ensemble des patients criblés. Le criblage moléculaire du gène *POT1* a ensuite été étendu à d'autres familles italiennes permettant de révéler deux autres variants rares de *POT1*<sup>111</sup>. Parallèlement à cette étude, une autre équipe a confirmé le rôle du gène *POT1* dans les prédispositions héréditaires aux mélanomes à l'aide également d'une étude





**Figure 29 :** Stratégie d'analyse exomique interfamiliale

L'objectif de cette stratégie est d'identifier le(s) gène(s) altéré(s) en commun entre plusieurs familles non-apparentées. Pour cela, les exomes de cas index sont séquencés. Puis bio-informatiquement, seuls les gènes altérés en commun sont conservés dans l'analyse.

exomique interfamiliale. Il s'agissait de 105 familles (anglaises, hollandaises, et australiennes) pour lesquelles aucune base moléculaire n'a pu être préalablement mise en évidence <sup>112</sup>.

#### La stratégie d'analyse exomique soustractive

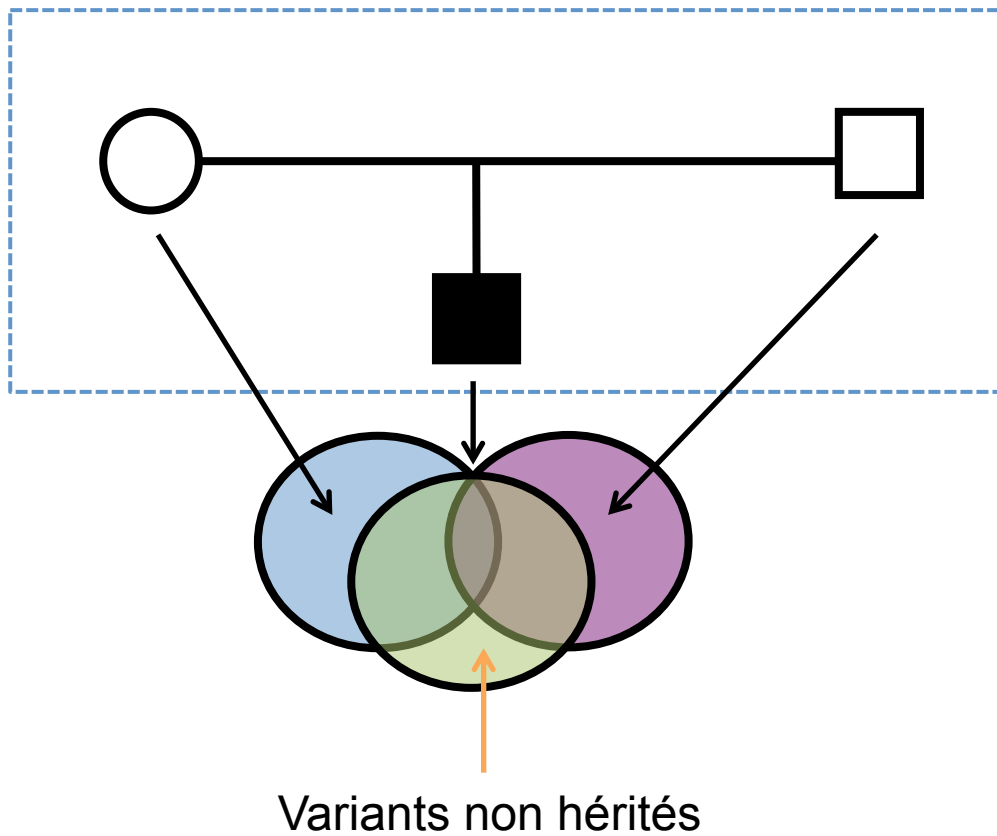
La stratégie des exomes soustractifs également dénommée stratégie des trios parents sains/enfant malade est une stratégie visant à identifier l'ensemble des variations présentes chez l'enfant malade et non héritées des parents indemnes de la maladie (Figure 30). Cette stratégie est majoritairement employée pour identifier des variants *de novo* mais peut être également utilisée pour identifier des variants hétérozygotes composites (résultants de l'héritage d'un variant du père et d'un variant de la mère dont l'effet cumulatif est responsable du phénotype) et des variants homozygotes chez l'enfant et hétérozygotes chez les parents (porteurs sains).

Cette stratégie est particulièrement intéressante lorsqu'un enfant développe une pathologie à un âge extrêmement précoce en l'absence de contexte familial de maladie. Ces phénotypes extrêmes sporadiques sont en effet, fortement évocateurs d'un événement génétique non présent chez les parents, comme une mutation *de novo*. En effet, dans le génome humain, le taux de mutations à chaque génération est compris entre  $1,5$  à  $2 \cdot 10^{-8}$  par paire de bases, soit 60 à 80 mutations *de novo* par génome. Cela représente 1 à 1,5 mutations *de novo* par exome <sup>75,76,113</sup>. Ces mutations sont théoriquement réparties en :

- 0,3 mutation synonyme
- 0,1 mutation perte de fonction
- 0,8 mutation faux sens

Il est important de noter que ces taux croissent en fonction de l'âge paternel <sup>78</sup>.

La stratégie exomique soustractive, initialement développée par l'équipe hollandaise de Lisenka Vissers et de Joris Veltman, est en réalité basée sur la soustraction bioinformatique de l'ensemble des variants exomiques identifiés chez les parents à ceux de l'enfant. Ainsi, ce type d'analyse permet, avec le séquençage de trois individus (l'enfant et les deux parents), de filtrer le nombre de variants à interpréter à 1 ou 2 parmi les 20 000 variants classiquement obtenus en WES <sup>72</sup>. Ce type d'approche a connu ses premiers succès dans le déterminisme du retard mental et l'autisme <sup>72,114-119</sup>



**Figure 30 :** Stratégie d'analyse exomique soustractive de trios

L'objectif de cette stratégie est d'identifier des variants *de novo* (non présents chez les parents) qui pourraient expliquer la pathologie observée. Pour cela, trois exomes sont réalisés : celui de l'enfant malade et de ses deux parents indemnes. Les variants hérités des parents sont ensuite soustraits bio-informatiquement à ceux de l'enfant afin d'identifier les variants *de novo* présents chez l'enfant.

de la schizophrénie <sup>120,121</sup> et très récemment dans le domaine cardiovasculaire <sup>122</sup>.

Dans le domaine de la cancérologie, cette approche a permis à notre unité d'identifier un nouveau gène de prédisposition au cancer de l'ovaire chez une femme ayant développé un adénocarcinome métastatique de l'ovaire à 21 ans en l'absence de contexte familial de cancer. Etant donné la précocité, le phénotype extrême et l'absence de base moléculaire identifiable, un séquençage exomique de trios a été réalisé. L'analyse de trio a révélé la présence d'une mutation *de novo* dans l'une des sous-unités des inhibines (*INHBA*). Puisque les inhibines participent au développement ovarien, des tests fonctionnels *in vitro* ont été entrepris dans cette famille de gènes ainsi qu'une étude de récurrence; conduisant à l'identification d'autres altérations constitutionnelles, notamment du gène partenaire d'*INHBA*: *INHA* <sup>123</sup>.

#### (iv) Stratégie exomique tissu sain / tissu tumoral

Dans le but d'identifier un événement constitutionnel à l'origine du phénotype, il est possible de réaliser le séquençage exomique du tissu tumoral et du tissu sain adjacent. Grâce à la soustraction bio-informatique {tissu tumoral - tissu sain}, il est possible d'obtenir l'ensemble des événements moléculaires apparus dans la tumeur. À partir de ces données et d'une bonne connaissance de la physio-pathologie tumorale, il est alors éventuellement possible d'identifier les bases moléculaires de la pathologie observée. L'une des premières publications relatant l'utilisation de cette stratégie date de 2009 et a été conduite par l'équipe de Jones. Les auteurs ont réalisé une comparaison entre le séquençage du génome d'un patient ayant développé un cancer pancréatique à un âge précoce sans base moléculaire identifiable et le séquençage exomique de sa tumeur. En postulant que l'allèle à l'origine du phénotype devait être présent constitutionnellement à l'état hétérozygote et altéré bi-alléliquement à l'état somatique, les auteurs ont identifié le gène *PALB2* comme étant potentiellement responsable. Etant donné le rôle interacteur de la protéine PALB2 avec la protéine BRCA2, les auteurs ont réalisé un séquençage ciblé de ce gène chez 96 patients ayant développé un cancer pancréatique et ont identifié 6 variations supplémentaires <sup>124</sup>. Ce type de stratégie a également été employé par l'équipe de Roberts pour le gène *ATM* dans la même pathologie <sup>125</sup>.

### **c. L'exome comme outil diagnostique**

Grâce à l'essor des technologies d'enrichissement et des coûts de séquençage décroissant, la question de l'utilisation de l'exome dans le contexte du diagnostic moléculaire des maladies Mendéliennes et notamment des prédispositions héréditaires aux cancers peut

être envisagée. Cette nouvelle approche qualifiée d'exome diagnostique (CES, *Clinical Exome Sequencing*), a l'avantage de séquencer l'ensemble des régions codantes et donc de ne plus être restreinte à un ensemble de gènes comme cela est le cas dans les panels de gènes. Cette propriété est particulièrement intéressante dans le domaine des prédispositions héréditaires aux cancers, étant donné l'évolution rapide des connaissances et la nécessité de séquencer plusieurs gènes en parallèle dû à l'hétérogénéité phénotypique.

La première description de l'utilisation du CES dans le diagnostic moléculaire de pathologie Mendélienne a été réalisé par le *Baylor College of Medicine* à Houston en 2013. Les auteurs ont réalisé un exome au titre du diagnostic pour une cohorte de 250 patients, majoritairement des enfants avec retards développementaux et des enfants avec un spectre autistique. Les auteurs ont pu identifier la mutation à l'origine de la maladie génétique chez 62 patients, soit un taux de détection de 25 %<sup>126</sup>. En 2013 également, une autre équipe s'est intéressée à l'étude de 186 patients présentant soit un retard mental, un trouble moteur, une cécité, une surdité, une maladie mitochondriale ou un cancer colorectal. Les auteurs ont démontré que cette approche est beaucoup plus sensible que les techniques classiquement utilisées jusqu'alors en diagnostic. Pour les patients atteints de surdité ou de troubles moteurs, le CES permet même un diagnostic quatre fois plus souvent comparativement au séquençage Sanger<sup>127</sup>. Au même titre que la stratégie d'analyse exomique soustractive en trio développée au titre de la recherche l'exome diagnostique peut être réalisé en trios dans le but d'identifier les variants *de novo* ou d'améliorer la filtration des variants délétères en fonction de la transmission supposée (ex : consanguinité)<sup>128</sup>.

L'avantage incontestable du CES est le criblage de l'ensemble des régions codantes en parallèle, ce qui représente 100 à 1000 fois plus de gènes analysés en comparaison avec un panel et 20 000 fois comparé au criblage d'un seul gène en Sanger. En terme d'avantage financier, il a été évalué qu'un CES est rentable au delà de trois gènes testés en Sanger (CES = 1900 €; Séquençage Sanger = 600 € / gène)<sup>127</sup>. Le CES permet donc à la fois de faire le diagnostic de maladies Mendéliennes dans le sens classique du terme, c'est à dire de cribler des gènes qui auraient été explorés dans le cadre d'un diagnostic moléculaire d'une pathologie en question, mais également d'explorer des gènes qui auraient été criblés en seconde intention en cas de résultat négatif et des gènes qui n'auraient pas été séquencé du tout, au sein d'une même analyse. A titre d'illustration, en 2016, l'équipe de Kansal a rapporté un enfant de 4 mois présentant un retard développemental sévère, des hémangiomes cutanés, des tumeurs bilatérales (du vestibules, de l'hypoglosse, cérébelleuses et lombaires) ainsi que des tâches café au lait pour lequel le séquençage des gènes *NF1*, *NF2*, *RASA1* et *SMARCB1* s'est avéré négatif. De même, l'immunohistochimie des protéines MLH1, MSH2, PMS2 et MSH6 n'avait pas montré d'inactivation. Une CGH-

array avait cependant montré une duplication de 133 kb emportant *FOXG1*, expliquant sûrement le retard mental observé mais pas la survenue de tumeurs. Ainsi, un CES a été réalisé et ce dernier a permis de mettre en évidence deux variants hétérozygotes de *MLH3* démontrant ainsi qu'une inactivation biallélique du système MMR était à l'origine du phénotype observé <sup>129</sup>.

#### **d. Analyse de panel *in silico* à partir de l'exome**

Étant donné que l'ensemble des régions codantes sont capturées, il est possible de créer artificiellement à façon un panel virtuel de gènes sur lequel l'analyse va être focalisée. Cette approche qualifiée de panel *in silico* est réalisée bioinformatiquement en utilisant une liste de gènes d'intérêt. Cela permet d'analyser un certain nombre de gènes en première intention, puis si besoin d'autres gènes *ad libitum* sans manipulation supplémentaire. Cette approche permet donc une réactivité importante et une grande flexibilité. Par exemple, lorsqu'un nouveau gène de prédisposition au cancer est identifié, il suffit d'ajouter informatiquement le gène au panel virtuel pour obtenir très rapidement les variants de ce gène dans la cohorte, y compris de façon rétrospective.

En 2015, Zhang et ses collaborateurs ont utilisé cette approche de panel *in silico* à partir de données d'exomes (n=456) et de génomes (n=595) de jeunes patients ayant développé une tumeur pédiatrique. Ainsi, les auteurs ont pu créer artificiellement trois panels à partir de leurs données de séquençage : un panel de 60 gènes impliqués dans des prédispositions héréditaires au cancer à transmission autosomique dominante, un panel de 29 gènes pour les transmissions autosomiques récessives de prédispositions héréditaires au cancer, ainsi qu'un panel de 476 gènes impliqués dans la cancérogenèse car présentant des altérations somatiques récurrentes. Cette approche a permis l'identification d'une altération constitutionnelle causale dans environ 8,5 % des individus analysés dans cette étude <sup>130</sup>.

#### **e. Remarques sur l'utilisation de l'exome et de ses approches**

Au delà de la limitation fondamentale de l'exome de ne pas inclure les régions non codantes, son utilisation en recherche fondamentale comme clinique se justifie pleinement. Gilissen et ses collaborateurs estiment qu'environ 60 % des projets trouvent une réponse à l'aide de l'exome <sup>131</sup>. Cela s'explique certainement par le fait que la majorité des maladies Mendéliennes résultent d'altérations se situant dans une région codante <sup>132</sup>. Deuxièmement, la nature non biaisée et sans a priori de ces approches en font des outils puissants dans le déterminisme de maladie Mendélienne. Grâce à la multitude de stratégies bioinformatiques développées, un nombre important de solutions techniques peuvent être apportées en

fonction de la famille pour limiter le nombre de variants à interpréter, y compris dans les maladies Mendéliennes sporadiques avec l'identification de mutations *de novo* rares. Ces nouvelles approches sont donc cruciales pour établir de nouvelles corrélations génotype-phénotype et pour améliorer le diagnostic moléculaire des familles ainsi que leurs conseils génétiques <sup>102</sup>.

### **Découvertes additionnelles**

Avec le CES notamment, il existe un risque d'identifier fortuitement une altération génétique non envisagée lors de l'indication de l'exome et pouvant avoir une relevance clinique. Ce type de découverte est dénommé "découverte incidentelle" ou "incidentalome". Cependant, le terme de découverte additionnelle semble être préféré au terme incidentalome pour son côté plus neutre <sup>133</sup>. En tout état de cause, le terme de découverte additionnelle est réservé à des découvertes réellement fortuites, et non à la recherche délibérée de variants pathogènes ne correspondant pas à l'indication première du test génétique, et dont la connaissance permettra la mise en œuvre de mesures de soin ou préventive.

Dans le contexte de séquençage d'exome CES, le collège américain de Génétique Médicale et de Génomique (*The American College of Medical Genetics and Genomics; ACMG*) a décidé de constituer un groupe de travail sur les découvertes additionnelles. Ce groupe de travail a formulé des recommandations ayant pour but d'aiguiller les biologistes et les cliniciens, en fournissant par exemple, une liste des gènes pour lesquels les altérations doivent être systématiquement rendues aux cliniciens. Ces recommandations s'appliquent pour les laboratoires travaillant avec du matériel constitutionnel (exome, génome, tissu "normal" lors de la stratégie de soustraction tumorale, etc.) <sup>134</sup>.

En France, il existe, depuis la loi du 7 juillet 2011 (article L.1131-1-2 du Code de la santé publique), une obligation d'information de la parentèle en cas de diagnostic d'une maladie d'origine génétique (héréditaire) grave. Cette obligation ne résout donc pas cette épineuse question éthique. Il perdure une difficulté entre établir un équilibre entre l'autonomie de la personne, le respect de cette autonomie et le devoir de solidarité qui s'exprime également dans la mise en garde et la prévention d'un risque ou d'un danger d'origine génétique.

La Commission présidentielle pour l'étude des questions de bioéthique aux États-Unis et la Société européenne de génétique humaine rappellent néanmoins la nécessité d'entreprendre une démarche de décision partagée avec les patients au sujet du périmètre de la communication des résultats des examens génétiques et des modalités à mettre en œuvre en cas de découvertes incidentes. Dès lors, les médecins doivent respecter la

préférence d'un patient de ne pas connaître les découvertes incidentes ou secondaires tout en restant cohérents avec leur devoir professionnel d'assistance.

Un ordre d'idée de la fréquence des découvertes additionnelles a pu être déterminé par l'étude d'une cohorte de 1000 exomes témoins issus de la cohorte ESP. En se focalisant sur une liste de 114 gènes dont les altérations sont fortement associées avec des maladies Mendéliennes, Dorschner et ces collaborateurs ont relevés 23 individus présentant une altération délétère dans l'un de ces 114 gènes, soit 2,3 % des individus testés (3,4 % chez les européens et 1,2 % chez les africains) <sup>74</sup>. Cette notion de découverte additionnelle est donc à spécifier explicitement dans le consentement du patient étant donné son côté non-anecdotique.

### **Vers le génome diagnostique ?**

Il y a fort à parier que dans le futur, l'exome diagnostique sera remplacé tôt ou tard par le génome diagnostique <sup>135</sup>. En effet, l'utilisation du séquençage tout génome présente l'avantage de donner accès aux séquences non codantes, d'avoir une uniformité de couverture importante, d'offrir une détection plus efficace des événements structuraux tels que les variations du nombre de copies (*Copy Number Variation* ; CNV), les translocations et les inversions et d'accéder aux variants introniques profonds par exemple. Cette transition ne pourra être effective que lorsque les coûts de séquençage seront plus faibles, que les débits des séquenceurs seront encore plus importants qu'il ne le sont aujourd'hui et que les laboratoires se seront équipés de matériels informatiques adéquats. De plus, de nouvelles publications sous entendent que le séquençage de génome (*Whole Genome Sequencing*; WGS) serait meilleur dans la détection de variants y compris dans la détection de variants exoniques <sup>136</sup>.





# **OBJECTIFS**



## Objectif des travaux de thèse

Notre unité de recherche travaille en étroite collaboration avec le laboratoire de Génétique Moléculaire du CHU de Rouen en charge du diagnostic moléculaire du syndrome de Li-Fraumeni au niveau national. Ce syndrome rare est l'une des formes héréditaires de cancer les plus graves caractérisée notamment par une hétérogénéité du spectre tumoral ; à savoir des ostéosarcomes, des sarcomes des tissus mous, des tumeurs cérébrales, des corticosurrénales, des cancers du sein pré-ménopausiques et de la femme jeune; et par la précocité de survenue de la première tumeur. D'un point de vue moléculaire, cette maladie est causée par des altérations constitutionnelles du gène *TP53*, gène codant pour le facteur de transcription p53. Sur l'ensemble des familles évocatrices d'un syndrome LFS qui sont adressées pour un criblage moléculaire du gène *TP53* au laboratoire de Génétique Moléculaire, 82 % des patients restent sans mutation détectée du gène *TP53* et sont donc orphelines de tout diagnostic moléculaire<sup>137</sup>. Il est possible qu'une partie de ces cas inexplicables puissent correspondre à des agrégations de cancers sporadiques. Une autre partie pourrait s'expliquer par la présence de mutations dans des régions du gène *TP53* non explorées dans le cadre du diagnostic (introns, régions non traduites en 5' et 3', régions régulatrices à distance). Il est également probable que des mutations dans d'autres gènes entraînant un phénotype similaire à celui observé chez les porteurs de mutations de *TP53* soient à l'origine de certains cas inexplicables.

Ainsi, de par ce recrutement, nous avons accès à une collection remarquable de patients évocateurs d'un syndrome de Li-Fraumeni pour lesquels aucune base moléculaire à l'origine du phénotype n'a pu être mise en évidence. Ce réservoir de patients évocateurs constitue une opportunité exceptionnelle d'identifier de nouvelles bases moléculaires du cancer et notamment des cancers affectant l'enfant.

Pour rechercher et identifier les altérations génétiques à l'origine de ces phénotypes, j'ai tiré parti du séquençage à haut-débit et des nouveaux outils qu'offrent l'enrichissement de régions d'intérêt. Ainsi, nous avons mis en place et appliqué quatre stratégies pour tenter d'identifier de nouvelles bases moléculaires des cancers à survenue précoce en fonction de la présentation familiale des patients :

1. une **stratégie exomique intrafamiliale** dans une famille présentant une agrégation de chondrosarcomes.
2. des **stratégies exomiques soustractives** chez des patients ayant développé une ou plusieurs tumeurs à un âge précoce et en l'absence de contexte familial de cancer.
3. une **stratégie exomique interfamiliale** pour de jeunes patients ayant développé à un âge précoce un corticosurrénales, une tumeur particulièrement exceptionnelle.

4. une stratégie de **séquençage d'un panel de 201 gènes impliqués dans le cancer** chez des patients ayant développé une ou plusieurs tumeurs à un âge pédiatrique dans un contexte familiale ou sporadique de cancer.

# **RESULTATS**



**Projet 1 : Identification par la stratégie d'analyse exomique intrafamiliale des bases moléculaires à l'origine d'une agrégation familiale atypique de chondrosarcomes costaux solitaires**

**Article "Familial Solitary Chondrosarcoma Resulting from Germline *EXT2* Mutation"**

Abdelkader Heddar<sup>1\*</sup>, **Pierre Fermey<sup>1\*</sup>**, Sophie Coutant<sup>1</sup>, Emilie Angot<sup>2</sup>, Jean-Christophe Sabourin<sup>2</sup>, Paul Michelin<sup>3</sup>, Nathalie Parodi<sup>1</sup>, Françoise Charbonnier<sup>1</sup>, Myriam Vezain<sup>1</sup>, Gaëlle Bougeard<sup>1</sup>, Stéphanie Baert-Desurmont<sup>1</sup>, Thierry Frébourg<sup>1</sup>, Isabelle Tournier<sup>1</sup>

<sup>1</sup> Inserm U1079, Faculty of Medicine, UNIROUEN, Normandie Univ and Department of Genetics, Rouen University Hospital, Normandy Centre for Genomic and Personalized Medicine

<sup>2</sup> Department of Pathology, Rouen University Hospital    <sup>3</sup> Department of Radiology, Rouen University Hospital

**\* Ces auteurs ont contribué de manière équivalente à ce travail.**

Publié dans Genes Chromosomes Cancer, 56: 128–134. doi:10.1002/gcc.22419 (2017)



## Le chondrosarcome (CS)

- ❑ Représente 20 % des tumeurs osseuses
- ❑ Est la tumeur osseuse la plus commune après le myélome et l'ostéosarcome
- ❑ Est majoritairement de localisation pelvienne et sur les os longs
- ❑ Peut être subdivisé en deux catégories :
  - les chondrosarcomes primaires
  - les chondrosarcomes secondaires se développant à partir de tumeurs bénignes préexistantes

### Chondrosarcomes primaires

- ❑ **Dans le contexte du syndrome Li-Fraumeni (*TP53*)**
  - 2% des 322 porteurs d'altérations *TP53* ont développé un CS primaire (Bougeard *et al.* ; 2015)

### Chondrosarcomes secondaires

- ❑ **Dans le contexte de la maladie des Ostéochondromes Multiples (*EXT1/EXT2*)**
  - Les ostéochondromes (tumeurs bénignes **périphériques**) peuvent dégénérer en CS secondaires
- ❑ **Dans le contexte de la maladie d'Ollier-Malfucci (*IDH1/IDH2*)**
  - Les enchondromes (tumeurs bénignes **centrales**) peuvent dégénérer en CS secondaires

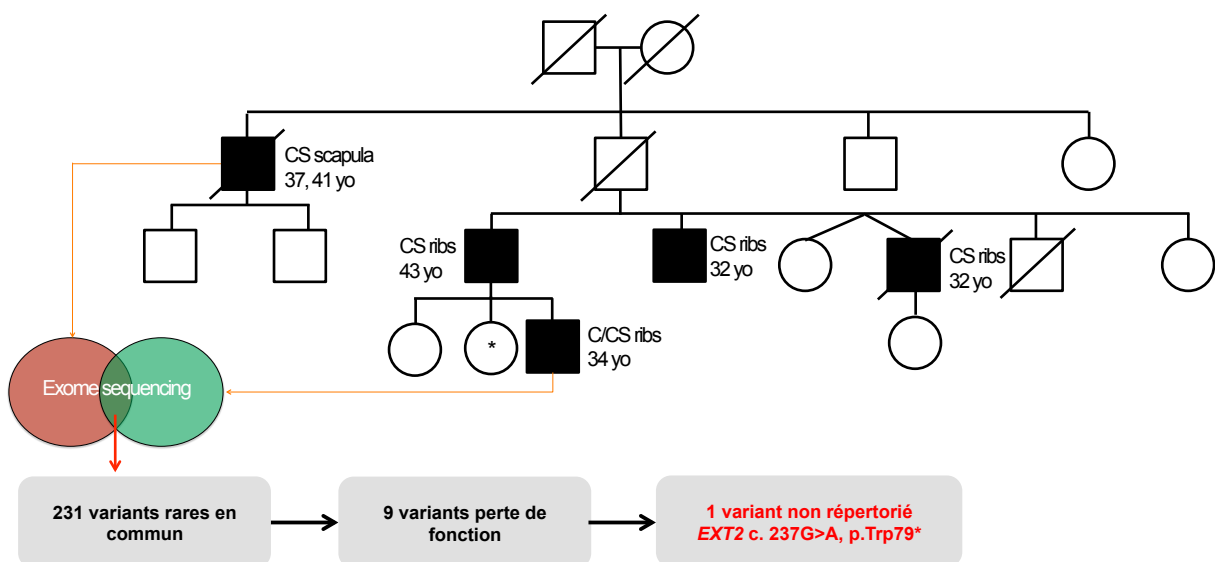


**Figure 31** : Génétique des chondrosarcomes

Bougeard *et al.* 2015 <sup>137</sup>

## I. Contexte scientifique

Le chondrosarcome (CS) est une tumeur maligne d'origine cartilagineuse qui représente plus de 20 % de l'ensemble des tumeurs osseuses. Il se localise préférentiellement au niveau du squelette axial et ne concerne la cage thoracique que dans moins de 15 % des cas. Plus de 90 % des CS sont des tumeurs primitives survenant généralement chez l'adulte de plus de 50 ans. Dans certains cas, le CS peut être secondaire à une lésion bénigne préexistante (enchondrome, ostéenchondrome). Cependant, le risque de transformation maligne de ces lésions bénignes varie de 1 à 4 % pour l'ostéenchondrome et l'enchondrome solitaires, de 3 à 5 % pour les ostéenchondromes multiples ou maladie des exostoses multiples, maladie autosomique dominante due à des mutations des gènes *EXT1* ou *EXT2* et peut atteindre jusqu'à 40 % en cas d'enchondromatose associée à des mutations somatiques récurrentes d'*IDH1* ou *IDH2* associées aux maladies d'Ollier et de Maffucci respectivement (Figure 31). Alors qu'aucune transmission familiale de CS n'a été rapportée à ce jour en dehors des rares cas liés au syndrome de Li-Fraumeni, nous rapportons ici une forme familiale de chondrosarcomes isolés de localisation exclusivement thoracique et d'évolution favorable. Etant donné la présentation atypique de cette grande famille, nous avons réalisé le séquençage exomique de deux individus apparentés atteints et employé la stratégie d'analyse exomique intrafamiliale pour élucider les bases moléculaires à l'origine de cette présentation familiale unique (Figure 32).



**Figure 32 :** Identification d'une mutation du gène *EXT2* par la stratégie d'analyse exomique intra-familiale

231 variants rares (fréquence inférieure à 0,0003) en commun ont été identifiés à l'aide du séquençage exomique de deux apparentés atteints au 3ème degré. Parmi ces variants, neuf sont des variants tronquant la protéine. Huit de ces variants étaient déjà présents dans notre base de données et correspondent à des artefacts. Le variant restant est un variant non-sens affectant le gène *EXT2*. CS =Chondrosarcome. C=Chondrome. Ribs = côtes.

## Familial Solitary Chondrosarcoma Resulting from Germline *EXT2* Mutation

Abdelkader Heddar,<sup>1†</sup> Pierre Fermey,<sup>1‡</sup> Sophie Coutant,<sup>1</sup> Emilie Angot,<sup>2</sup> Jean-Christophe Sabourin,<sup>2</sup> Paul Michelin,<sup>3</sup> Nathalie Parodi,<sup>1</sup> Françoise Charbonnier,<sup>1</sup> Myriam Vezain,<sup>1</sup> Gaëlle Bougeard,<sup>1</sup> Stéphanie Baert-Desurmont,<sup>1</sup> Thierry Frébourg,<sup>1\*</sup> and Isabelle Tournier<sup>1</sup>

<sup>1</sup>Inserm U1079, Faculty of Medicine, UNIROUEN, Normandie Univ and Department of Genetics, Rouen University Hospital, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

<sup>2</sup>Department of Pathology, Rouen University Hospital, Rouen, France

<sup>3</sup>Department of Radiology, Rouen University Hospital, Rouen, France

Germline mutations of *EXT2*, encoding Exostosin Glycosyltransferase 2, are associated with multiple osteochondromas (MO), an autosomal dominant disease characterized by the development of multiple peripheral cartilaginous benign tumors with a weak risk of malignant transformation. We report here a family with a remarkable clinical presentation characterized by the development of isolated chondrosarcomas, mostly located in ribs. Comparative analysis of exomes from two third-degree affected relatives led us to identify a single common disruptive variation, corresponding to a stop mutation (c.237G > A, p.Trp79\*; (NM\_000401.3); c.138G > A, p.Trp46\*; (NM\_207122.1)) within exon 2 of the *EXT2* gene. Interestingly, no obvious sign of MO was detected in affected members by radiological examination. This report shows that germline mutations of *EXT2* can result, not only in the development of multiple benign osteochondromas, but also in the development of isolated malignant cartilaginous tumors including central tumors, and that the presence of germline *EXT2* mutation should be considered in patients suspected to have an inherited predisposition to chondrosarcoma, even in the absence of MO. © 2016 Wiley Periodicals, Inc.

### INTRODUCTION

Chondrosarcoma (CS) is one of the most common malignancies of bones after myeloma and osteosarcoma and represents more than 20% of all bone tumors (Fletcher et al., 2002). They are frequently located in the pelvis and long bones and, in contrast, their thoracic location is uncommon. Ninety percent of these tumors are primary CS that arise in adults older than 50. Alternatively, CS can be secondary to a preexisting benign lesion, mainly osteochondromas and enchondromas and usually affects younger adults.

Osteochondromas - previously called cartilaginous exostoses - are the most frequent peripheral benign tumors and correspond to cartilage-capped bony projections arising on the external surface of bones. They usually arise from the metaphyseal regions of long bones (Bovéé et al., 2010). Osteochondromas are solitary lesions, but multiple osteochondromas (MO) can occur in the context of hereditary multiple osteochondromas, an autosomal dominant disease ([MIM]#133700 and 133701) (Bovéé, 2008), resulting from inactivating mutations within one of the two tumor suppressor genes, *EXT1* or *EXT2*, encoding the exostosins 1

and 2 (Wuyts et al., 1993; Bovee, 2008). Enchondromas, located within the medullary cavity of bones, are one of the most frequent benign tumors and are usually asymptomatic. Typical locations for enchondromas include the small bones of the hands and feet, followed by the meta- and diaphysis of the long bones. Enchondromas most often occur as solitary lesions, but they may also present as multiple lesions (enchondromatosis). These conditions include Ollier disease ([MIM] #166000) and Maffucci syndrome ([MIM] #614569), resulting from mosaic mutations of the *IDH1* and *IDH2* genes (Amary et al., 2011; Pansuriya et al., 2011).

Both subtypes of benign cartilaginous tumors can progress to malignancy, but this progression

Additional Supporting Information may be found in the online version of this article.

†Abdelkader Heddar and Pierre Fermey contributed equally to this work.

\*Correspondence to: Thierry Frébourg, Inserm U1079, Faculty of Medicine, Normandie University, 22 Boulevard Gambetta, 76183 Rouen Cedex, France; Phone: 33 2 32 88 81 82; Fax: 33 2 32 88 80 80; Mail: Frébourg@chu-rouen.fr

Received 7 June 2016; Revised 9 September 2016; Accepted 12 September 2016

DOI 10.1002/gcc.22419

Published online 25 October 2016 in Wiley Online Library (wileyonlinelibrary.com).

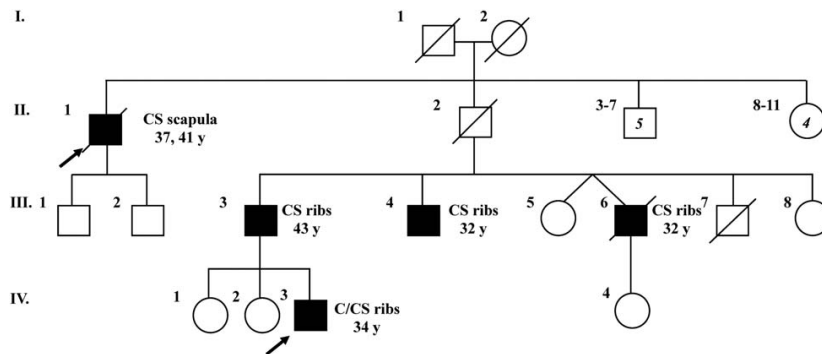


Figure 1. Pedigree of the family. The black filled squares represent patients with chondroma or chondrosarcoma and the open symbols represent unaffected individuals. Italicized numbers indicate the number of individuals. Ages in years (y) at diagnosis are indicated. The arrows show the relatives who have been subjected to whole exome sequencing. C, chondroma; CS, chondrosarcoma.

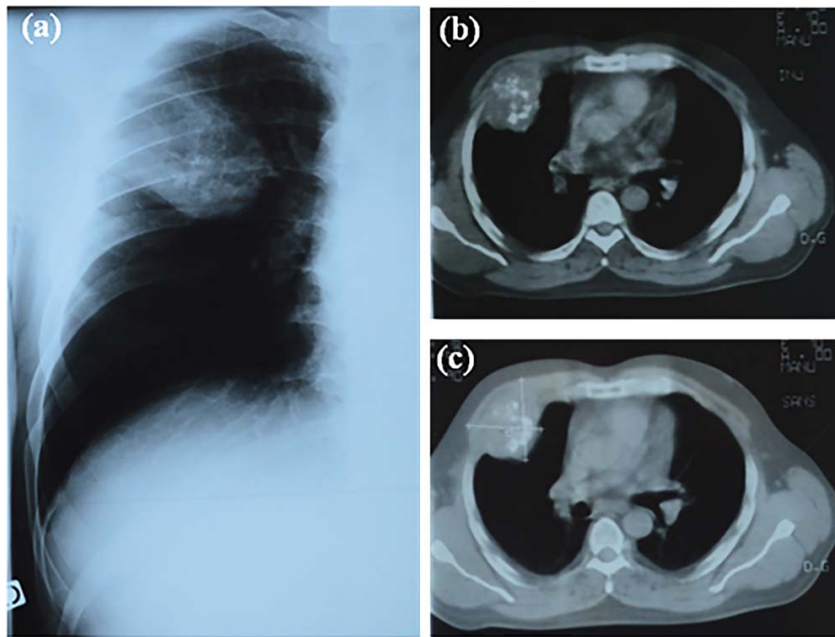


Figure 2. Preoperative images in patient III.3. (a) Chest wall radiography showing a large parietal mass with punctuate calcifications. (b) and (c), CT scan transverse views showing a well circumscribed mass of the third rib with confluent calcifications typical of a chondroid matrix. No imaging feature suggestive of the transformation of a

preexisting osteochondroma could be detected: the mass symmetrically extends to soft tissues surrounding the rib it originates from and no cortical deformation is seen. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

occurs more commonly with enchondromatosis than with osteochondroma: indeed, malignant transformation is observed in 1-5% of MO cases, 4% of solitary enchondroma cases and in up to 35% of cases of Ollier disease and over 50% of Maffucci syndrome cases (Altay et al., 2007, Bovee et al., 2010). Beside MO and enchondromatosis, the only Mendelian condition predisposing to CS

is the Li-Fraumeni syndrome (LFS; [MIM] #151623), a rare syndrome resulting from *TP53* mutations and characterized by a wide tumor spectrum. Nevertheless, in a large series of LFS patients including 322 affected *TP53* mutation carriers, we observed CS in only 2% of the *TP53* mutation carriers (Bougeard et al., 2015). In contrast, to our knowledge, no genetic predisposition

exposing only to solitary CS has been described. We report here, using exome sequencing, the identification of a germline *EXT2* mutation in a family with a remarkable clinical presentation characterized by inherited solitary CS.

## MATERIALS AND METHODS

### Clinical Data and Medical History

The pedigree of the family is shown in Figure 1. The index case (patient III.4) was diagnosed at the age of 32 with a chondrosarcoma of the 5th right rib, which was treated by surgical resection. Radiographs of the upper and lower limbs, pelvis and spine did not reveal any osteochondroma, and no recurrence or metastasis was detected in over 30-years' follow-up. His brother (patient III.3) had presented at age 43, with a large tumor (6 x 4.5 cm) arising from the 3<sup>rd</sup> right rib (Fig. 2). The tumor, that had partially destroyed the 3<sup>rd</sup> and 4<sup>th</sup> ribs, corresponded to a CS (Fig. 3 (a)). This chondrosarcoma of the ribs was treated by large surgical resection and reconstruction with bone graft. The patient was followed-up for more than 16 years and physical examination, radiography, thoracic computed tomography and bone scintigraphy revealed neither recurrence nor metastasis. Another brother (patient III.6) died in the context of lung metastases of a costal chondrosarcoma, 2 years after the diagnosis at the age of 32. In patient IV.3, radiography performed at the age of 34 for back pain revealed incidentally on the right 9<sup>th</sup> rib a tumoral lesion which was shown to be an enchondroma or a very small and well differentiated central CS (Fig. 4). No signs of osteochondroma or bone modeling defect could be detected by full body radiographies of this patient (Supporting Information Figure 1). A paternal uncle of the index case (patient II.1) had presented at age 37 a chondrosarcoma of the left scapula revealed by a rapidly progressing swelling on the upper back (Fig. 5). This tumor was treated successfully by surgical resection. However, 4 years later, the mass on the scapula increased in size, reaching 15 x 5 cm. Two resected specimens revealed a chondrosarcoma with extension to soft tissues but without metastasis (Fig. 3b). The patient was treated at that time with three courses of radiotherapy. Progression was marked by a gradual installation of persistent pain and reduction in the shoulder motion range that required a scapulothoracic disarticulation. No exostosis had been noticed in the physician's report or in the last scintigraphy.

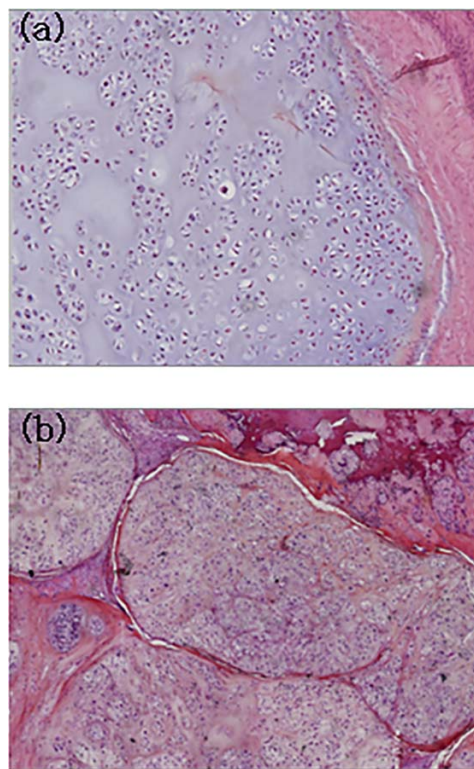


Figure 3. (a) Hematein, eosin plus saffron stain of patient III.3 tumor showing a well-differentiated chondrosarcoma with mild cytologic atypia and chondroid matrix (X 100). (b) Hematein, eosin plus saffron stain of patient II.1 tumor showing a well differentiated chondrosarcoma with mild cytologic atypia and chondroid matrix (X 25). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Histological examination of the surgical specimen revealed a large tumor measuring 24x19x10 cm and confirmed the diagnosis of chondrosarcoma. Follow-up of the patient for more than 17 years after surgery did not show recurrence or metastasis. No germline *TP53* mutation was identified in this family.

### Exome Analysis

Exome analysis was performed on genomic DNA extracted from peripheral blood of patients II.1 and IV.3. Exon enrichment was performed on 3  $\mu$ g of DNA, using the SureSelect Human All Exons V4 kit (Agilent Technologies, Santa Clara, USA). Exon-enriched libraries were subjected to a 75 bp paired-end sequencing on an Illumina GAIIx, according to the manufacturer's protocol and raw image files were processed using an Illumina pipeline (Off-line Basecaller 1.9.4). The alignment of the reads to the human reference genome (Human hg19/GRCh37) and the variant

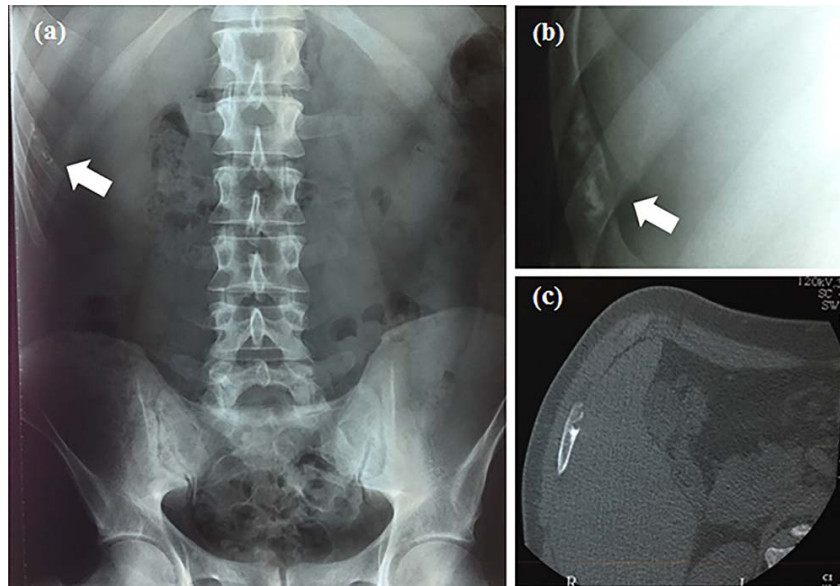


Figure 4. Preoperative radiographies in patient IV3. (a, b) Chest wall radiography showing a finely calcified lytic lesion of the 9<sup>th</sup> right rib. (c) Transverse computed tomography image showing a small chondroid lesion centrally located in the medulla of the 9<sup>th</sup> right rib. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

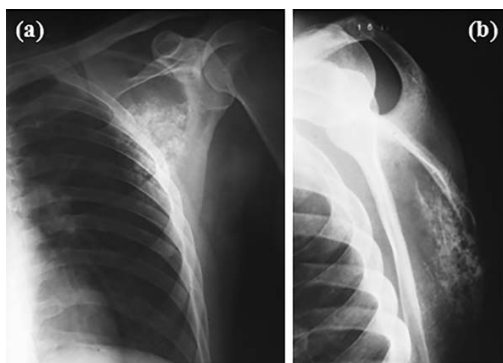


Figure 5. Preoperative anteroposterior (a) and lateral (b) radiographies of the scapula in patient II.1 showing a large chondroid lesion suggestive of a superficial chondrosarcoma. No cortical deformity of the scapula suggestive of a preexisting osteochondroma can be seen.

calling were performed using CASAVA 1.8.2 (Illumina, San Diego, USA). The annotation was performed using Annovar ([www.openbioinformatics.org/annovar](http://www.openbioinformatics.org/annovar)). Visualization of the BAM files was performed using the Alamut Visual software (Interactive Biosoftware, Rouen, France).

#### Sanger Sequencing Analysis

The DNA fragment from *EXT2* exon 2 was PCR-amplified from 50 ng of genomic DNA extracted

from peripheral blood using the following primers: EXT2-Ex2-F 5' TCCTGCCCTCATCCCAAGAA 3' and EXT2-Ex2-R 5' CTTGATGCGCAGTGT GTTCT 3' and from 5 to 10 ng of DNA extracted from the tumor with the following primers: EXT2-F2 5' TTCTCCATTGTCCCTCCTGGG 3' and EXT2-R2 5' CCTAACAACCGGCACATCAC 3'. Sanger sequencing of these fragments were performed using the BigDye® Terminator v3.1 Cycle sequencing Kit (Life Technologies, Carlsbad, USA). Sequencing reactions were migrated on a 3100 xl Genetic Analyzer (Life Technologies) and analyzed using the Sequencing analysis software 5.2.0 (Life Technologies).

#### RESULTS

In order to characterize the molecular basis of the inherited solitary chondrosarcomas observed in this family, we used the intra-familial exome sequencing strategy. Exome sequencing was performed on peripheral blood DNA from patients II.1 and IV.3, two third-degree affected relatives (Fig. 1). Across the two exomes, we obtained an average of 6 Gb with 98% of mappable sequences, a mean read depth of 75x, 96% of bases were covered to a minimum depth of 10x and 91% of the read bases had a Qscore above 30. First, variants

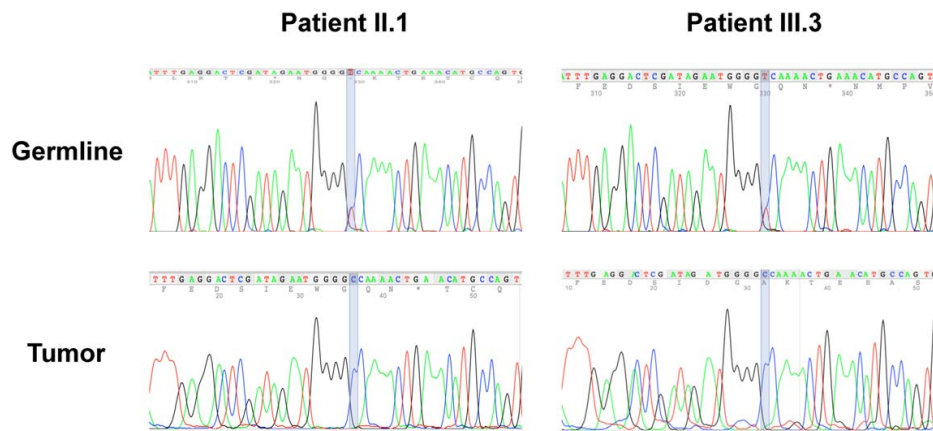


Figure 6. Sanger electropherograms from germline DNA (upper panel) and tumoral DNA (lower panel) from patient II.1 and III.3 centered on the position of the c.237G > A variant of *EXT2* exon 2 (NM\_000401.3). The light blue box indicates the position of the variation. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

with a read coverage of less than 10x and a Qscore below 20 were filtered out. Then, we filtered the variations against the 1000 Genomes Project data set (May 2011, 20101123 release, <http://www.1000genomes.org>) and the ESP cohort data set (ESP6500, <http://evs.gs.washington.edu/EVS/>) using an allelic frequency filter of 0.0003. Using these filters, we detected 231 exonic variants (exons and splice junctions) common between patient II.1 and patient IV.3. Among these variants, 9 corresponded to frameshift deletions and insertions and stop gain and were considered as disruptive. Eight of these variants were already present in our in-house database and careful examination of the corresponding BAM files indicated that these variants likely correspond to sequencing artifacts. The only remaining disruptive variation shared by individuals II.1 and IV.3 was a heterozygous stop mutation (c.237G > A, p.Trp79\*; (NM\_000401.3); c.138G > A, p.Trp46\*; (NM\_207122.1)) within exon 2 of the *EXT2* gene encoding Exostosin Glycosyltransferase 2, an enzyme involved in the biosynthesis of heparan sulfate which regulates chondrocyte proliferation (Lind et al., 1998). This mutation is absent from the general population variant databases (1000 Genomes Project, ESP6500, ExAC (<http://exac.broadinstitute.org/>)) and was never reported in tumor mutation databases such as Cosmic (<http://cancer.sanger.ac.uk/cosmic>) or TCGA (<http://www.cbioportal.org/>). It is also absent from the MO LOVD database that includes 223 different reported *EXT2* mutations to date ([http://medgen.uantwerpen.be/LOVDv.2.0/home.php?select\\_db=EXT2](http://medgen.uantwerpen.be/LOVDv.2.0/home.php?select_db=EXT2)).

Sanger sequencing confirmed the presence of this truncating mutation in these two patients and in the other affected relatives III.3 and III.4. No DNA was available for patient III.6 but this affected relative was an obligate carrier of the mutation since his daughter (patient IV.4) was shown to carry the mutation. Therefore, within this family, the *EXT2* mutation was present in the 5 relatives who had developed chondrosarcoma. Presymptomatic testing performed in this family revealed that patient III.3's daughter (individual IV.2) had inherited this pathogenic mutation. At the age of 43, she is still asymptomatic and recent full body radiography showed the absence of any cartilaginous tumor (Supp. Fig. S2). The tumoral DNA could be examined for patients II-1 and patient III-3 and surprisingly, the germline *EXT2* mutation could not be detected in these samples suggesting that the mutated allele was either lost by loss of heterozygosity or by reverse mutation (Fig. 6).

## DISCUSSION

The identification in this family of a truncating *EXT2* mutation segregating with solitary chondrosarcoma was unexpected, since germline *EXT2* mutations are commonly associated with MO (Wuyts et al., 1993; Bovee, 2008). If numerous reports have previously highlighted the great phenotypic variability of germline *EXT2* mutations generally less severe than *EXT1* mutations (Porter et al., 2004; Pedrini et al., 2011), no study has reported that *EXT2* germline mutations can result in only solitary chondrosarcoma. The diagnosis of

MO had been ruled out in this family since imaging examinations of affected carriers showed no trace of osteochondroma. Moreover, no other phenotypic features of MO such as skeletal deformity or growth disturbance could be detected in affected members of this family. Nearly all affected individuals with MO are diagnosed by the age of 12 (Wuyts et al., 1993), making it unlikely that osteochondromas were missed in our affected patients with an average age at diagnosis of 35 years. In agreement with the atypical phenotypic expression of the *EXT2* mutation detected in this family, we found that the female carrier IV.2 had no detectable cartilaginous tumor at the age of 43. This observation contrasts with the penetrance of truncating mutations of the *EXT* genes which is close to 100% and might reflect the difference of severity observed between male and female *EXT* mutation carriers (Pedrini et al., 2011). Additionally, examination of the surgical specimen from patient IV.3 by different experienced pathologists excluded the diagnosis of osteochondroma and suggested that this central lesion was an enchondroma, these two types of benign bone tumors being easily distinguishable by a projection of the cortex in continuity with the underlying bone observed in osteochondroma and not in enchondroma. This observation raises the possibility that the CS observed as primary tumors in the 4 affected relatives of this family could correspond to the malignant transformation of enchondromas, which is unusual, since progression of enchondroma to central chondrosarcoma is considered an infrequent event observed in only 4% of cases (Bovee et al., 2010). The distinction between enchondroma and low grade central CS based on histological and radiological examination is a very challenging issue (Murphey et al., 1998). Therefore, although the lesion in patient IV.3 was identified three times as an enchondroma, we cannot exclude the possibility that this lesion was already a very small and well differentiated central CS. Interestingly such atypical central lesions were identified in a small fraction of MO patients with germline *EXT1* or *EXT2* mutations (7 out of 195, 3.6%) (Goud et al., 2015). Of these 7 patients with MO and central lesions, 5 also developed a secondary peripheral CS on a known osteochondroma which represents a higher malignant transformation rate than the one usually associated with MO (1-5%). In the hypothesis of a hereditary form of CS secondary to enchondromas, the malignant transformation rate in our family would also be above the usual one (4/5).

The second atypical feature observed within this family carrying this deleterious *EXT2* mutation was the tumor location corresponding to ribs in 4/5 cases, whereas this location is observed in 4% of osteochondromas, 5% of enchondromas and 9% of central chondrosarcomas (Brien et al., 1997; Porter et al., 2004; Shah et al., 2010; Zarqane et al., 2013). The tumoral DNA could be examined for patients II.1 and patient III.3 and surprisingly, the germline *EXT2* mutation could not be detected in these samples suggesting that the mutated allele was either lost by loss of heterozygosity or by reverse mutation (Fig. 6). These findings are in agreement with recent studies that showed in the context of MO, that osteochondroma cells exhibit *EXT* homozygous inactivation whereas secondary chondrosarcoma initiating cells are *EXT*-expressing cells (de Andrea et al., 2012; Musso et al., 2015).

The atypical phenotypical expression of the germline *EXT2* mutation observed in this family could suggest a specific effect of the mutation. Over 200 pathogenic variants of *EXT2* have been reported to date in the context of MO (MO LOVD database). These mutations are mostly loss of function mutations, frequently located in the first eight exons (Wuyts et al., 1993). Almost a quarter of the *EXT2* mutations reported in the MO LOVD database affect exon 2. The mutation detected in this family also hits exon 2 and is predicted to trigger Nonsense-mediated mRNA Decay (NMD). Because exon 2 is the first coding exon in the predominant transcript (NM\_207122.1), one could speculate that the mutated transcript could escape degradation by NMD by translation reinitiation at a downstream initiation codon but the functional impact of such event remains unclear. The mutation described in our CS family was recently reported elsewhere in a patient presenting a peripheral CS which had progressed from a solitary osteochondroma (Tarpey et al., 2013), but with few clinical data and with no familial history available.

The alternative explanation for the atypical phenotype observed in this family could be the existence of genetic modifying factors. Such factors would need to co-segregate in the 5 affected individuals. No obvious candidate modifier variant could be extracted from the exome sequencing data. In particular, no *TP53* variant reported to be associated with poor prognosis and more aggressive forms of malignancies was identified. The homogeneity of the phenotype (i.e., the absence of MO) observed in the 5 affected mutation carriers might more likely be explained by a genetic modifier variation located in *cis* within or in the



vicinity of the *EXT2* gene, therefore cosegregating with the mutation. Interestingly, exome sequencing data revealed that patients II.1 and IV.3 share two polymorphic variants within intron 1 of *EXT2*. Segregation analysis revealed that these single nucleotide variants (SNVs) cosegregate with the *EXT2* mutation therefore defining a haplotype (data not shown). The first of these variants, rs112082531, with a minor allele frequency (MAF) estimated to 3% in the European non-Finnish population according to the ExAC database, affects position +5 of intron 1 within transcript NM\_000401.3 (c.69 + 5G > A (NM\_000401.3) corresponding to c.-31 + 474G > A (NM\_207122.1)). Bioinformatic predictions indicate that this variant could impact the splicing by strongly weakening the donor splice site of exon 1. The second polymorphism detected (rs11037860) with a MAF estimated to 69% in the European non-Finnish population affects position +27 of the same intron and could slightly strengthen a cryptic donor site at position +23/+24. Therefore, although clearly non pathogenic by itself, the combination of these two SNVs could modify the splicing pattern of this transcript. In the context of the mutated allele, this could subsequently modify the residual amount of functional *EXT2* protein. It is noteworthy that germline biallelic mutations of the *EXT2* gene do not result in MO but in a completely different phenotype corresponding to the Seizures-Scoliosis-Macrocephaly Syndrome, with no signs of MO (Farhan et al., 2015) suggesting that different levels of wild-type *EXT2* protein could determine different phenotypes.

In conclusion, this study shows that germline mutations of *EXT2* can result, not only in the development of multiple benign osteochondromas, but also in the development of isolated malignant cartilaginous tumors including central tumors, and that the presence of germline *EXT2* mutation should be considered in patients suspected to have an inherited predisposition to chondrosarcoma, even in the absence of evidence of MO.

## REFERENCES

- Altay M, Bayrakci K, Yildiz Y, Erekuş S, Saglik Y. 2007. Secondary chondrosarcoma in cartilage bone tumors: Report of 32 patients. *J Orthop Sci* 12:415–423.
- Amary MF, Damato S, Halai D, Eskandarpour M, Berisha F, Bonar F, McCarthy S, Fantin VR, Straley KS, Lobo S, Aston W, Green CL, Gale RE, Tirabosco R, Futreal A, Campbell P, Presneau N, Flanagan AM. 2011. Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations of *IDH1* and *IDH2*. *Nat Genet* 43:1262–1265.
- Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermei P, Belotti M, Gauthier-Villars M, Stoppa-Lyonnet D, Consolino E, Brugieres L, Caron O, Benusiglio PR, Bressac-de Paillerets B, Bonadona V, Bonaiti-Pellie C, Tinat J, Baert-Desurmont S, Frebourg T. 2015. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *J Clin Oncol* 33:2345–2352.
- Bovée JV. 2008. Multiple osteochondromas. *Orphanet J Rare Dis* 3:3.
- Bovée JV, Hogendoorn PC, Wunder JS, Alman BA. 2010. Cartilage tumours and bone development: molecular pathology and possible therapeutic targets. *Nat Rev Cancer* 10:481–488.
- Brien EW, Mirra JM, Kerr R. 1997. Benign and malignant cartilage tumors of bone and joint: their anatomic and theoretical basis with an emphasis on radiology, pathology and clinical biology. I. The intramedullary cartilage tumors. *Skeletal Radiol* 26:325–353.
- de Andrea CE, Reijnders CM, Kroon HM, de Jong D, Hogendoorn PC, Szuhai K, Bovée JV. 2012. Secondary peripheral chondrosarcoma evolving from osteochondroma as a result of outgrowth of cells with functional *EXT*. *Oncogene* 31:1095–1104.
- Farhan SM, Wang J, Robinson JF, Prasad AN, Rupar CA, Siu VM, Hegele RA. 2015. Old gene, new phenotype: mutations in heparan sulfate synthesis enzyme, *EXT2* leads to seizure and developmental disorder, no exostoses. *J Med Genet* 52:666–675.
- Fletcher CDM, Unni KK, Mertens F, editors. 2002. Pathology and Genetics of Tumours of Soft Tissue and Bone. World Health Organization Classification of Tumours. IARC Press: Lyon
- Goud AL, Wuyts W, Bessems J, Brammer J, van der Woude HJ, Ham J. 2015. Intraosseous atypical chondroid tumor or chondrosarcoma grade 1 in patients with multiple osteochondromas. *J Bone Joint Surg Am* 97:24–31.
- Lind T, Tufaro F, McCormick C, Lindahl U, Lidholt K. 1998. The putative tumor suppressors *EXT1* and *EXT2* are glycosyltransferases required for the biosynthesis of heparan sulfate. *J Biol Chem* 273:26265–26268.
- Murphey MD, Flemming DJ, Boyea SR, Bojescul JA, Sweet DE, Temple HT. 1998. Enchondroma versus chondrosarcoma in the appendicular skeleton: Differentiating features. *Radiographics* 18:1213–1237. quiz 1244-5.
- Musso N, Caronia FP, Castorina S, Lo Monte AI, Barresi V, Condorelli DF. 2015. Somatic loss of an *EXT2* gene mutation during malignant progression in a patient with hereditary multiple osteochondromas. *Cancer Genet* 208:62–67.
- Pansuriya TC, van Eijk R, d'Adamo P, van Ruler MA, Kuijjer ML, Oosting J, Cleton-Jansen AM, van Oosterwijk JG, Verbeke SL, Meijer D, van Wezel T, Nord KH, Sangiorgi L, Tokar B, Liegl-Atzwanger B, San-Julian M, Sciort R, Limaye N, Kindblom LG, Daugaard S, Godfraind C, Boon LM, Vikkula M, Kurek KC, Szuhai K, French PJ, Bovée JV. 2011. Somatic mosaic *IDH1* and *IDH2* mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. *Nat Genet* 43:1256–1261.
- Pedrini E, Jennes I, Tremosini M, Milanese A, Mordenti M, Parra A, Sgariglia F, Zuntini M, Campanacci L, Fabbri N, Pignotti E, Wuyts W, Sangiorgi L. 2011. Genotype-phenotype correlation study in 529 patients with multiple hereditary exostoses: identification of “protective” and “risk” factors. *J Bone Joint Surg Am* 93:2294–2302.
- Porter DE, Lonie L, Fraser M, Dobson-Stone C, Porter JR, Monaco AP, Simpson AH. 2004. Severity of disease and risk of malignant change in hereditary multiple exostoses. A genotype-phenotype study. *J Bone Joint Surg Br* 86:1041–1046.
- Shah AA, D'Amico TA. 2010. Primary chest wall tumors. *J Am Coll Surg* 210:360–366.
- Tarpey PS, Behjati S, Cooke SL, Van Loo P, Wedge DC, Pillay N, Marshall J, O'Meara S, Davies H, Nik-Zainal S, Beare D, Butler A, Gamble J, Hardy C, Hinton J, Jia MM, Jayakumar A, Jones D, Latimer C, Maddison M, Martin S, McLaren S, Menzies A, Mudie L, Raine K, Teague JW, Tubio JM, Halai D, Tirabosco R, Amary F, Campbell PJ, Stratton MR, Flanagan AM, Futreal PA. 2013. Frequent mutation of the major cartilage collagen gene *COL2A1* in chondrosarcoma. *Nat Genet* 45:923–926.
- Wuyts W, Schmale GA, Chansky HA, Raskind WH. 1993. Hereditary Multiple Osteochondromas. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2015. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK1235/>.
- Zarqane H, Viala P, Dallaudiere B, Vernhet H, Cyteval C, Larbi A. 2013. Tumors of the rib. *Diagn Interv Imaging* 94:1095–1108.

### III. Discussion

Sur le plan scientifique, la découverte de la co-ségrégation d'une mutation tronquante de l'exon 2 du gène *EXT2* avec une forme atypique de chondrosarcomes familiaux était un événement surprenant. *EXT2* est un gène suppresseur de tumeurs jouant un rôle dans la prolifération des chondrocytes. Les mutations perte de fonction d'*EXT2* documentées à l'état hétérozygote sont à l'origine de la maladie des Ostéochondromes Multiples (OM), des lésions cartilagineuses bénignes ne dégénérant en CS périphériques que dans moins de 5 % des cas. Alors qu'aucun argument clinique, radiologique ou anatomopathologique n'était en faveur d'une OM dans cette famille, nous avons identifié une mutation perte de fonction du gène *EXT2* co-ségrégant avec la pathologie. Ces résultats laissent à penser que le phénotype associé aux mutations perte de fonction du gène *EXT2* ne se limite pas aux tumeurs bénignes périphériques mais inclut également des tumeurs malignes dont certaines tumeurs centrales. Les mutations du gène *EXT2* doivent donc être recherchées dans les formes familiales de chondrosarcomes y compris en l'absence de signe clinique d'Ostéochondromes Multiples.

Il est frappant de noter que la mutation détectée dans cette famille ne semble pas différente des autres mutations perte de fonction de l'exon 2 du gène *EXT2* détectée chez les patients atteints de OM. Pourtant le phénotype observé dans notre famille est très différent de celui observé chez les patients OM. L'homogénéité du phénotype observé chez les cinq individus de cette famille pourrait être expliquée par la présence d'un modificateur génétique situé en *cis*, dans ou à proximité du gène d'*EXT2*, et donc co-ségrégant avec la mutation. Nous avons détecté plusieurs variants du gène *EXT2* formant un haplotype avec la mutation stop :

- Le premier variant touche la position +5 du site donneur de l'exon 1 c.69+5G>A (MAF 3 %). La prédiction bioinformatique de l'effet de ce variant est un affaiblissement du site donneur de l'exon 1.

- Le deuxième variant est une variation fréquente (MAF 70 %) en position 27 de l'exon 1 (c.69+27T>G). Les prédictions bioinformatique indiquent que ce variant pourrait renforcer légèrement un site donneur cryptique à la position +23/24.

Bien que non pathogènes en eux-mêmes, ces variants en combinaison pourraient modifier le profil d'épissage et *in fine* la quantité de protéine *EXT2* disponible. Il est intéressant de noter que l'inactivation biallélique constitutionnelle du gène *EXT2* entraîne un syndrome de Convulsions-Scoliose-Macrocéphalie sans signe de OM, ce qui suggère que différents niveaux de protéine *EXT2* pourraient donc déterminer différents phénotypes.

Nos résultats démontrent la puissance de la stratégie d'analyse exomique intrafamiliale pour identifier les bases moléculaires à l'origine de la prédisposition observée dans une famille. À partir d'un nombre restreint d'exomes (n=2), nous avons pu mettre en évidence un nombre réduit de variants potentiellement impliqués dans le phénotype. Ces résultats démontrent l'intérêt d'une approche exhaustive sans *a priori* telle que l'exome et démontre l'efficacité de la stratégie exomique intrafamiliale. En outre, ces résultats soutiennent l'hypothèse qu'un nombre limité de gènes participerait aux mécanismes d'oncogenèse : à un gène altéré peut correspondre plusieurs phénotypes.

Néanmoins, ce type de présentation familiale atypique est exceptionnelle et cela constitue une limite importante à l'analyse exomique intra-familiale.

**Projet 2 : Identification de mutations *de novo* par analyse exomique de trio chez un jeune patient ayant développé sporadiquement un médulloblastome et un méningiome**

## I. Contexte scientifique de l'étude

Entre les cancers familiaux associés à un âge de survenue précoce et l'ensemble des cancers dits « sporadiques » qui résultent de l'accumulation de mutations somatiques et qui sont associés à un âge de survenue tardif, il existe des patients développant de façon très précoce une ou plusieurs tumeurs en l'absence de contexte familial et sans altération dans les gènes de prédisposition précédemment connus. De par leur précocité et leur sévérité, ces cas sporadiques sont très fortement évocateurs d'une altération génétique constitutionnelle dans un gène de prédisposition au cancer. L'hypothèse que nous avons développée au laboratoire est que ces cas pourraient résulter de mutations *de novo*. En effet, grâce à l'essor du séquençage de nouvelle génération, un taux élevé de mutations *de novo* a été mis en évidence chez l'homme. Le taux de mutations *de novo* est estimé à  $1,8 \times 10^{-8}$  par nucléotide et par génome, représentant en moyenne 1,58 par exome <sup>138</sup>.

Ainsi, dans le but d'identifier des mutations *de novo* potentiellement responsables de la prédisposition génétique au cancer observée chez ces cas sporadiques, nous avons utilisé la stratégie décrite par Vissers <sup>72</sup> basée sur l'analyse d'exome différentielle à partir de trios cas index/parents sains. Pour rappel, cette stratégie est basée sur la sélection d'un individu présentant un phénotype extrême de survenue très précoce suggérant une cause génétique, avec des parents indemnes de la maladie (présentation sporadique). Ensuite, le séquençage de l'ensemble de l'exome chez ces trois individus constituant le « trio » est réalisé. Grâce à la bio-informatique, l'ensemble des variants détectés chez le père et chez la mère sont soustraits aux variants détectés chez l'individu atteint. Ainsi, dans l'analyse, les variants hérités sont éliminés laissant place uniquement aux variants non hérités et propres à l'individu ou mutations *de novo*.

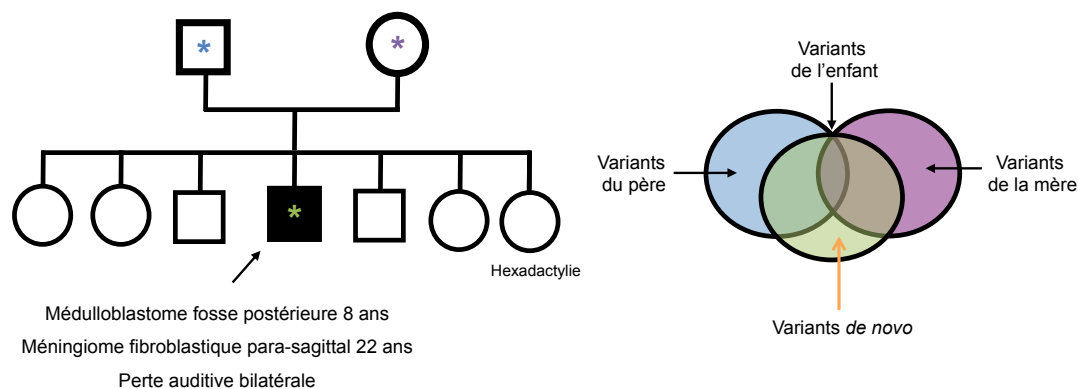
Étant donné que cette approche nous a permis en 2014 de mettre en évidence un nouveau gène candidat de prédisposition au cancer de l'ovaire <sup>123</sup>, nous avons décidé de transposer cet outil aux tumeurs pédiatriques afin d'identifier de nouveaux gènes de prédisposition génétique aux cancers de l'enfant. En effet, ce type de tumeurs sont de par l'âge de survenue particulièrement évocatrices d'un accident constitutionnel. L'identification de mutations *de novo* impliquées dans le développement précoce de tumeurs pourrait permettre de mettre en évidence de nouveaux gènes ou voies biologiques impliquées dans les tumeurs pédiatriques. Ainsi, nous avons appliqué la stratégie d'analyse exomique de trios enfant malade-parents sains à un jeune patient qui a développé de façon sporadique, un médulloblastome à l'âge de 8 ans suivi d'un méningiome à l'âge de 22 ans. Le tableau clinique de ce jeune patient se complétait par une perte auditive bilatérale considérée comme une séquelle du traitement du médulloblastome. L'Organisation Mondiale de la

Santé reconnaît cinq sous-groupes histologiques pour le médulloblastome : la forme dite classique (70 % des cas), la forme desmoplasique de meilleur pronostic (15 % des cas), la forme à nodularité extensive, la forme anaplasique et la forme dite à grandes cellules. Il est également possible de classer le médulloblastome selon quatre événements moléculaires somatiques : avec activation de WNT (10 % des cas), avec activation de la voie de signalisation Sonic Hedgehog SHH (30 % des cas), avec amplification du gène *MYC* (25 % des cas), avec gain de la région 17q (35 % des cas). Malheureusement, nous ne disposons pas de ces informations pour ce patient étant donné que le médulloblastome a été développé en 1996 et qu'aucun matériel n'a été congelé à ce moment là.

## II. Matériels et Méthodes

### Patient analysé

Cette étude a porté sur un patient ayant développé un médulloblastome de la fosse postérieure à l'âge de 8 ans ainsi qu'un méningiome fibroblastique para-sagittal à 22 ans, accompagnés d'une perte progressive de l'audition attribuée à son traitement par radiothérapie sans antécédent familiaux de cancers (Figure 33). Le médulloblastome est la tumeur embryonnaire la plus fréquente chez l'enfant. Il représente environ 20 % des tumeurs cérébrales pédiatriques. Cette tumeur cérébrale d'origine neuro-ectodermique primitive se développe généralement dans la fosse cérébrale postérieure au niveau du cervelet.

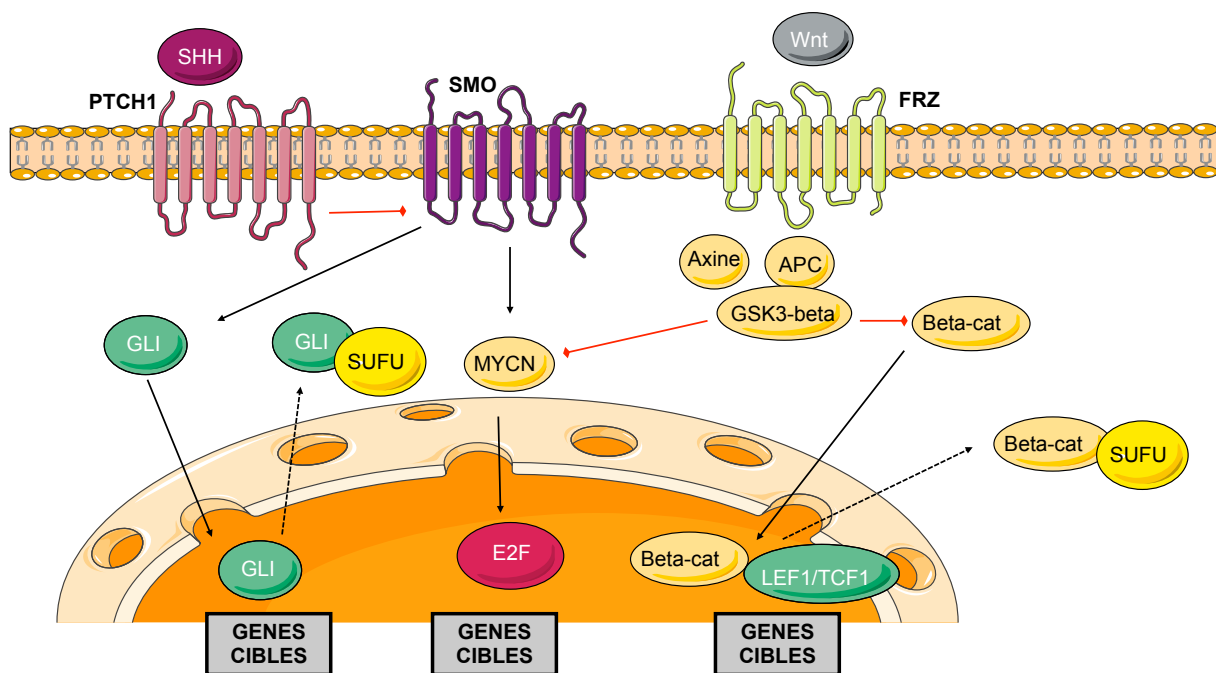


**Figure 33** : Arbre génétique de la famille analysée et la stratégie exomique employée dans cette famille

Etant donné le contexte sporadique dans lequel le cas index (flèche noire) a développé un médulloblastome à 8 ans ainsi qu'un méningiome à 22 ans, nous avons employé une stratégie exomique soustractive. Les variants des parents identifiés par exome (schématiquement en bleu et violet) ont été soustraits à ceux détectés chez l'enfant (schématiquement en vert) afin d'identifier les variants non hérités potentiellement responsables de la pathologie observée chez ce jeune patient.

D'un point de vue moléculaire, le médulloblastome peut être la résultante d'altérations constitutionnelles dans les gènes *TP53* (associé au syndrome de Li-Fraumeni), *SUFU* et *PTCH1* (dans le contexte d'un syndrome de Gorlin)<sup>139</sup> (Figure 34). Ce patient avait bénéficié

d'un criblage moléculaire des gènes *TP53* et *SUFU* qui se sont avérés négatifs. Nous ne possédons pas l'information concernant l'analyse du gène *PTCH1* dans cette famille, cependant, ce patient ne présentait pas les manifestations cliniques d'un syndrome de Gorlin comme la présence de naevi basocellulaires, ou une dysmorphie cranio-faciale. Cette information a été vérifiée à postériori sur les fichiers BAM après le séquençage exomique.



**Figure 34** : La voie biologique Sonic Hedgehog et Wingless

La signalisation SHH est initiée lorsque SHH se lie à son récepteur PTCH1, ceci libère SMO de l'inhibition. Cette libération entraîne l'activation de GLI et MYCN. Lorsque WNT se lie à son récepteur FRZ, le complexe APC est déstabilisé, libérant la bêta-caténine. Cette dernière peut alors entrer dans le noyau et activer les facteurs de transcription en aval. APC, *Adenomatous Polyposis Coli*; Beta-cat, Beta-caténine; FRZ, *Frizzled*; GSK3beta, *Glycogene Synthase Kinase 3 bêta*; LEF1 / TCF1, Facteur Activateur Lymphoïde / facteur spécifique de cellules T1; PTCH1, *Patch 1*; SHH, *Sonic hedgehog*; SMO, *Smoothened*; SUFU, Suppresseur de fusion; WNT, *Wingless*.

### **Séquençage exomique et filtration des variants**

Les bibliothèques d'ADN ont été réalisées à partir de 3µg d'ADN génomique extrait à partir du sang périphérique, grâce au kit Agilent SureSelect XT. L'enrichissement des régions exoniques a été réalisé à partir de la version V4 du kit SureSelect All Exons. Les bibliothèques ont ensuite été soumises à un séquençage *paired-end* de 2x75pb sur le séquenceur Illumina GAIIx. Le séquençage de ces trois exomes a permis d'obtenir une profondeur moyenne de 60x. 95 % des bases étaient au minimum lues 10 fois et 93 % des bases lues possédaient un Qscore supérieur à 30. L'alignement et le *variant calling* ont été réalisés à l'aide du logiciel CASAVA (version 1.8.2). L'annotation a été réalisée à l'aide du logiciel Annovar. Les 20 009 variants identifiés chez le cas index ont ensuite été filtrés à l'aide du logiciel EVA (*Exome Variation Analyzer*)<sup>140</sup> en éliminant les variants présents chez les parents (variants

hérités), les variants présents dans les bases de données 1000 GENOMES PROJECT et ESP à une fréquence supérieure à 0,1 %, les variants présents dans une base interne au laboratoire de 72 exomes, les variants présents à une faible profondeur (< 10 x) et de mauvaise qualité (Qscore < 30) et les variants ayant une balance allélique inférieure à 20 %.

Cette filtration a permis d'identifier 309 variants potentiellement *de novo* parmi les 20 009 séquencés chez le cas index. L'analyse a ensuite été complétée par un curage des fichiers BAM consistant principalement à éliminer les artefacts de séquençage comme des artefacts de biais de brin non éliminés par les versions logicielles utilisées au moment de l'étude. À la fin de ces étapes, deux variants *de novo* ont été retenus. Le caractère *de novo* de ces deux variations a été ensuite confirmé par séquençage Sanger.

### **Séquençage Sanger du gène *HID-1***

Les 19 exons du gène *HID-1* ont été amplifiés par PCR à partir de 50 à 100 ng d'ADN génomique extrait à partir du sang périphérique. Le séquençage Sanger de ces fragments a été réalisé à l'aide du kit BigDye® Terminator v3.1 Cycle sequencing (*Life Technologies*). Les réactions de séquence ont été séparées par électrophorèse sur le séquenceur 3100xl Genetic Analyzer (*Life Technologies*) et analysées à l'aide du logiciel *Sequencing analysis software 5.2.0 (Life Technologies)*.

### **Quantitative Multiplex PCR of Short Fluorescent Fragments ou QMPSF**

La recherche de réarrangements tels que des délétions ou duplications d'exons a été réalisée grâce à une analyse ciblée : la QMPSF (*Quantitative Multiplex PCR of Short Fluorescent fragments*). Cette technique consiste à amplifier de manière simultanée des courts fragments d'ADN de tailles différentes en utilisant des amorces fluorescentes (ajout d'un groupement 6-FAM en 5' de l'amorce sens). L'amplification se fait avec un nombre limité de cycles de PCR (ici, n=26) permettant de rester dans la phase exponentielle de la PCR. La séparation des produits de PCR se fait selon la taille par électrophorèse sur séquenceur automatisé. L'analyse comparative des données est réalisée par superposition informatique des profils de fluorescence obtenus pour un patient et un individu témoin, normalisés sur des amplicons contrôles présents en deux copies chez les témoins comme chez les patients. Une délétion hétérozygote sera visualisée par une diminution de 50 % de la hauteur du pic et une duplication hétérozygote sera visualisée par une augmentation de 50 % de la hauteur du pic.

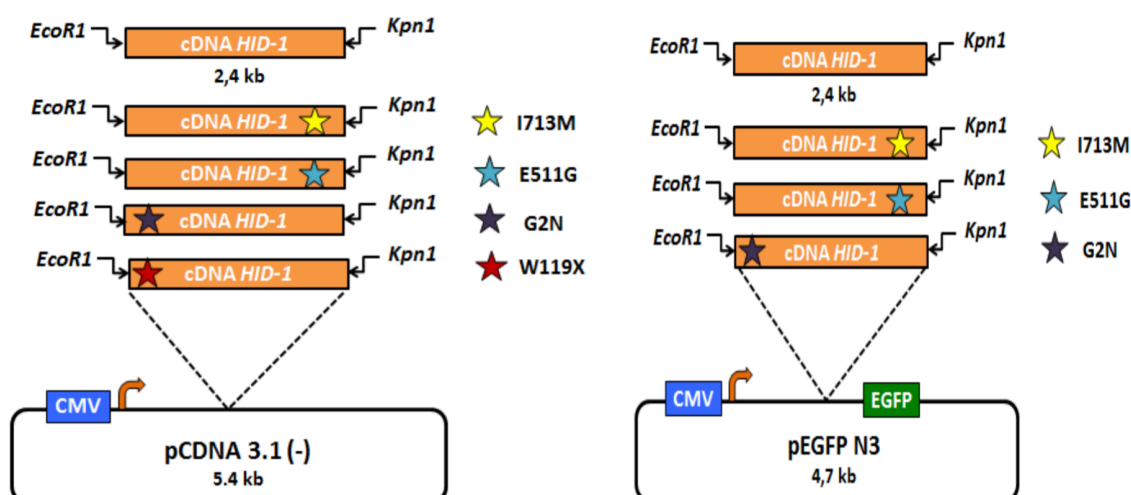
Les QMPSF ont été réalisées à partir de 50 ng d'ADN génomique. Chaque QMPSF a été réalisée avec deux amplicons contrôles (*HPRT* et *HMBS*). Pour chaque amplicon, nous utilisons un mélange équimolaire (10 µM) d'amorces sens et antisens. La concentration de chaque amorce a été ajustée pour obtenir des hauteurs de pics comparables. 1 µL de



produit de cette PCR multiplexe est ensuite mélangé avec 10 µL de formamide et 0,4 µL de marqueur de taille (*GeneScan TM 400HD standard ROX*, Life Technologies TM). Ce mélange est dénaturé pendant deux minutes à 95°C avant d'être séparé par électrophorèse sur séquenceur automatisé (*ABI Prism 3100 Genetic Analyser*, Life technologies TM). Le profil de fluorescence obtenu pour un patient est superposé de façon informatique avec celui d'un individu témoin. À noter que l'amplicon contrôle *HPRT* est situé sur le chromosome X, permettant de confirmer le sexe du patient.

### **Construction des plasmides et transfection cellulaire**

Cent nanogrammes issus d'une banque d'ARN de cerveau humain ont été rétrotranscrits grâce au kit Verso cDNA (Thermo-Scientific). L'ADNc en entier codant pour *HID-1* sauvage a été ensuite amplifié par PCR grâce à la *Taq Phusion High Fidelity DNA Polymerase* (Thermo-Scientific) et à des amorces spécifiques de l'ARNm du gène *HID-1*. Après digestion par les enzymes de restriction *EcoRI* et *KpnI*, l'ADNc d'*HID-1* a été inséré dans deux plasmides pcDNA3.1(-) et pEGFP-N3 au moyen de la T4 DNA ligase (New England Biolabs). Les plasmides ainsi générés ont été amplifiés dans des bactéries compétentes XL1B. Les bactéries résistantes à l'antibiotique (50 µg/ml de carbénicilline pour pcDNA3.1(-) et 50 µg/ml de kanamycine pour le pEGFP-N3) ont ensuite été repiquées dans du milieu LB liquide pendant 16h à 37°C. Ensuite, chaque plasmide été purifié grâce au kit d'extraction Nucleospin Plasmid (Macherey-Nagel) selon les recommandations du fabricant. Les mutations identifiées (I713M, E511G) ainsi que des mutations contrôles artificielles (G2N, W119X) ont été introduites dans chacun des deux plasmides à l'aide du kit *QuickChange II XL* de mutagenèse dirigée (Figure 35). Les plasmides des formes sauvages et mutantes du gène *HID-1* ont été transitoirement transfectés dans des cellules SH-SY5Y et dans des PC12 à l'aide de la Lipofectamine 2000 (Invitrogen) selon les instructions du fabricant.



**Figure 35** : Schéma des différentes constructions plasmidiques utilisées dans ce projet

Dans le but d'étudier l'impact des mutations détectées dans le gène *HID-1* chez des patients présentant un médulloblastome, différentes constructions plasmidiques ont été réalisées. Pour cela, le cDNA du gène *HID-1* (2,4kb) a été amplifié par RT-PCR sur ARN totaux de cerveau humain. Il a été cloné dans les vecteurs pCDNA3.1(-) et le pEGFP-N3 en utilisant les sites de restriction *EcoRI* et *KpnI*. Le vecteur pEGFP-N3 permet d'exprimer une protéine de fusion entre une protéine d'intérêt et la EGFP (*Enhanced Green Fluorescent Protein*) en position C-terminal. Les différentes mutations exoniques détectées précédemment (I713M et E511G) ainsi que les mutations contrôles (G2N et W119X) ont été introduites par mutagenèse dirigée. CMV = promoteur du CytoMégaloVirus; EGFP = *Enhanced Green Fluorescent Protein*

### **Extraction protéique et Western Blot**

Les cellules transfectées ont été récoltées par traitement à la trypsine-EDTA (Invitrogen) puis culotées. Les protéines ont été ensuite extraites à l'aide du tampon d'extraction RIPA (50 mM Tris-HCl pH8, 150 mM NaCl, 20 mM EDTA, 1% NP-40) additionné d'un cocktail d'inhibiteurs de protéases 1 % (Sigma-Aldrich, St. Louis, MO, USA) et d'un cocktail d'inhibiteurs de phosphatases 1 % (Halt phosphatase, ThermoScientific, Waltham, MA, USA). Le dosage des protéines a été réalisé par la méthode de Lowry (dosage colorimétrique) à l'aide du kit DC Protein Assay (Bio-Rad laboratories, Hercules, CA, USA), selon les recommandations du fabricant. De 20 à 15 µg de protéines ont été déposées sur un gel de polyacrylamide 10 % (TGX Stain-Free™ FastCast™ Acrylamide Solutions, Bio-Rad) en conditions dénaturantes pour une électrophorèse. Après transfert, les membranes ont été saturées avec une solution PBS-Tween 0,05 % contenant 2,5 % de BSA (Bovine Serum Albumin) pendant 1h à température ambiante. Elles ont été ensuite incubées une nuit à 4°C en présence de l'anticorps primaire Anti-*HID-1* monoclonal (Novus Biological – NBP2-02667 ; dilution 1/1000ème) dirigé contre la protéine d'intérêt *HID-1* ou l'anticorps primaire Anti-néomycine polyclonal (Anti-Neomycin Phosphotransferase II, Millipore, 06-747 ; dilution 1/1000ème) dirigé contre la protéine de néomycine. Les deux anticorps ont été dilués en

PBT-BSA 2,5 % (Phosphate Buffer Saline 0,05% Tween-BSA). Après 3 lavages de 15 minutes en PBT à température ambiante, les membranes sont incubées 1 heure à température ambiante avec un anticorps secondaire couplé à la *Horse Radish Peroxydase* (HRP) et dirigé contre les immunoglobulines (Ig) des deux anticorps primaires précédemment utilisés. Après 3 nouveaux lavages en PBT, une acquisition numérique du signal fluorescent (Trihalo) présent sur la membrane est réalisée (UV, GelDoc™ EZ system, Bio-Rad). Cette dernière permettra de normaliser le signal de western blot par rapport à la quantité de protéine totale présente dans chaque piste (ImageLab, Bio-Rad). Le signal correspondant à la protéine d'intérêt a finalement été obtenu par chemiluminescence avec le kit ECL (Pierce, Rockford, IL, USA). L'acquisition de l'image a été effectuée au moyen du système G-Box (Syngene, Cambridge, UK) grâce au logiciel GeneSnap (Syngene) et les quantifications sont réalisées avec le logiciel GeneTools (Syngene).

### **Immunocytochimie et microscopie confocale**

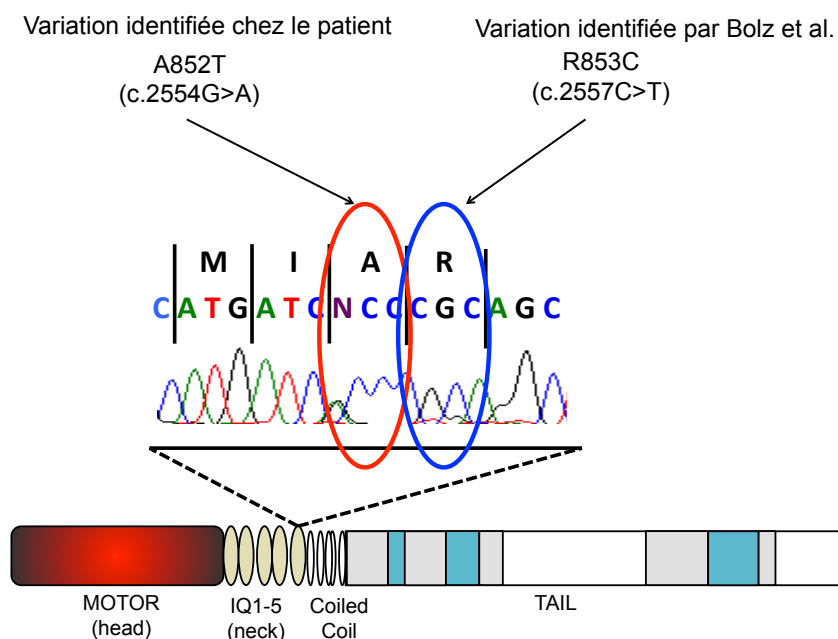
24 heures après transfection, les cellules ont été rincées au PBS 1X, puis fixées au paraformaldehyde (PFA) 4 %, pendant 15 minutes. Les cellules sont ensuite rincées au PBS 1X et incubées 30 minutes avec le tampon de saturation contenant du PBS, NDS (*Normal Donkey Serum*) au 1/50ème, Triton X-100 0,3 % afin de perméabiliser les membranes et réduire le marquage non spécifique des anticorps. Les cellules non transfectées ont ensuite été incubées 2h à température ambiante avec l'anticorps primaire anti-HID1 généré chez la souris (Anti-HID-1 1 monoclonal, Novus Biological – NBP2-02667). Après un rinçage au PBS, les cellules sont incubées avec l'anticorps secondaire anti-IgG de souris généré chez l'âne (Donkey Anti-Mouse, DAM) et couplé à un fluorochrome (DAM-Alexa 488, Promega). Chaque anticorps est dilué dans du tampon de blocage, et incubé 1 heure à l'obscurité. Les cellules transfectées avec les plasmides pEGFP-N3-HID-1 ont été incubées 2 heures à température ambiante avec l'anticorps primaire TGN46 ciblant la protéine TGN46, un marqueur du réseau trans-Golgien humain. Après un rinçage au PBS, les cellules ont été incubées avec l'anticorps secondaire anti-IgG de mouton généré chez l'âne (Donkey Anti-Sheep, DAS) et couplé à un fluorochrome (DAS-Alexa 594). L'anticorps est dilué dans du tampon de blocage, et incubé 1 heure à l'obscurité. Les cellules sont ensuite rincées au PBS 1X. Le marquage du noyau se fait à l'aide du DAPI (Sigma), incubé 30 secondes à température ambiante. Les lamelles de verre ont été ensuite rincées à l'eau distillée et montées sur lame avec du mowiol (Sigma), un milieu de montage spécifiquement adapté à l'immunofluorescence. Les lames ont été ensuite observées avec un microscope confocal droit (Leica SP2, plate-forme PRIMACEN, Mont-Saint Aignan). Ce microscope confocal est équipé d'objectifs à immersion à huile (40X, 63X), ainsi que de lasers permettant d'exciter la GFP, et l'Alexa 488, et d'un laser UV pour exciter le DAPI.

### III. Résultats

#### ❖ Séquençage exomique du trio enfant malade-parents sains

Afin d'identifier les variations *de novo* présentes chez le cas index, nous avons réalisé le séquençage de son exome ainsi que celui de ses parents. A l'aide du logiciel EVA<sup>140</sup>, l'ensemble des variants hérités détectés chez les parents ont été soustraits à ceux détectés chez le cas index pour ne conserver que les variants *de novo*. Cette analyse a permis de révéler l'existence de deux variants *de novo* hétérozygotes non répertoriés en population générale (ESP, 1000GENOMES, ExAC) parmi les 20 009 variants présents chez le cas index.

La première, impactant le gène *MYO7A* (NM\_000260.3; c.2554G>A; p.Ala852Thr) est susceptible de contribuer à la perte de l'audition également observée chez ce patient. En effet, le gène *MYO7A* code pour la myosine 7A, une myosine nécessaire à la transformation mécanique du son en signal moléculaire. Nous avons identifié chez ce patient une mutation faux-sens de ce gène à l'état hétérozygote, non répertoriée en population générale et prédite délétère par 3/3 des logiciels de prédiction de pathogénicité (SIFT, PolyPhen2, MutationTaster). Cette altération est située dans l'un des domaines de liaison à la calmoduline de la myosine 7A, nécessaire à son activité. L'équipe allemande de Bolz et ses collaborateurs a décrit une altération de ce gène située dans le même domaine de liaison à la calmoduline, à un codon de l'altération identifiée chez notre patient<sup>141</sup> (Figure 36). Cette altération invalide la fonction et cause une forme particulière autosomique dominante de surdité (*DeaFNess Autosomal dominant 11*; DFNA11). Ainsi, cette altération est très probablement responsable de la perte de l'audition observée chez ce patient.

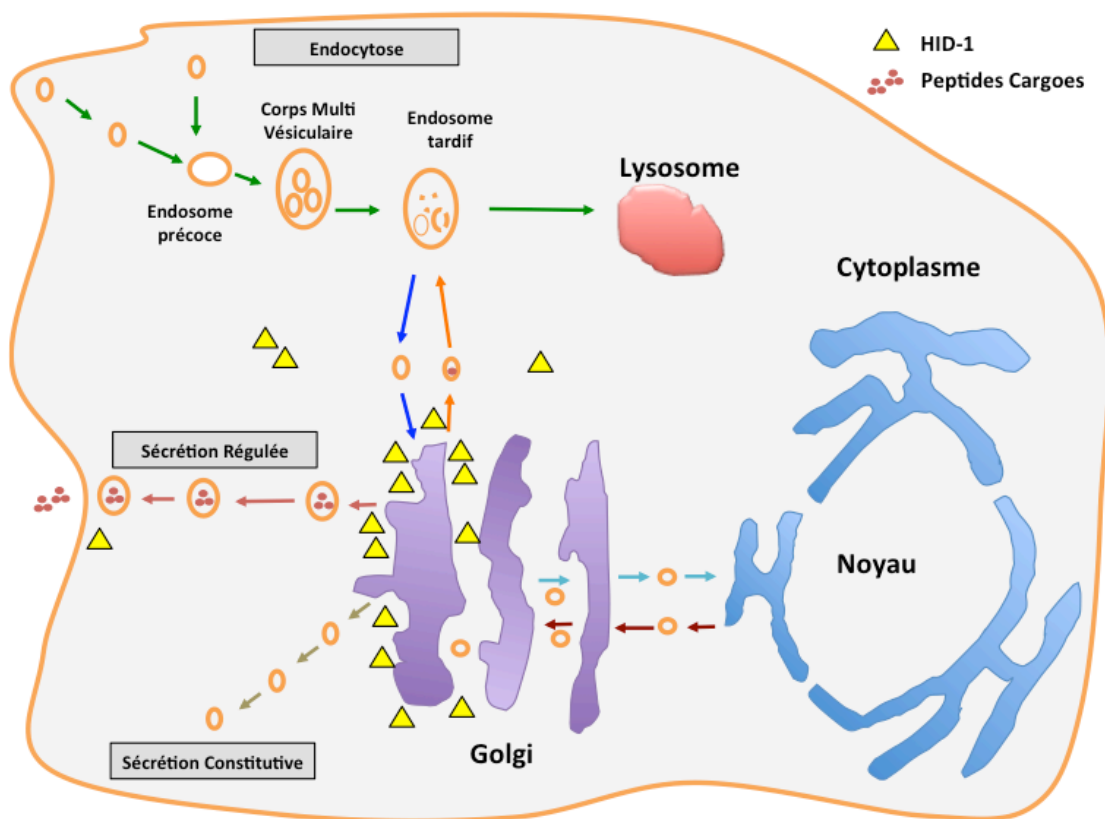


**Figure 36** : Position du variant de novo dans le gène *MYO7A* identifié chez le patient analysé

Grâce au séquençage exomique et à une stratégie exomique soustractive, nous avons identifié une altération dans le gène *MYO7A* (NM\_000260.3; c.2554G>A; p.Ala852Thr). Cette altération impacte un site de liaison à la calmoduline située dans le domaine IQ de la protéine. Bolz et ses collaborateurs ont décrit une altération située à un codon de l'altération que nous avons détectée et ont démontré son implication dans une forme autosomique à transmission dominante de perte de l'audition (DNFA11).

Néanmoins, il est intéressant de noter que les altérations bi-alléliques du gène *MYO7A* sont également impliquées dans le syndrome de Usher. Ce syndrome de transmission autosomique récessive entraîne une perte de l'audition ainsi qu'une cécité progressive. Or, ici, ce patient ne présente pas de cécité. De plus, j'ai pu vérifier sur les fichiers BAM l'absence de second événement dans le gène *MYO7A*. J'ai par ailleurs vérifié l'absence de mutation dans les autres gènes décrits comme étant impliqués dans le syndrome de Usher. Ainsi, il est fort probable que la mutation *de novo* dans le gène *MYO7A* puisse expliquer la surdité observée chez ce patient qui avait préalablement attribuée au traitement du médulloblastome par radiothérapie. L'identification de cette altération sera potentiellement d'intérêt pour la prise en charge médicale et le conseil génétique de ce patient.

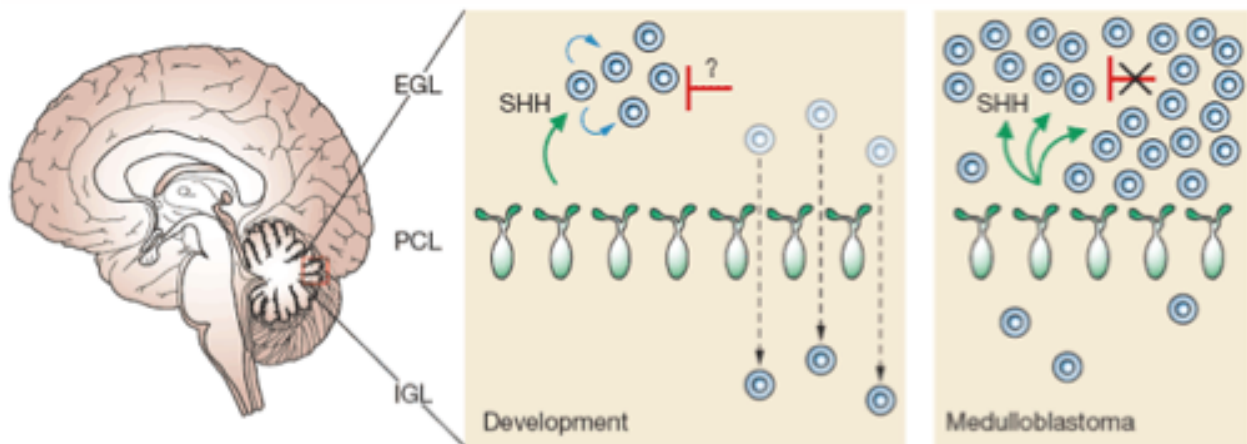
La seconde variation *de novo* identifiée est une mutation faux sens touchant le gène *HID-1* (*High-Temperature-Induced Dauer Formation*) (c.2139T>G ; p.Ile713Met ; NM\_030630.2). Cette mutation non répertoriée en population générale est prédite délétère par 3/3 des logiciels de prédictions de pathogénicité. Le gène *HID-1*, anciennement nommé *DMC1* (*Down-regulated In Multiple Cancer-1*)<sup>142</sup>, a été identifié à l'origine par Aillion et Thomas chez le ver *Caenorhabditis elegans*<sup>143</sup>. Ce gène code pour une protéine qui régule la décision d'entrer dans un stade larvaire appelé « dauer ». Ce gène est hautement conservé de *C.elegans* à l'Homme. Cependant, sa fonction cellulaire exacte chez l'humain n'est pas claire. Il a été démontré que *HID-1* est fortement exprimé dans les neurones ainsi que dans le tissu digestif et dans les cellules sécrétrices<sup>143-146</sup>. La protéine HID-1 codée par le gène *HID-1*, serait impliquée dans l'exocytose vésiculaire des granules de sécrétion à cœur dense ou DCV (*Dense Core Vesicles*). Cette libération du contenu des DCV est une sécrétion régulée c'est-à-dire déclenchée par un signal provoquant un influx calcique. Elle diffère de la sécrétion constitutive qui sécrète en continu. De plus, le contenu des vésicules de ces deux types de sécrétion est différent. Les produits de la voie régulée sont plutôt peptidiques. Ces peptides appelés cargos sont triés et rassemblés au niveau du réseau *trans*-Golgien (TGN) pour être stockés dans les DCV. Sous l'effet du signal, il y a déclenchement d'une cascade d'événements intracellulaires (augmentation de la concentration cytosolique de calcium) entraînant la libération du contenu de ces DCV par fusion avec la membrane plasmique. La protéine HID-1 est majoritairement péri-golgienne et est retrouvée au niveau de la membrane du TGN grâce à un motif de N-myristoylation<sup>144</sup> (Figure 37).



**Figure 37** : Schéma représentant la localisation de la protéine HID-1 au sein du trafic vésiculaire cellulaire

La protéine HID-1 (ici représentée par un triangle jaune) est majoritairement péri-golgienne (autour du *trans*-Golgi). Elle est également présente au niveau de la membrane cytoplasmique. Chez *C.elegans*, elle est impliquée dans la sécrétion vésiculaire régulée.

Malgré le faible nombre de publications relatives à la fonction du gène *HID-1*, ce rôle supposé serait cohérent avec une implication dans la genèse du médulloblastome puisque certains médulloblastomes se développent à partir des précurseurs neuronaux du cervelet fortement sécrétants. En effet, physiologiquement, les précurseurs cellulaires granulaires (*Granule Cell Precursor; GCP*) reçoivent un cocktail moléculaire incluant la protéine SHH (*Sonic HedgeHog*), leur permettant de s'expandre dans la couche granulaire externe (*External Granule Layer; EGL*). Lorsque le signal SHH s'arrête, les GCP entament alors une maturation et une migration à travers la couche cellulaire de Purkinje vers la couche interne granulaire (*Internal Granule Layer; IGL*). Ni la nature, ni l'origine cellulaire de cette abolition du signal SHH n'est aujourd'hui connue. Cependant, il semble que ce signal ne soit plus aboli chez les patients développant un médulloblastome<sup>139,147</sup> (Figure 38).



**Figure 38** : Schéma illustrant la communication moléculaire fine entre les différentes couches cellulaires du cervelet médiée par la molécule SHH pendant le développement et dans le contexte du médulloblastome

Pendant le développement, les précurseurs cellulaires granulaires (GCP) subissent une prolifération massive dans la couche granulaire externe après avoir reçu le signal SHH de la cellule de Purkinje. Par la suite, les GCP vont se différencier et migrer vers le bas pour former l'IGL (couche de granule interne). Dans le cas d'un médulloblastome, les cellules de granules dérégulées vont proliférer notamment à cause de l'absence de signaux d'arrêt de prolifération. EGL, Couche Granulaire Externe; GCP, Précurseur de Cellules Granulaires; IGL, Couche Granulaire Interne; SHH, *Sonic Hedgehog*; PCL, Couche Cellulaire de Purkinje.

Ainsi, il semble exister une communication cellulaire fine entre ces différents tissus cérébelleux qui participent au développement normal du cervelet. Cette communication altérée participerait au développement d'un médulloblastome. L'hypothèse que nous formulons ici est que les altérations identifiées du gène *HID-1* modifieraient cette homéostasie moléculaire en perturbant l'adressage de molécules normalement sécrétées ou en modifiant la présence de récepteurs membranaires à la surface cellulaire, conduisant alors à une dérégulation de la communication de ces précurseurs neuronaux. Cette hypothèse est corroborée par le fait que les mutations du gène *PTCH1* entraînent le développement d'un médulloblastome, car empêchant la reconnaissance ou la transduction du signal SHH et donc la migration physiologique des GCP (Figure 38). A ce titre, le gène *HID-1* représente alors un nouveau gène candidat.

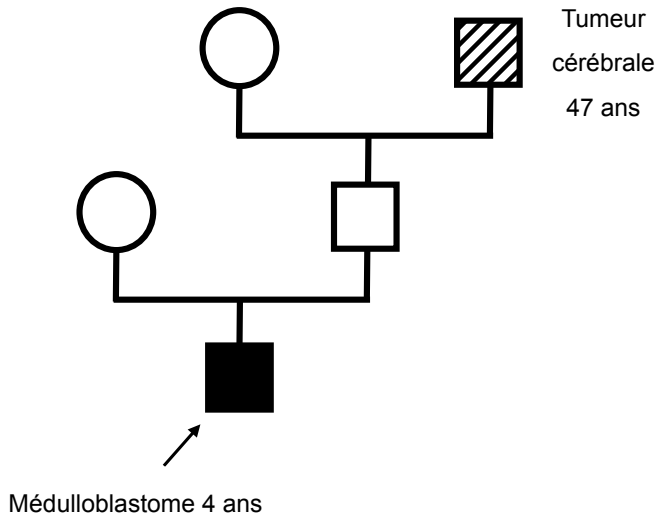
L'identification d'un variant *de novo* chez un patient est un élément fort dans l'implication d'un gène dans une pathologie mais ne permet pas à elle seule de démontrer son implication. Cette démonstration passe par un faisceau d'arguments comprenant des arguments génétiques : (i) la récurrence de mutations du même gène chez des patients présentant un phénotype similaire, (ii) la co-ségrégation d'une altération de ce gène dans une autre famille présentant le même phénotype, (iii) l'enrichissement en variants rares de ce gène de la cohorte de patients par rapport à une cohorte d'individus témoins ; et des arguments fonctionnels : (i) la prédiction *in silico* de l'effet délétère des mutations (ii) la démonstration de l'impact fonctionnel des mutations dans des modèles *in vitro* ou *ex vivo*,

(iii) l'observation du phénotype dans un modèle animal portant l'altération.

❖ Recherche de mutations ponctuelles et de réarrangements du gène *HID-1* dans une cohorte de réplication

Pour vérifier l'hypothèse d'une implication des mutations constitutionnelles du gène *HID-1* dans la genèse du médulloblastome, j'ai entrepris, avec l'aide de Sabine Raad pendant son stage de Master1, de rechercher une récurrence des mutations constitutionnelles de ce gène dans une cohorte de patients présentant un phénotype proche du patient étudié. Pour cela, nous avons constitué une première cohorte de 32 patients ayant développé un médulloblastome (n=25) ou un méningiome (n=6) ou les deux (n=1) en présence ou non d'un contexte familial de cancer mais sans mutation détectée des gènes *SUFU* ou *TP53*. Nous avons mis au point la recherche des mutations ponctuelles de ce gène par séquençage Sanger et la recherche des remaniements géniques de grande taille par la technique de QMPSF (*Quantitative Multiplex PCR of Short Fluorescent fragments*). Deux QMPSF ont été développées spécialement pour cette expérience : la première ciblant les exons pairs du gène *HID-1* et la seconde les exons impairs. Cette technique n'a pas permis de mettre en évidence un réarrangement de ce gène. Par la suite, nous avons recherché des variations dans chacun des 19 exons du gène *HID-1* par séquençage Sanger. Cette analyse nous a permis de mettre en évidence une deuxième variation faux-sens non répertoriée en population générale (NM\_030630.2; c.1532A>G; p.Glu511Gly), également prédite délétère par les trois logiciels de prédiction de pathogénicité. Cette mutation a été détectée chez un patient ayant développé un médulloblastome à l'âge de 4 ans. Malheureusement, nous ne disposons d'aucune information sur le sous-type histologique du médulloblastome présenté par ce patient. Le grand-père paternel du cas index a développé une tumeur cérébrale à 47 ans, cependant nous ne disposons pas de l'ADN de cet apparenté nous permettant d'effectuer une analyse de co-ségrégation (Figure 39). Cette variation intervient à la fin du seul domaine connu de *HID-1*, le domaine «Dymécline». Ce domaine est un domaine d'homologie à une autre protéine associée au Golgi et impliquée dans le trafic vésiculaire, la Dymécline.





**Figure 39** : Arbre génétique du second patient présentant une altération constitutionnelle du gène *HID-1* (NM\_030630.2; c.1532A>G; p.Glu511Gly).

❖ Arguments fonctionnels en faveur de la pathogénicité des mutations du gène *HID-1* détectées chez des patients ayant développé un médulloblastome

Etant donné le rôle suggéré de *HID-1* dans le trafic vésiculaire, nous avons entrepris une collaboration avec l'équipe du Professeur Maité Montero-Hadjadje (Inserm U982, Mont Saint Aignan) experte dans le domaine de la sécrétion régulée. Nous avons dans un premier temps utilisé les cellules PC12 (cellules dérivées de phéochromocytome de rat) et les cellules COS-7 (cellules dérivées de rein de singe) exprimant de façon stable la Chromogranine A (COS-7-CgA) et contenant des vésicules à cœur dense induites par l'expression de la CgA.

Après clonage de l'ADN complémentaire de *HID-1* dans un vecteur d'expression, les différentes mutations détectées chez les patients ont été introduites par mutagenèse dirigée. Deux vecteurs ont été employés : pcDNA3.1(-) et pEGFPN3 permettant de fusionner la protéine fluorescente EGFP en position C-terminale. Par la suite, la variation *de novo* (c.2139T>G ; p.Ile713Met ; NM\_030630.2) identifiée par exome sera dénommée I713M et la variation identifiée (NM\_030630.2; c.1532A>G; p.Glu511Gly) par Sanger sera dénommée E511G. Ces deux variants ont été introduits dans les plasmides d'expression. Nous avons également généré deux autres variants particuliers : (i) un mutant non-sens (W119X) pour permettre l'analyse de l'effet des mutations perte de fonction de *HID-1* et (ii) un mutant de localisation subcellulaire (G2N). Ce dernier a été obtenu en mutant le site de N-myristoylation, impliqué dans la localisation péri-golgienne de la protéine *HID-1* comme

préalablement décrit dans la littérature <sup>144</sup>.

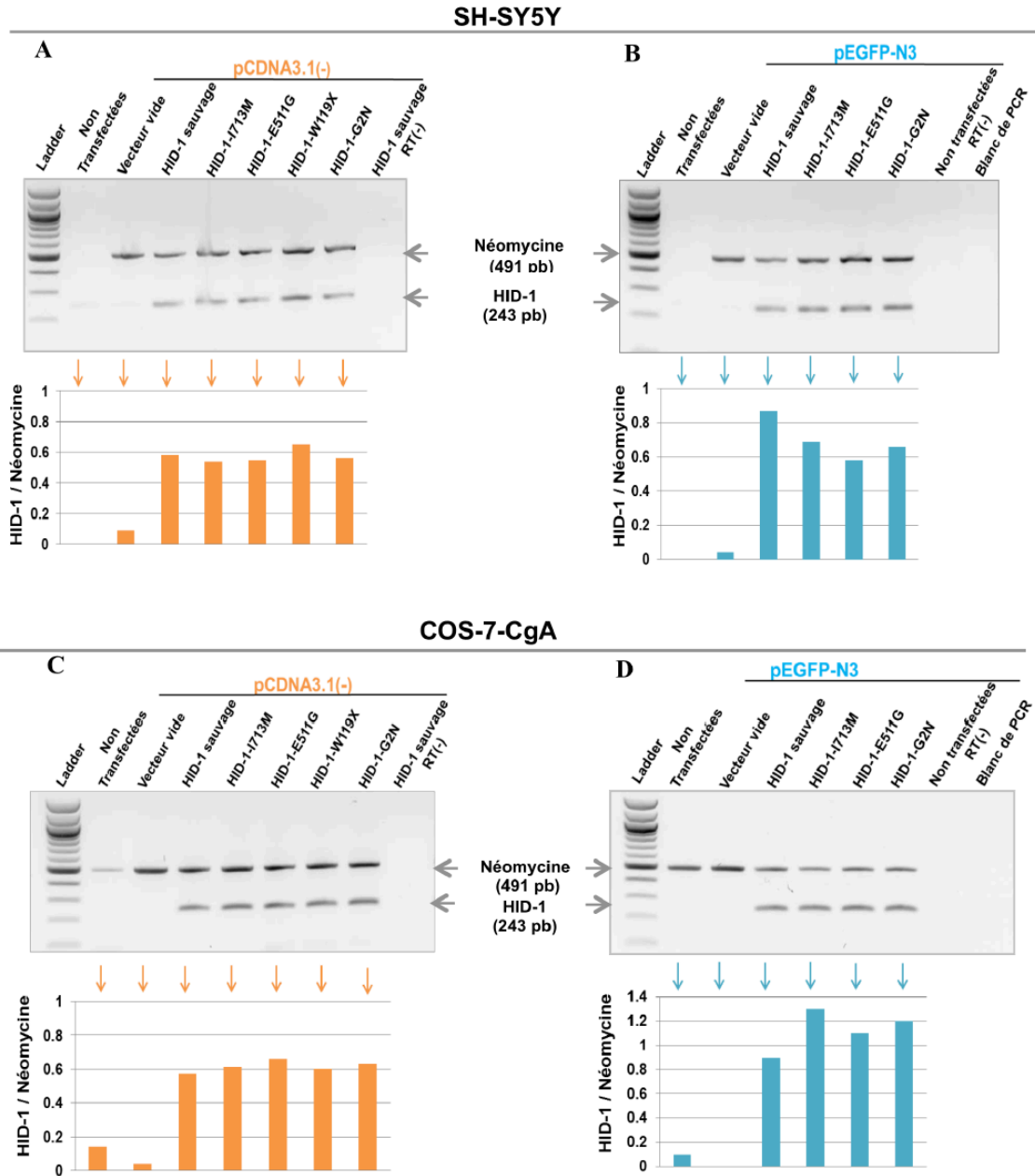
Dans un premier temps, **l'impact des mutations détectées sur l'expression de l'ARNm du gène *HID-1*** a été analysé. Après transfection transitoire de la construction sauvage ou mutée dans les cellules SH-5YSH ou COS-7-CgA, le niveau d'expression des transgènes mutés a été analysé par RT-PCR en comparaison avec celui du transgène sauvage. L'amplicon choisi pour la détection d'*HID-1* a une taille de 243 pb. Il est à noter d'ailleurs une bande à cette taille dans les échantillons correspondant aux cellules non transfectées ou transfectées avec le plasmide vide. Cette bande correspond à l'expression endogène du gène *HID-1* (Figure 40). Lorsque les cellules sont transfectées avec la forme sauvage de *HID-1* ou les formes mutantes (I713M, E511G, W119X et G2N), il y a une surexpression de la protéine *HID-1* à ce niveau dans les échantillons. Une seconde bande à environ 490 pb correspond à la néomycine. Cette bande est observée dès que l'on transfecte les cellules avec nos constructions. On peut noter la présence également de cette bande dans les cellules COS7-CGA non transfectées. La présence de cette bande correspond à l'intégration dans le génome de ces cellules du gène de résistance à la néomycine qui a été utilisé pour sélectionner les clones ayant intégré de façon stable la CgA. Pour chaque condition, le niveau d'expression de *HID-1* a été rapporté à celui de la néomycine afin de s'affranchir des artefacts liés à des variations dans l'efficacité de transfection. Puis, nous avons vérifié la surexpression de *HID-1* au niveau des transcrits en comparant les niveaux obtenus dans les cellules transfectées par rapport aux cellules non transfectées. Les résultats montrent que le niveau d'expression de *HID-1* sauvage est de 6 à 8 fois plus élevé que celui de *HID-1* endogène pour les deux types cellulaires et les deux constructions (pcDNA3.1(-) et pEGFP-N3). La comparaison des résultats obtenus entre la transfection avec les vecteurs *HID-1* sauvage et les vecteurs portant les mutations I713M, E511G, W119X et G2N semblent indiquer que ces mutations n'impactent pas les niveaux d'ARN de *HID-1* dans notre système (Figure 40 A, B, C, D).

Par la suite, dans l'objectif d'évaluer l'impact des mutations détectées sur l'expression de la protéine *HID-1* nous avons transfecté les cellules SH-SY5Y et les cellules CO7-CgA avec nos constructions *HID-1* sauvages ou mutées et analysé par western blot l'expression protéique de *HID-1* dans ces différentes conditions. Pour les deux types cellulaires, lors de la transfection avec le vecteur pcDNA3.1-*HID-1* sauvage, une bande à environ 95 kDa (Figure 41, A, C) est visible. Cette bande est aussi observée après transfection des constructions contenant les variants I713M, E511G, G2N. Comme attendu, cette bande n'est plus détectée avec la mutation W119X car cette construction conduit à une protéine tronquée en C-terminal, partie reconnue par l'anticorps. Les mêmes résultats sont observés après

transfection des cellules avec les constructions de pEGFP-N3 (Figure 41, B, D) mais pour ces conditions la bande observée correspond à notre protéine de fusion HID-1:EGFP dont le poids moléculaire est d'environ 130kDa. D'autres bandes de plus faible poids moléculaires peuvent être observées. Ces dernières correspondent *a priori* des produits de dégradation de la protéine HID-1.

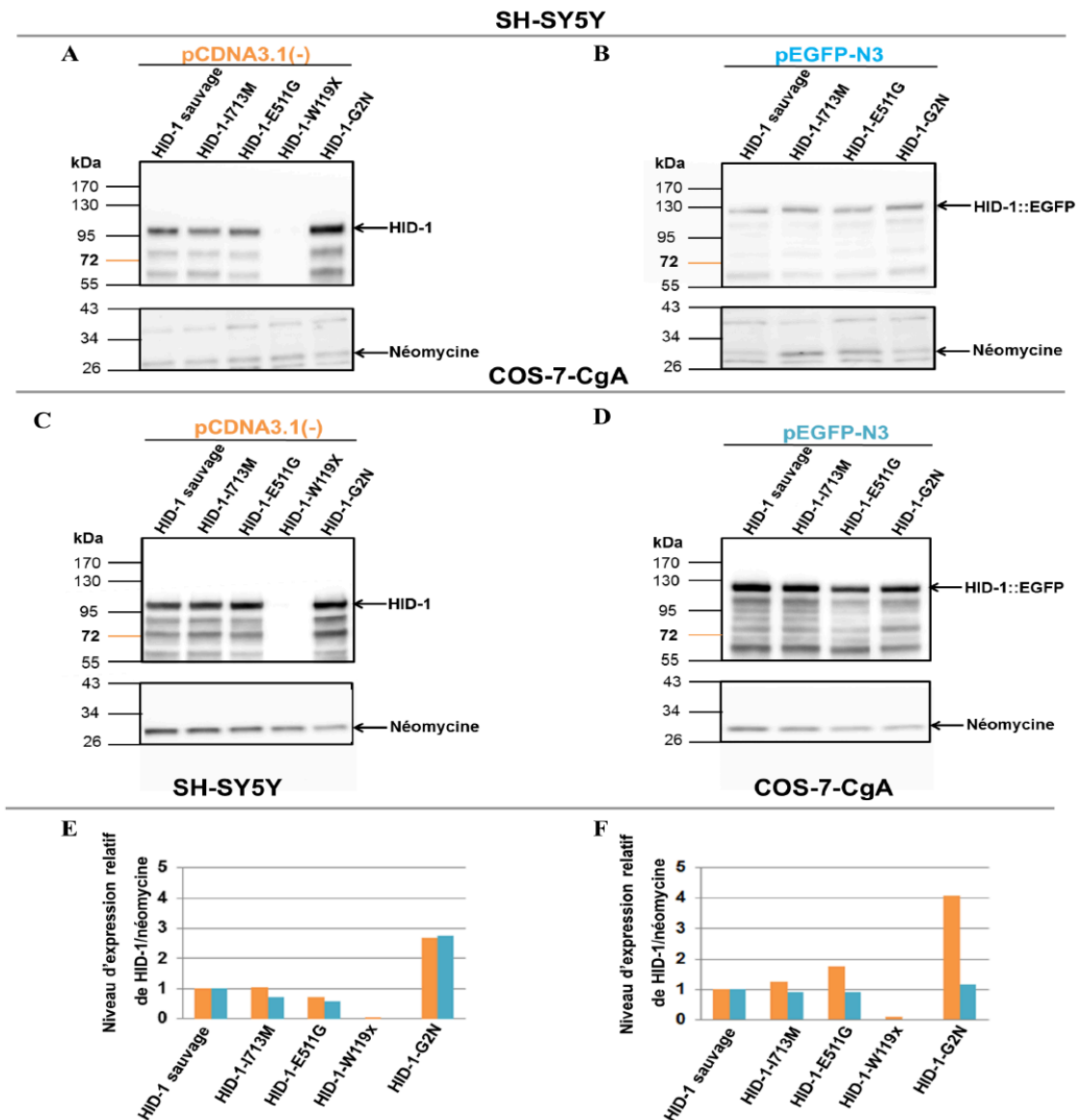
Afin de quantifier les protéines HID-1 produites à partir des constructions sauvages ou mutées, les résultats ont été normalisés à l'aide du système Stain Free (Bio-Rad) mais également en réalisant un marquage néomycine permettant d'éliminer les biais de quantification liés aux efficacités de transfection. Ce marquage néomycine se traduit par une bande à environ 30 kDa (Figure 41 A, B, C, D). Les résultats de quantification semblent indiquer que le niveau d'expression de HID-1 après transfection des deux mutants I713M et E511G est similaire en comparaison avec HID-1 sauvage dans les deux types cellulaires et avec les deux plasmides (Figure 41 , E, F). En ce qui concerne la mutation contrôle G2N (mutant de localisation) les résultats, qui sont à reproduire, semblent suggérer une augmentation de la quantité de protéine mutée par rapport à la condition référence.

Un autre impact potentiel des mutations détectées du gène *HID-1* pourrait-être un **défaut de localisation empêchant ainsi une fonction correcte**. Dans l'objectif de tester cette hypothèse, nous avons transfecté de façon transitoire les cellules SH-SY5Y avec des plasmides permettant d'exprimer la protéine de fusion HID-1::EGFP sauvage ou mutée et analysé par microscopie confocale (Plate-forme PRIMACEN). Préalablement, la distribution subcellulaire de la protéine HID-1 endogène décrite dans la littérature à savoir une localisation péri-golgienne ainsi qu'un marquage punctiforme cytosolique a été confirmés dans les deux modèles cellulaires employés. Dans notre modèle de surexpression, la localisation subcellulaire de la forme sauvage semble suivre celle de l'endogène, validant notre modèle. Comme contrôle de localisation, nous avons utilisé le mutant G2N impactant le site de N-myristoylation nécessaire à l'ancrage peri-golgien de la protéine HID-1. Nos expériences confirment l'importance de ce motif pour l'ancrage péri-golgien. En effet, la localisation de cette forme mutée de la protéine HID-1 est maintenant diffuse dans le cytoplasme. Concernant la distribution des formes mutées I713M et E511G, ces dernières semblent suivre la distribution observée de l'endogène et de la forme sauvage suggérant que les mutations identifiées n'impactent pas la localisation de HID-1 dans notre système (Figure 42).



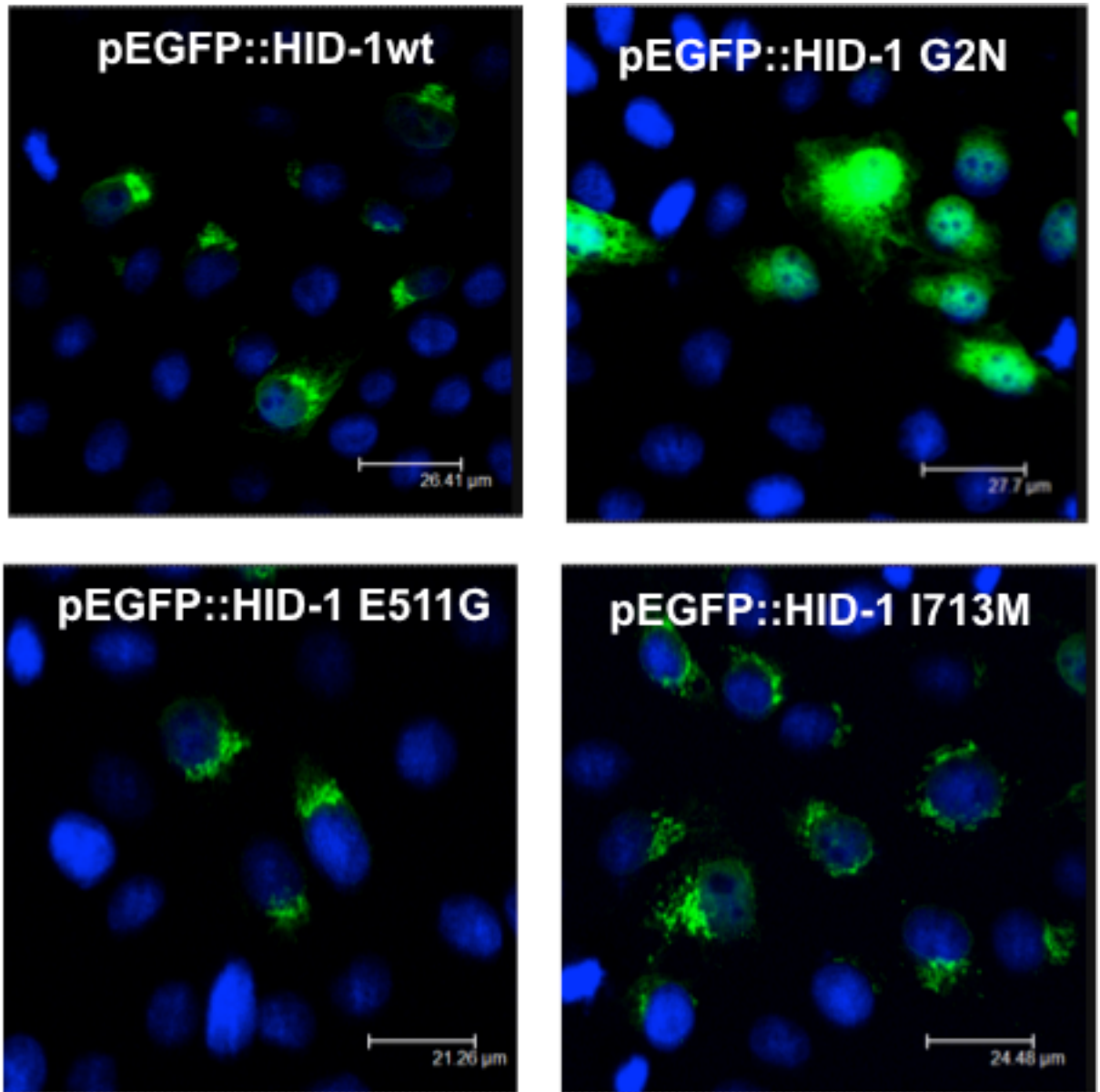
**Figure 40:** Analyse de l'effet des mutations sur l'expression des transcrits de HID-1 par RT-PCR

Amplification par RT-PCR semi-quantitative du transcrit HID-1 (exons 6-7) et du transcrit de la néomycine après transfection des cellules SH-SY5Y (A, B) ou COS-7-CgA (C, D) avec les constructions pCDNA3.1(-) (A, C) ou pEGFP-N3 (B, D) sauvages ou mutées. Les niveaux d'expression relatifs de HID-1 ont été calculés en divisant la quantité de transcrit HID-1 détectée sur gel par celle de la néomycine.



**Figure 41** : Impact des mutations sur l'expression protéique du gène *HID-1*

Analyse par western blot des protéines *HID-1* et des protéines de fusion *HID-1::EGFP* dans les cellules SH-SY5Y (A,B) et COS-7-CgA (C,D) après transfection avec les constructions pCDNA3.1(-) (A, C) ou pEGFP-N3 (B, D) sauvages ou mutées et quantification en stain free (Bio-Rad) et par rapport à la néomycine des niveaux relatifs d'expression des protéines *HID-1* ou *HID-1::EGFP* détectées dans les SH-SY5Y (E) ou COS-7-CgA (F), normalisé par rapport à la quantité de protéines *HID-1* produites par les constructions *HID-1* sauvage.



**Figure 42** : Analyse de l'impact des mutations G2N, E511G, I713M sur la localisation subcellulaire d'HID-1

Analyse par microscopie confocale de la localisation des protéines de fusion pEGFP::HID-1 sauvage (en haut à gauche), pEGFP::HID-1 G2N (en haut à droite), pEGFP::HID-1 E511G (en bas à gauche), pEGFP::HID-1 I713M (en bas à droite) dans les cellules SH-SY5Y. Les protéines d'intérêt sont visualisées en vert, les noyaux sont marqués au DAPI (bleu).

## IV. Discussion

Devant la précocité de l'apparition de la tumeur chez ce jeune patient ainsi que la survenue d'une seconde tumeur, une prédisposition génétique a été supposée. La configuration familiale nous a permis d'effectuer un séquençage exomique du trio enfant malade/parents sains. Cette analyse a révélé la présence de deux variations nucléotidiques *de novo*. La première affectant le gène *MYO7A*, semble expliquer la surdit e observ e chez ce jeune patient, surdit e qui avait  t e pr ealablement attribu e au traitement par radioth erapie. La seconde, quant   elle, affecte le g ene *HID-1*. Etant donn e le faible nombre de publications relatives au g ene *HID-1*, nous avons entrepris dans un premier temps une analyse de r ecurrence chez des patients pr esentant un ph enotype similaire   celui du patient cribl e. Cette analyse nous a permis d'identifier une seconde variation dans le g ene *HID-1*. Malheureusement, nous n'avons pas acc es   l'ADN de l'apparent e de ce patient qui pr esente une tumeur c erebrale   l' age de 47 ans afin de tester la co-s egr egation de l'alt eration et de la pathologie. N eanmoins, il est possible qu'il s'agisse d'une ph enocopie  tant donn e l' age plus tardif de d eveloppement de cette tumeur et le fait que nous n'avons aucune autre information dans le dossier m edical sur la nature de cette tumeur. A l'heure actuelle, nous n'avons pas connaissance d'autres alt erations constitutionnelles du g ene *HID-1*. Trois publications rapportent des s equençages   haut-d ebit de tumeurs de m edulloblastomes. N eanmoins, aucune ne rapporte d'alt eration somatique du g ene *HID-1* dans ce contexte (<http://www.cbioportal.org/>). Il aurait  t e possible de poursuivre cette  tude en  largissant notre cohorte de r eplication et notamment en collaborant avec le consortium international sur le m edulloblastome (Michael Taylor, Olivier Delattre) dans le but d'identifier d'autres  v enements constitutionnels mol eculaires du g ene *HID-1*.

Etant donn e le r ole sugg er e du g ene *HID-1* dans la s ecr etion, nous avons entrepris une collaboration avec le Professeur Mait e Montero-Hadjadje et son  tudiante en th ese Oph elie Carmon (Mont Saint Aignan, Inserm U1239). Malheureusement, ces exp erimentations ont  t e difficiles   mettre en place. Nous avons pour projet de r ealiser une exp erience d'analyse de l'exocytose des v esicules   c oeur dense (DCV) en  valuant la s ecr etion de CGA par des approches de *western blot*. N eanmoins, nous avons  t e frein es par de nombreuses difficult es techniques comme l'extraction prot eique de la fraction s ecr et ee et la normalisation de cette derni ere.

Enfin  tant donn e que la prot eine *HID-1* est extr emement conserv ee au cours de l' volution, nous aurions pu envisager la construction d'un mod ele animal. Par exemple, il est connu que chez *C. elegans*, une lign ee d eficiente pour *hid-1* pr esente une alt eration de la voie de s ecr etion r egul ee entrainant un ph enotype l ethargique, mod er ement constip e, insensible au toucher et avec des probl emes de coordination de la motricit e <sup>143</sup>. De plus,

l'homologie entre la protéine humaine et celle de *C. elegans* est telle que la protéine humaine est capable de compléter le phénotype de la lignée *C. elegans* déficiente pour *hid-1* pour restaurer un phénotype sauvage.

En conclusion, la démonstration fonctionnelle des variations identifiées en exome puis en séquençage ciblé chez des patients présentant un médulloblastome est un long processus. Cela souligne toute la difficulté d'interprétation des données. Ce travail confirme la puissance de l'analyse d'exome de trios pour identifier très rapidement des variants *de novo* potentiellement responsables d'une prédisposition génétique au cancer. A partir de trois exomes, nous avons été en mesure d'identifier des variations *de novo* dont l'une est vraisemblablement à l'origine de la surdité observée chez ce jeune patient. Ce travail illustre aussi la très grande difficulté de démontrer l'implication des mutations détectées dans la pathologie étudiée. Le séquençage massif d'individu témoins a récemment permis de mettre en évidence une variabilité insoupçonnée du génome humain avec un taux de variants rares ou de singletons très élevé ainsi qu'un taux de mutation *de novo* par génération générant en moyenne entre 0 et 2 mutations par exome. A la lumière de ces nouvelles découvertes, le caractère *de novo* d'une mutation ainsi que le fait qu'elle n'ait jamais été répertoriée en population générale peuvent constituer des éléments en faveur de la pathogénicité d'un variant mais ne peuvent suffire à déterminer le caractère causal de celui-ci. Il est donc indispensable de disposer d'autres arguments qui peuvent être d'ordre génétique avec une analyse de co-ségrégation de la pathologie avec le variant ou une récurrence des mutations dans ce gène chez d'autres patients, mais également d'ordre fonctionnel. Ce projet illustre également toute la difficulté de démontrer la causalité du variant lorsque le rôle biologique du gène candidat est peu ou pas connu. Cette démonstration s'avère d'autant plus difficile dès lors que le gène n'est pas lié à une quelconque pathologie tumorale.





**Projet 3 : Recherche de nouvelles bases moléculaires du corticosurréalome de l'enfant par analyse d'exomes**

## I. Contexte

Le troisième volet de ma thèse visait à identifier de nouvelles bases moléculaires du corticosurréalome survenant chez l'enfant, et notamment chez l'enfant de moins d'un an, en l'absence d'altération détectée de *TP53*. En effet, le corticosurréalome est une tumeur maligne du cortex surrénalien qui affecte généralement l'enfant avec un pic d'incidence à 3 ans. Ce type de tumeur est particulièrement rare (1 à 2/10<sup>6</sup> naissances) et survient généralement de façon sporadique. Cependant, il existe quelques cas où le corticosurréalome est associé à des maladies héréditaires rares et en particulier le syndrome de Li-Fraumeni (*TP53*), le syndrome de Wiedemann-Beckwith (dû à des altérations épigénétiques et/ou génétiques qui dérégulent l'empreinte parentale des gènes du chromosome 11p15.5). Il est important de souligner d'emblée que le taux de détection d'une altération de *TP53* chez les enfants ayant développé un corticosurréalome est de 45 %, ce qui en fait le gène majeur de prédisposition à cette tumeur<sup>137</sup>.

Du fait du recrutement national du laboratoire de Génétique Moléculaire du CHU de Rouen pour le diagnostic moléculaire du syndrome de Li-Fraumeni, nous disposons d'une collection exceptionnelle de prélèvements de patients atteints d'un corticosurréalome pour lesquels aucune base moléculaire n'a pu être mise en évidence. Ainsi, nous avons mis à profit cette collection pour tenter d'identifier de nouvelles bases moléculaires du corticosurréalome survenant chez l'enfant, et particulièrement de l'enfant de moins d'un an, en l'absence de mutation constitutionnelle de *TP53*. Pour cela, nous avons employé deux approches basées sur l'exome : la stratégie exomique soustractive et la stratégie exomique comparative interfamiliale.

## II. Matériels et méthodes

### ❖ Patients analysés

Dans ce projet, nous avons analysé 10 patients ayant développé un corticosurréalome à un âge précoce (avant l'âge de 1 an pour 8 cas). Les ADN parentaux étaient disponibles pour deux familles (patient #1 et patient #2). Les informations concernant les âges de survenue et le caractère sporadique est récapitulé dans le tableau 6.

**Tableau 6 : Récapitulatif des informations concernant les familles ciblées dans le projet 3**

Individu	Sexe	Age de survenue du corticosurrénalome	Présentation	Séquenceur	Pipeline Bioinformatique	Moyenne de séquençage (x)	% de bases >30x	Total Reads mappés	On-target (%)
#01	M	1an	sporadique	GAIIX	CASAVA	65.2	86.40	73,520,655	71,7
Père du #01	M	-	-	GAIIX	CASAVA	62.1	84.48	68,377,251	73.5
Mère du #01	F	-	-	GAIIX	CASAVA	57.0	82.39	62,228,135	73.9
#02	F	9 ans	sporadique	GAIIX	CASAVA	55.9	80.65	59,583,563	75.5
Père du #02	M	-	-	GAIIX	CASAVA	60.2	82.78	65,84,2283	73.8
Mère du #02	F	-	-	GAIIX	CASAVA	50.7	77.22	56,285,567	72.9
#03	F	1an	non	NextSeq500	BWA-GATK	88,3	90,92	100,470,366	71,5
#04	F	7 mois	non	NextSeq500	BWA-GATK	72,6	86,33	81,279,035	76,3
#05	M	6 mois	sporadique	NextSeq500	BWA-GATK	74,8	87,44	84,068,451	76,8
#06	F	5mois	non	NextSeq500	BWA-GATK	69,3	85,43	77,392,534	76,5
#07	F	10 mois	sporadique	NextSeq500	BWA-GATK	84,6	90	94,409,210	74,6
#08	M	anténatal	sporadique	NextSeq500	BWA-GATK	66,1	83,41	74,940,464	75,7
#09	M	2 mois	sporadique	NextSeq500	BWA-GATK	76,7	87,16	87,095,594	73,8
#10	F	3mois	sporadique	NextSeq500	BWA-GATK	74	86,13	86,736,193	73,6

#### ❖ Séquençage exomique et filtration des variants

Les librairies d'ADN ont été synthétisées à partir de 3 µg d'ADN génomique extrait à partir du sang périphérique, grâce au kit Agilent SureSelect XT. L'enrichissement des régions exoniques a été réalisé à partir de la version V5 du kit SureSelect All Exons.

Les librairies ont ensuite été soumises à un séquençage *paired-end* de 2 x 75 pb sur un séquenceur Illumina GAIIx ou un séquenceur Illumina NextSeq 500 en 2 x 150 pb. Le séquençage de ces exomes a permis d'obtenir une profondeur en moyenne de 68 x. En moyenne, 96 % des bases étaient lues au minimum 10 fois.

Les deux trios analysés ont été traités par un double pipeline CASAVA et BWA-GATK. L'alignement a été réalisé par CASAVA (version 1.8.2) et BWA (version 0.7.5.a). Le variant calling a été réalisé par CASAVA (version 1.8.20) et GATK (version 3.1.1). L'annotation a été réalisée par Annotvar. La filtration a été réalisée par des scripts perl pour éliminer les variants présents chez les parents (variants hérités), les variants présents dans les bases de données 1000 GENOMES PROJECT et ESP à une fréquence supérieure à 0,1 %, les variants présents dans une base interne au laboratoire de 72 exomes, les variants présents à une faible profondeur (< 5 x) et de mauvaise qualité (Qscore < 30) et les variants ayant une balance allélique inférieure à 20 %. Cette filtration a ensuite été complétée par un curage manuel des fichiers BAM consistant principalement à éliminer les artefacts de séquençage comme des artefacts de biais de brin non éliminés par les versions logicielles utilisées au moment de l'étude.

Pour l'étude de récurrence, les 8 exomes ont été alignés au génome de référence (hg19) à l'aide de BWA-GATK (version 0.7.5a et 3.1.1 respectivement). L'étape de *variant calling* a été réalisée à l'aide de GATK HaplotypeCaller puis l'annotation a été réalisée à l'aide d'Annotvar. La filtration a été réalisée à partir de script perl pour générer un fichier de comptage de récurrence en éliminant les variants avec une profondeur < 5 x, une balance allélique d'au moins 20 % pour les hétérozygotes et une MAF supérieure à 1%.

#### ❖ Confirmation Sanger des variants identifiés

Pour confirmer l'existence des variants détectés en exome, un séquençage Sanger a été réalisé. Pour cela, des amplifications par PCR autour des variants ont été réalisées en utilisant de 50 à 100 ng d'ADN génomique extrait à partir du sang périphérique. Le séquençage Sanger de ces fragments a ensuite été réalisé à l'aide du kit BigDye® Terminator v3.1 Cycle sequencing (Life Technologies). Les réactions de séquence sont été

séparées par électrophorèse sur séquenceur 3100xl Genetic Analyzer (Life Technologies) et analysées à l'aide du logiciel Sequencing Analysis Software 5.2.0 (Life Technologies).

### III. Résultats

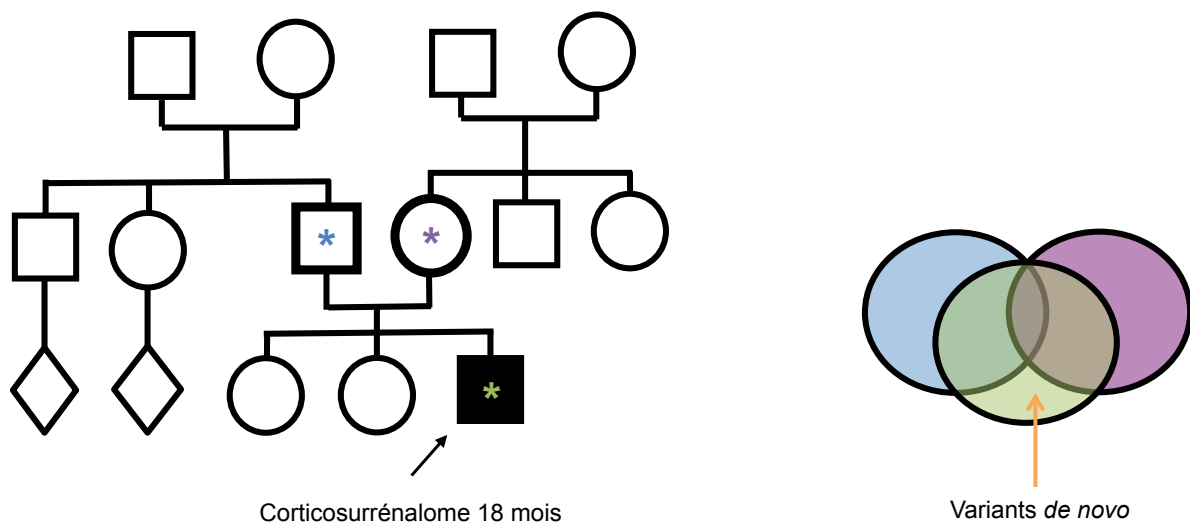
#### ❖ Analyse exomique soustractive du cas index #01

Ce projet a débuté par l'étude d'un jeune patient ayant développé un corticosurréalome à l'âge de 18 mois en dehors de tout contexte familial de cancer (Tableau 6 et figure 43). L'analyse extensive du gène *TP53* chez ce patient s'est avérée négative. Ainsi, devant ce phénotype extrême de présentation sporadique, l'existence d'une mutation *de novo* dans un gène de prédisposition au cancer a été suspectée. Un séquençage exomique a été réalisé chez ce patient ainsi que chez ses parents, afin de mettre en œuvre une stratégie exomique soustractive (Figure 43). Cette analyse soustractive a été réalisée selon le protocole détaillé dans la partie matériel et méthode. A la fin de l'analyse bioinformatique, douze variants répondaient aux critères d'une variation *de novo* rare. Cependant, après un curage des données par analyse visuelle des fichiers BAM, aucun variant n'a été conservé, ces derniers étant majoritairement des artefacts.

Néanmoins, nous avons donc décidé d'explorer les variants sous d'autres hypothèses de transmission. Dans un premier temps, nous avons envisagé la possibilité d'une pathologie à transmission autosomique récessive liée à des variants hétérozygotes composites, c'est à dire la transmission d'un variant rare par le père et un variant rare transmis par la mère dans le même gène à des positions génomiques différentes. Arbitrairement, nous avons décidé d'éliminer les variants présentant une fréquence supérieure à 1/1000 étant donné la fréquence du corticosurréalome ( $1/10^6$  naissances). Ainsi, 446 variants potentiellement hétérozygotes composites ont été identifiés. Etant donné qu'il s'agit d'une collection extrêmement rare et précieuse, nous avons analysé visuellement ces 446 variants impactant 223 gènes. Malgré cette longue analyse, nous n'avons pas pu mettre en évidence de variants hétérozygotes composites potentiellement délétères. Il s'agissait d'artefacts de séquençage, d'artefacts de biais de brin, ou de variants dont la fréquence n'est pas concordante avec celle du corticosurréalome. Cette analyse serait sûrement aujourd'hui simplifiée grâce à l'évolution des pipelines bio-informatiques permettant de filtrer par exemple les artefacts de séquence.

Dans un second temps, même s'il ne s'agissait pas d'une famille consanguine, nous avons quand même exploré la possibilité d'une transmission récessive de variants homozygote rares. Pour cela, nous avons éliminé les variants présents dans les bases de données avec une fréquence supérieure à 1/1000 et les variants de mauvaise qualité. Ainsi, 89 variants répondaient à ces critères. Après analyse visuelle des fichiers BAM et analyse bibliographique, aucun variant n'a été retenu.

En conclusion, malgré une analyse extensive des données exomiques issues de ce trio, nous n'avons pas pu mettre en évidence de variants *de novo*, homozygotes rares ou hétérozygotes composites rares pouvant expliquer la pathologie observée chez ce jeune patient de 18 mois.



**Figure 43** : Arbre génétique de la famille du cas index #01 et la stratégie exomique employée

Etant donné le contexte sporadique dans lequel le cas index (flèche noire) a développé un corticosurréalome à 18 mois, nous avons employé une stratégie exomique soustractive. Les variants identifiés par exome des parents (schématiquement en bleu et violet) étaient soustraits à l'exome de l'enfant (schématiquement en vert) afin d'identifier les variants non hérités potentiellement responsables de la pathologie observée chez ce jeune patient. Losange = sexe non spécifié.

#### ❖ Analyse Exomique Soustractive du cas index #02

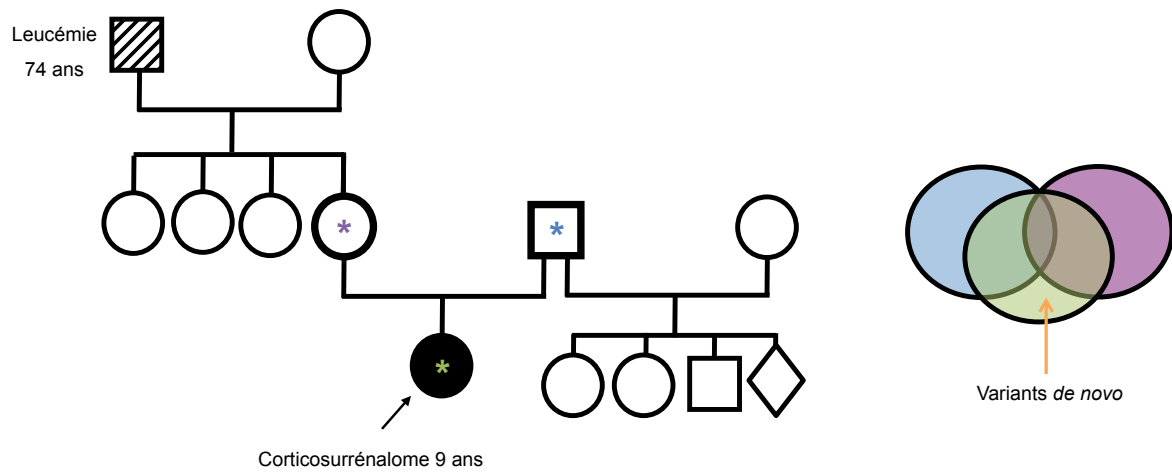
Ce projet s'est poursuivi par l'analyse exomique soustractive d'une autre jeune patiente ayant développé un corticosurréalome à l'âge de 9 ans d'emblée métastatique au diagnostic, en l'absence de contexte familial de cancer (Tableau 6 et Figure 44). Le

caractère explosif, la précocité d'apparition de la tumeur et la présentation sporadique prêtent à penser à une mutation *de novo* à l'origine du phénotype observé (Figure 44). Un protocole analogue à l'étude précédente tant en termes de préparation d'échantillon, de séquençage, de stratégie de filtration, a été appliqué. Vingt-sept variants potentiellement *de novo* et rares ont été identifiés. Néanmoins, après cure manuelle par analyse visuelle des fichiers BAM, aucun variant n'a été retenu. Il s'agissait également principalement d'artefacts de séquençage.

Similairement au trio précédent, d'autres modes de transmission ont été envisagés. Premièrement, nous avons envisagé la possibilité d'une transmission autosomique récessive liée à des variants rares hétérozygotes composites comme précédemment. Ainsi, 745 variants hétérozygotes composites ont été identifiés. Malgré le nombre important de variants et étant donné le caractère exceptionnel de la pathologie, une cure manuelle par analyse visuelle des fichiers BAM a été entreprise. Cependant, aucune altération hétérozygote composite n'a été retenue. Beaucoup de variants étaient en réalité des artefacts de séquençage, des variants impactant des gènes très polymorphiques ou des variants dont les fréquences n'étaient pas concordantes avec la fréquence du corticosurréalome.

Secondairement, bien qu'il ne s'agissait pas d'une famille consanguine, une transmission récessive de variants homozygotes rares a été explorée. Après élimination des variants de mauvaise qualité (Qscore < 30, bases lues moins de 10 fois) et des variants dont la fréquence allélique dépassait les 1/1000 (1000GENOMES, ESP), 58 variants homozygotes rares ont été identifiés. Cependant, après analyse visuelle des fichiers BAM, aucune de ces variations ne s'est avérée être pertinente pour les mêmes raisons que précédemment. En conclusion, malgré une analyse extensive de ces données exomiques issues du trio, aucun variant à effet causal n'a pu être identifié. Ce type d'analyse est maintenant grandement simplifié à l'aide des nouveaux outils bioinformatiques développés, outils dont nous ne disposons pas au moment de l'analyse. Il peut s'agir d'amélioration d'outils préexistants comme par exemple la prise en compte de l'environnement de la base lue pour évaluer sa véracité (plus la base lue est dans une région difficile à lire, plus la probabilité qu'il s'agissent d'un faux positif est grande) ou la création de nouveaux pipelines avec de nouveaux outils.





**Figure 44** : Arbre génétique de la famille du cas index #02 et la stratégie exomique employée dans cette famille

Etant donné le contexte sporadique dans lequel le cas index (flèche noire) a développé un corticosturréalome à 9 ans, nous avons employé une stratégie exomique soustractive. Les variants identifiés par exome (schématiquement en bleu et violet) ont été soustraits à l'exome de l'enfant (schématiquement en vert) afin d'identifier les variants non hérités potentiellement responsables de la pathologie observée chez cette jeune patiente.

❖ Etude par analyse exomique comparative interfamiliale de 8 patients ayant développé un corticosurréalome avant l'âge de 1 an.

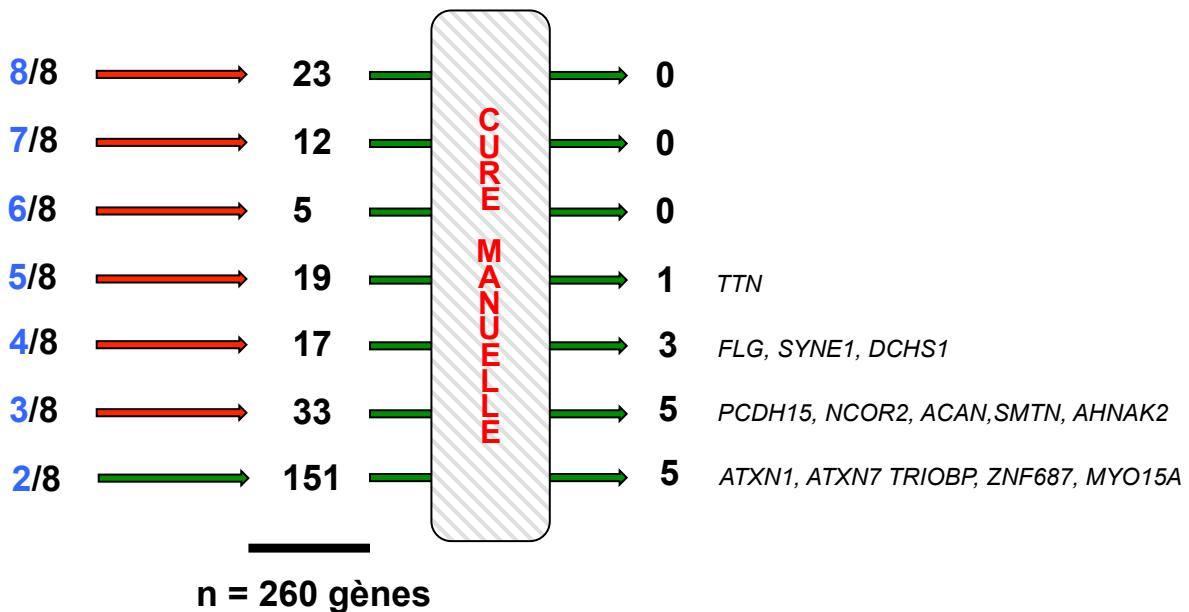
Nous disposons au laboratoire d'une collection extrêmement précieuse de huit ADN de patients ayant développé d'un corticosurréalome avant l'âge de 1 an (Tableau 6). Il s'agit d'une cohorte exceptionnelle dans la mesure où la fréquence du corticosurréalome est estimée à 1 cas pour un à deux millions de naissance. En sus, ici, il s'agit de corticosurréaumes de l'enfant de moins d'1 an dont la fréquence est extrêmement difficile à déterminer.

Cette analyse a porté sur 8 patients de moins d'un an : cinq présentaient un corticosurréalome dans un contexte sporadique et trois dans un contexte familial de cancer. Etant donné la forte contribution du gène *TP53* dans les prédispositions au corticosurréalome, ces patients ont bénéficié préalablement d'une analyse ciblée à haut-débit visant à détecter les altérations du gène *TP53* y compris les mosaïques, qui s'est avérée négative. L'analyse en trio étant impossible du fait que nous ne disposions pas de l'ADN des parents, nous avons donc mis en place une analyse exomique comparative interfamiliale pour identifier une base génétique commune à ces cas exceptionnels.

Pour rappel, dans la stratégie exomique comparative interfamiliale, l'hypothèse développée est qu'un gène pourrait être altéré en commun chez des individus non apparentés présentant le même phénotype. Ainsi, le séquençage exomique des cas index a été conduit. Les variants de mauvaise qualité (<Q30 et bases lues moins de 10 fois) et ceux dont la fréquence allélique dépasse 1/1000 (1000GENOMES, ESP) ont été éliminés de l'analyse. Puis, les variants ont été triés en fonction du nombre de patients présentant à minima une altération dans un gène donné. Par exemple, 8/8 signifie que l'ensemble des huit patients présentent au moins une altération dans un gène commun.

L'analyse bioinformatique a révélé qu'environ 260 gènes répondent à une hypothèse de récurrence (de 2/8 à 8/8). Dans un premier temps, les gènes mutés chez 3 des 8 patients jusqu'à chez 8 des 8 patients (3/8 à 8/8) ont été analysés manuellement. La difficulté de ce type d'analyse est que chaque artefact identifié provoque un changement de catégorie du gène (ex : si le gène x originellement classé 7/8 présente un artefact chez l'un des patients, il sera donc réattribué à la catégorie 6/8). Avec les pipelines informatiques utilisés à ce moment de ma thèse, beaucoup d'artefacts de séquençage étaient malheureusement conservés dans l'analyse et sont donc venus fausser les chiffres. Après élimination des artefacts, les gènes ont donc changé de catégorie. Ainsi, après recalibrage manuel dans la

catégorie correspondante, aucun gène n'est altéré en commun chez les huit patients, ni chez sept sur huit et six patients sur huit (Figure 45).



**Figure 45 :** Résultats de l'analyse interfamiliale et stratégie d'analyse employée par la suite

L'analyse exomique interfamiliale a révélé que 260 gènes pouvaient être gènes candidats selon l'hypothèse de récurrence. Une cure manuelle a été entreprise pour éliminer les artefacts de séquençage. Finalement 14 gènes répondaient à l'hypothèse d'une récurrence. Ces gènes ont été analysés finement un à un selon la fréquence des altérations, les effets prédits des variations identifiées, ainsi que selon la fonction du gène.

Les altérations dans les 14 gènes candidats restants ont été analysées selon leurs fréquences, leurs prédictions bio-informatiques et également leurs fonctions biologiques. Au final, aucun des gènes proposés n'a été retenu comme gène candidat : soit la fréquence des altérations n'est pas cohérente avec la fréquence du corticosurréalome, soit les variants n'étaient pas prédits comme délétères, ou enfin parce que les altérations constitutionnelles de ces gènes n'étaient pas pertinentes dans une pathologie comme le cancer. A titre d'illustration, les gènes *TTN*, *NCOR2*, *SYNE1* sont de très grands gènes et sont donc très polymorphes; les altérations de *FLG* sont impliquées dans l'*ichthyosis vulgaris*, une maladie de la peau; les altérations de *TRIOBP*, *PCDH15*, *MYO15A* dans la surdité, les altérations du gène *ACAN* sont impliquées dans des dégénérationes spinales ou encore *SMTN* n'est pas pertinent étant donné que sa localisation est exclusivement dans le muscle lisse et donc pas en concordance avec un corticosurréalome affectant le tissu surrénalien. La fréquence des altérations du gène *AHNAK2* n'était pas concordante avec la fréquence du corticosurréalome (MAF respectives des trois altérations : 0,008; 0,0032; 0,005). Les altérations dans les gènes des ataxines (*ATXN1*, *ATXN7*) étaient prédites délétères par un seul des trois logiciels de prédictions que nous utilisons (uniquement Mutation Taster). Les

altérations dans le gène *ZNF687* n'ont pu être incriminée étant donné le faible nombre d'article relatif à la fonction de la protéine codé par ce gène.

Devant le caractère exceptionnel de cette cohorte, nous avons également procédé à une analyse individuelle de chaque exome pour les huit patients. Afin d'identifier des éléments moléculaires à l'origine de la prédisposition observée parmi les 20 000 variants de chaque exome, nous avons employé un filtre basé sur l'absence des variants dans les bases de données de fréquence allélique (MAF = 0% dans ExACall ou ESP ou 1000GENOMES) ainsi qu'un filtre de pathogénicité prédite (3/3 logiciels prédisant une pathogénicité). Puis, seuls sont gardés les variants présentant soit une altération tronquant la protéine (comme une altération décalant le cadre de lecture, les altérations provoquant un gain ou une perte d'un codon stop, les altérations affectant les sites canoniques d'épissage jusqu'à deux nucléotides des bornes exoniques) ou soit une altération "strictement délétère", c'est à dire prédite délétère par l'ensemble des trois logiciels de prédiction classiquement employés dans la littérature (SIFT, PolyPhen2, MutationTaster). Cette analyse a conduit à l'identification de 296 variants candidats. Après analyse visuelle des fichiers BAM, aucun variant potentiellement causal n'a été retenu. Il s'agissait soit d'artefacts de séquençage, soit de gènes dont les altérations documentées n'ont pas de lien évident avec une quelconque pathologie tumorale.

#### IV. Conclusion et discussion

L'objectif général de ce volet de ma thèse était consacré à l'identification de nouveaux gènes impliqués dans le corticosurréalome du très jeune sujet à l'aide de l'exome. Étant donné que nous possédons au laboratoire une collection exceptionnelle d'ADN de patients jeunes ayant développé un corticosurréalome et pour lesquels aucune base moléculaire n'avait pu être identifiée, nous avons entrepris une double analyse basée sur l'exome pour réaliser cet objectif.

Dans un premier temps, nous disposons de l'ADN des parents pour deux patients ayant développé un corticosurréalome en l'absence de contexte familial de cancer. Une stratégie exomique soustractive a pu être conduite chez ces deux familles. Cependant, aucun variant causal n'a pu être incriminé à l'aide de cette analyse y compris sous d'autres hypothèses qu'une mutation *de novo*. Ce résultat décevant soulève plusieurs questions. Premièrement, étant donné la forte prévalence des altérations constitutionnelles dans le corticosurréalome de l'enfant de l'ordre 45 %<sup>137</sup>, il est légitime de se questionner sur l'utilisation de l'exome. En effet, l'exome de par sa nature ne permet pas l'identification de

mutations introniques profondes ou localisées dans les régions promotrices du gène *TP53* pouvant être à l'origine de la pathologie. De plus, il n'est pas exclu qu'au delà des mutations ponctuelles, d'autres événements moléculaires tels que des réarrangements équilibrés ou des inversions puissent être à l'origine de la prédisposition observée. Cependant, l'exome ne permet pas *a priori* leur détection, bien qu'aujourd'hui de nombreux logiciels soient développés dans ce sens. Deuxièmement, notamment dans le cas des présentations sporadiques, l'utilisation du sang comme matériel peut être questionné puisque ce type de présentation peut résulter d'une mutation *de novo* en mosaïque. Les profondeurs de lecture obtenues dans le cadre de ce travail n'excèdent généralement pas les 100x en moyenne, il est donc possible de manquer dans l'analyse un événement moléculaire présent à bas bruit. Ainsi, il serait pertinent d'envisager d'utiliser le tissu surrénalien sain adjacent et une profondeur de lecture plus importante dans cette situation.

Dans un second temps, étant donné que nous disposions d'une collection d'ADN de patients ayant développé des corticosurrénales avant l'âge d'un an, nous avons entrepris une stratégie d'analyse exomique comparative interfamiliale entre ces patients. Malheureusement, cette dernière s'est également avérée infructueuse. Face au caractère extrêmement exceptionnel de cette cohorte, nous avons analysé individuellement les exomes de chaque patient. Toute la difficulté de ce type d'analyse réside dans être à la fois suffisamment stricte pour filtrer les 20 000 variants classiquement identifiés par exome sans pour autant exclure le ou les variants causaux. Il est donc possible qu'à ce titre, notre stratégie de filtration ait été trop drastique. De plus, les zones les moins bien couvertes de l'exome sont exclues de l'analyse et contiennent potentiellement des variants d'intérêt.

Ce double insuccès soulève plusieurs questions sur l'exome et son utilisation dans des pathologies comme le corticosurrénales de l'enfant où un seul gène, en l'occurrence *TP53*, possède une contribution importante. L'exome est-il l'outil adapté ? N'est-il pas nécessaire d'employer des outils comme le panel de gènes ou le séquençage de génome, afin de séquencer la totalité du gène de contribution majeure et d'identifier des événements moléculaires plus difficilement détectables comme les inversions, les CNV, ou encore les altérations introniques profondes ? Il est également envisageable que ces pathologies relèvent en réalité d'autres types d'altérations que l'analyse de l'exome ou du génome ne peut détecter à l'instar de la méthylation et autres modifications épigénétiques.

En conclusion, ces résultats démontrent toute la difficulté à identifier de nouvelles bases moléculaires dans le contexte des prédispositions héréditaires au cancer y compris au moyen de nouvelles techniques pangénomiques comme l'exome. De par sa nature focalisée

sur les régions codantes du génome, l'exome ne permet donc pas d'identifier les bases moléculaires en toutes circonstances. Il convient peut-être d'employer des techniques comme le séquençage de génome permettant d'identifier les altérations situées en dehors des 2 % que représente l'exome ainsi que l'ensemble des réarrangements structuraux. En attendant de pouvoir réaliser des génomes à bas coût et que les infrastructures informatiques nécessaires à ce type d'analyse se développent, une solution intermédiaire consiste à utiliser des techniques ciblées comme le séquençage de panels de gènes afin d'obtenir des profondeurs de lecture plus importante et permettre d'identifier les événements moléculaires présents à bas bruit.

La fréquence élevée des mutations constitutionnelles du gène *TP53* dans le corticosurréalome et l'identification récente par notre unité en cours de publication de mutations en mosaïque du gène *TP53* chez des patients ayant développé un corticosurréalome (voir annexe 2), nous laisse supposer qu'il n'existe pas d'autre gène majeur de prédisposition au corticosurréalome et que la voie majeure de l'oncogène de cette tumeur est l'altération du gène *TP53* qu'il s'agisse d'altérations constitutionnelles, d'altérations constitutionnelles en mosaïque ou d'altérations somatiques.



**Projet 4 : Recherche de nouvelles bases moléculaires des tumeurs pédiatriques par séquençage à haut-débit de 201 gènes impliqués dans le cancer.**



## I. Contexte et objectifs de l'étude

**Un des arguments les plus évocateurs d'une forme héréditaire de cancer est leur précocité.** Or, le bilan d'activité des laboratoires d'oncogénétique révèle **qu'une fraction importante des jeunes patients ayant développé une tumeur restent sans mutation détectable dans les gènes connus faisant l'objet d'une analyse diagnostique.** Même si une fraction importante des cas sans mutation détectable s'explique probablement par des altérations somatiques ou des agrégations familiales non Mendéliennes, il est logique de penser que certains patients présentent des mutations rares constitutionnelles à ce jour non détectables. Le constat demeure donc celui d'une méconnaissance des bases moléculaires des cancers héréditaires.

Le séquençage d'exome s'est révélé être une approche puissante pour identifier de nouvelles bases moléculaires de maladies mendéliennes à partir d'un nombre limité de patients et même à partir d'un seul patient. Compte-tenu du nombre de SNV dans l'exome humain et du défi que représente l'interprétation biologique des variants, cette approche nécessite de comparer les exomes pour isoler les variants d'intérêt. Des stratégies de comparaison différentes ont été développées : comparaison des exomes entre différents cas index, entre apparentés affectés à distance dans une famille ou entre le cas index et ses parents non affectés. Bien que ces différentes stratégies aient permis d'identifier les bases moléculaires de nombreuses maladies rares, elles n'ont pas été très efficaces dans le domaine de la prédisposition au cancer comme le souligne le nombre limité de publications rapportant de nouvelles mutations constitutionnelles et comme l'illustrent nos travaux ci-dessus. Notre groupe a été l'un des rares à identifier par séquençage de l'exome de nouvelles bases moléculaires aux cancers de début précoce<sup>123</sup>. D'après notre expérience, la difficulté majeure de ces approches réside dans l'hétérogénéité génétique des cancers, dans la difficulté d'interprétation biologique des variants détectés, et dans la démonstration finale de l'implication des gènes identifiés dans la pathologie. Compte-tenu de la fréquence des cancers dans la population générale générant des phénocopies, les études de ségrégation peuvent également se révéler difficiles. L'ensemble de ces éléments nous a amenés à envisager une autre stratégie.

Comme exposé dans la partie introductive de ce document, des publications récentes basées sur des analyses pangénomiques ont révélé que le nombre de gènes affectés par des mutations somatiques contribuant au cancer est plus limité qu'initialement anticipé et qu'ils appartiennent souvent aux mêmes voies biologiques. En outre, près de la moitié des gènes prédisposant au cancer décrits jusqu'à présent sont également affectés par des mutations somatiques dans les tumeurs, ce qui confirme que les voies biologiques affectées

dans le cadre des cancers héréditaires et dans celui des cancers sporadiques sont au moins partiellement superposées. En se basant sur ces observations, nous avons formulé les hypothèses suivantes : (i) le nombre de gènes dont les mutations constitutionnelles contribuent au déterminisme génétique des cancers est également limité, (ii) les gènes potentiellement altérés à l'état constitutionnel sont identiques à, ou impliqués dans les mêmes voies biologiques que les gènes fréquemment mutés à l'état somatique dans les cancers, (iii) les corrélations génotype/phénotype orientant actuellement les analyses moléculaires peuvent être en défaut (concept d'extension phénotypique), (iv) une fraction de ces altérations moléculaires *de novo* peuvent exister à l'état de mosaïque, dès lors que l'évènement mutationnel survient au stade post-zygotique, et de ce fait uniquement détectables par analyse NGS à très grande profondeur. Dans cette dernière hypothèse, seuls certains tissus de l'organisme sont concernés par la mutation, avec de ce fait une fraction allélique très inférieure à 50 % lorsque l'analyse est effectuée à partir d'ADN extrait du sang périphérique. Le NGS à grande profondeur (500 à 1000x) est un outil puissant permettant la détection de ces mosaïques.

Ainsi pour ne se focaliser que sur les variations les plus probablement délétères, l'objectif de ce projet a été **d'étudier par NGS** non plus l'ensemble des 23 000 gènes mais **201 gènes qualifiés de gènes du cancer** dont la contribution au cancer a été étayée de façon non ambiguë. L'identification d'une mutation constitutionnelle de l'un de ces gènes sera d'emblée un argument fort de causalité qu'il conviendra d'étoffer par des études de prédiction de l'effet de la mutation et des analyses statistiques. De plus restreindre l'analyse à 200 gènes permettra de disposer d'une **profondeur de lecture très élevée** (chaque base de l'ADN étant séquencée plusieurs centaines à un millier de fois) permettant la détection de mutations *de novo* à l'état de mosaïque.

Ce projet a été appliqué à **des cas sélectionnés de cancers évocateurs d'un déterminisme génétique du fait de leur précocité** (cancers pédiatriques, cancer du sein avant 31 ans, cancer de l'ovaire avant 41 ans, cancer colorectal avant 31 ans), sans mutation détectable des gènes connus de prédisposition génétique au phénotype considéré. Il est important de noter que ce projet est partagé entre les deux sites de Rouen et de Caen. Etienne Muller (Caen) et Jean-Christophe Thery (Rouen) étaient en charge de la partie relative à l'étude des cancers du sein et de l'ovaire de la femme jeune de moins de 31 ans; Françoise Charbonnier (Rouen) des cancers colorectaux de moins de 31 ans. **Enfin j'avais à charge d'étudier les tumeurs pédiatriques.**

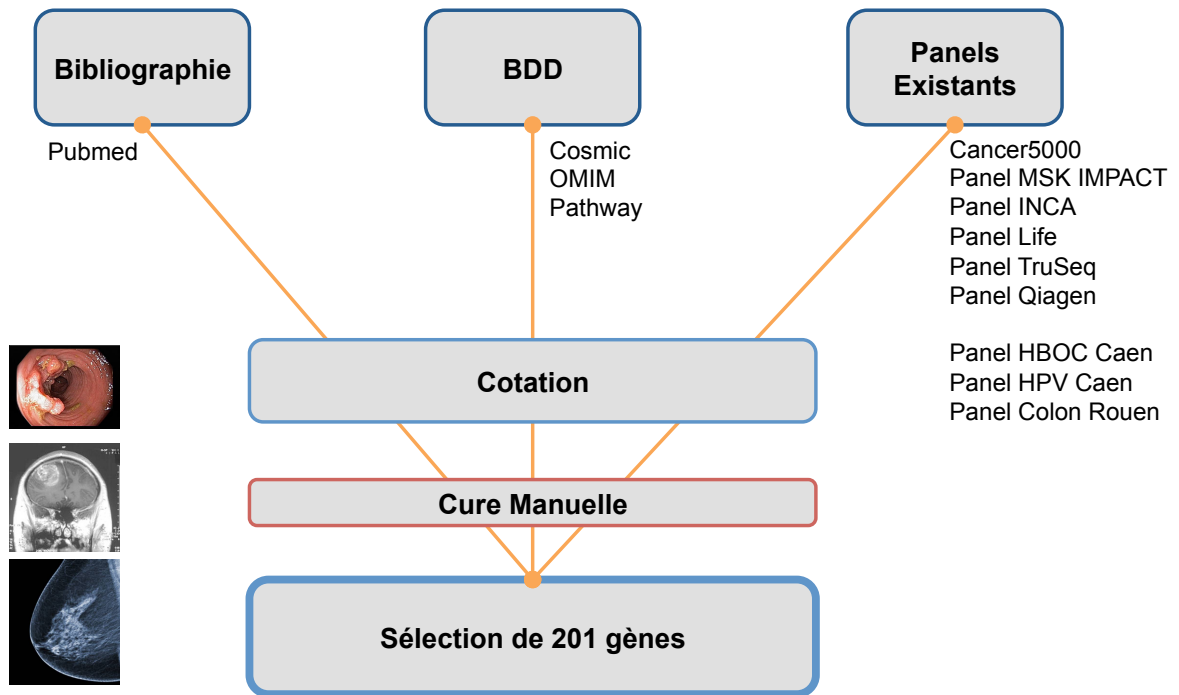
Ce projet vise *in fine* à détecter des variations constitutionnelles rares délétères ou possiblement délétères et d'évaluer leur causalité dans le phénotype tumoral observé, par

des approches *in silico*, des approches fonctionnelles, des études de ségrégation et des séries de réplication. Il vise également à déterminer la prévalence des variations constitutionnelles délétères ou probablement délétères, détectables sous forme de mosaïques. Si cette interprétation aboutit de manière non ambiguë à une démonstration de causalité, après avis d'expert (réunion de concertation pluridisciplinaire d'oncogénétique régionale), le patient sera informé du diagnostic de variant délétère et des conséquences en termes de prise en charge et de conseil génétique.

## II. Matériel et Méthodes

### ❖ Sélection des gènes

Durant le dernier trimestre 2014, après revue extensive des données bibliographiques, nous avons établi une liste de plus de 800 gènes impliqués de manière somatique et/ou constitutionnelle dans la survenue du cancer. Ces gènes ont été extraits de plusieurs sources, comprenant (i) les bases COSMIC recensant les altérations somatiques identifiées dans les cancers, (<http://cancer.sanger.ac.uk/cosmic>) (ii) le portail du *Broad Institute* (<http://www.tumorportal.org>) recensant les données issues des travaux de Lawrence *et al.* <sup>148</sup> (iii) les multiples panels académiques ou commerciaux de gènes impliqués dans le développement des cancers ou faisant l'objet d'un ciblage thérapeutique (ex: Life Comprehensive Cancer) (iv) le panel diagnostique proposé par l'INCa concernant les formes héréditaires de cancer, (<http://www.e-cancer.fr/Professionnels-de-sante/L-organisation-de-l-offre-de-soins/Oncogenetique/Les-predispositions-genetiques>) (v) les bases OMIM interrogées sur la coexistence de formes syndromiques de cancer et d'altérations génétiques, ainsi que les publications majeures relatives aux séquençages par NGS des tumeurs (*The Cancer Genome Atlas; TCGA* <sup>149</sup>). Afin d'obtenir une liste hiérarchisée, nous avons réuni et critiqué les données épidémiologiques, biologiques et fonctionnelles publiées pour l'ensemble de ces gènes, en pondérant les différents arguments (Figure 46). Une liste de 201 *cancer genes* a été ainsi retenue pour définir le set de capture utilisé pour le séquençage (Tableau 7).



**Figure 46 :** Etapes de sélection des 201 gènes impliqués dans le Cancer

Afin de sélectionner les gènes à séquencer à haut-débit, les gènes ont été implémentés à partir d'une analyse bibliographique (notamment via Pubmed), de différentes bases de données d'altérations constitutionnelles et somatiques du cancer, de bases de voies biologiques (KEGG *pathway*) ainsi qu'à partir de panels de gènes déjà existants. Les quelques 800 gènes ainsi présélectionnés ont été cotés selon plusieurs critères comme par exemple le niveau d'implication dans le cancer (somatique ou constitutionnel). Enfin, une cure manuelle a été entreprise pour constituer une liste de 201 gènes impliqués dans le Cancer.

Cancer5000 : panel publié par Lawrence *et al.* <sup>148</sup>

Panel MSK IMPACT : panel *Memorial Sloan Kettering Cancer Integrated Mutation Profiling of Actionable Cancer Targets*

Panel INCa : panel de l'Institut National du Cancer

Panel Life : panel Life Comprehensive Cancer publié par la société Life

TruSeq Cancer Panel: Panel publié par la société Illumina

Panel Qiagen : Panel Clinical Insight for Hereditary Cancer publié par la société Qiagen

Panel HBOC Caen : panel publié par l'équipe caennaise de notre unité pour l'étude des prédispositions génétiques au cancer du sein et de l'ovaire <sup>87</sup>

Panel HPV Caen : panel publié par l'équipe caennaise de notre unité pour l'étude du papillomavirus. <sup>150</sup>

Panel Colon Rouen : panel de notre unité employé pour l'étude des prédispositions héréditaires aux cancers colorectaux.

BDD: Base De Données

**Tableau 7 : Liste des 201 gènes sélectionnés**

AIP	EPHB2	KIT	RAD21
AKT1	ERBB2	KMT2B	RAD50
AKT2	ERCC1	KMT2C	RAD51
AKT3	ERCC2	KRAS	RAD51B
ALK	ERCC3	MALT1	RAD51C
AMER1	ERCC4	MAP3K1	RAD51D
APC	ERCC5	MAX	RAF1
ARAF	EXT1	MDM2	RB1
ARID1A	EXT2	MDM4	RECQL4
ARID5B	FAM157A	MEN1	RET
ATM	FANCA	MET	RINT1
ATR	FANCB	MGMT	RNASEL
ATRX	FANCC	MITF	RUNX1
AURKA	FANCD2	MLH1	SBDS
AXIN2	FANCE	MPL	SDHA
BAP1	FANCF	MRE11A	SDHAF2
BARD1	FANCG	MSH2	SDHB
BCL2	FANCI	MSH6	SDHC
BCL2L11	FANCL	MUTYH	SDHD
BCL9	FANCM	MYC	SH2D1A
BIRC5	FAS	MYCN	SLX4
BLM	FAT1	NBN	SMAD3
BMPR1A	FBXW7	NF1	SMAD4
BRAF	FGFR1	NF2	SMARCA4
BRCA1	FGFR2	NKX2-1	SMARCB1
BRCA2	FGFR3	NOTCH1	SMC1A
BRIP1	FGFR4	NRAS	SMO
BUB1B	FH	NSD1	SOS1
CARD11	FHIT	NTRK1	SPEN
CASP8	FLCN	PALB2	SPOP
CCND1	GATA2	PAX7	SPRED1
CDH1	GNA11	PDGFB	STAG2
CDK4	GNAQ	PDGFRA	STK11
CDKN1A	GNAS	PDGFRB	SUFU
CDKN1B	H3F3A	PHOX2B	TCF7L2
CDKN2A	HNF1A	PIK3CA	TERT
CDKN2B	HOXB13	PIK3CD	TGFBR2
CDKN2C	HRAS	PIK3R1	TMEM127
CHEK2	IDH1	PIK3R2	TNFAIP3
CREBBP	IDH2	PML	TP53
CTNNA1	IGF2	PMS2	TP63
CTNNB1	IKZF1	POLE	TRIM33
CYLD	INHA	POLH	TSC1
DDB1	INHBA	POT1	TSC2
DDB2	INSR	PPM1D	VHL
DICER1	IRF4	PRKAR1A	WRN
DNMT3A	IRS1	PTCH1	WT1
EGFR	JAK1	PTEN	XPA
EGLN1	JAK2	PTPN11	XPC
EPCAM	JAK3	PTPRD	XRCC2
			XRCC3

❖ Patients de l'étude

Les critères d'inclusion à l'étude sont les suivants :

- Age supérieur à 18 ans (au moment de l'étude) ou accord parental ou des responsables légaux en cas de patient mineur
- Tumeur solide pédiatrique d'âge de survenue inférieure à 18 ans OU patients ayant développé une ou plusieurs tumeurs rares à un âge précoce pour lesquelles une recherche de mutation du gène *TP53* a été prescrite (suspicion clinique de syndrome de Li-Fraumeni)
- Non porteurs d'une altération constitutionnelle du gène *TP53* détectée par les techniques diagnostiques usuelles (Sanger, NGS et QMPSF pour la recherche de réarrangements de grande taille)
- Consentement libre et éclairé

Une cohorte de 82 patients ayant développé une ou plusieurs tumeurs à un âge précoce, évocateurs d'un syndrome de Li-Fraumeni pour lesquels aucune base moléculaire n'a pu être mise en évidence a été constituée. Cette dernière est majoritairement constituée de patients ayant développé des tumeurs emblématiques du Li-Fraumeni comme : un rhabdomyosarcome (44 %), un corticosurréalome (24 %) ou une tumeur cérébrale (23 %) (Tableau 8). Il est intéressant de noter que 17 % de ces patients ont développé des tumeurs primitives multiples.

**Tableau 8** : Distribution des phénotypes des patients analysés à l'aide du Panel 201 gènes.

Type de Tumeur Pédiatrique	n	%
Rhaddomyosarcome	36	44%
Corticosurréalome	20	24 %
Tumeur Gliale	6	7 %
Autres tumeurs cérébrales	5	6 %
Médulloblastome	4	5 %
Néphroblastome	3	4 %
Tumeur des plexus choroïdes	2	2 %
Tumeur Germinale	1	1 %
Tératome	1	1 %
Ostéosarcome	1	1 %
Neuroblastome	1	1 %
Pinéoblastome	1	1 %
Cancer du sein	1	1 %
Total	82	100 %

Ce projet a été accepté auprès du Comité de Protection des Personnes Nord Ouest I sous le nom de protocole 2015/160/HP.

❖ Séquençage à haut-débit d'un panel de gènes

Pour la capture des 201 gènes ainsi sélectionnés, un design des sondes a été réalisé grâce au logiciel *SureDesign* de chez Agilent. Les coordonnées génomiques de chacun des gènes, ainsi que les références de chacun des transcrits décrits de ces gènes ont été renseignées. Les exons codants flanqués de 25 paires de bases introniques définissent les séquences d'intérêt, pour lesquelles des sondes sont définies avec un tiling d'au moins 5x. Les régions riches en GC ou connues pour être difficilement capturées sont enrichies en sondes après vérification *in silico* de la position des sondes. Une vérification biologique a été effectuée sur les premiers patients pour confirmer la qualité de la couverture (toutes les régions codantes sont bien séquencées) et les profondeurs obtenues. La taille des régions séquencées représente environ 1 Mb.

Les échantillons d'ADN des patients inclus ont été analysés sur les plateformes de séquençage à haut-débit de Rouen (Inserm U1245) et de Caen (Laboratoire de Biologie Clinique et Oncologique). Ces plateformes comportent des séquenceurs de nouvelle génération de type NextSeq 500 (illumina). Les librairies d'ADN ont été effectuées manuellement ou de manière robotisée selon le protocole SureSelect<sup>QXT</sup> d'Agilent à Rouen. A Caen, les librairies ont été réalisées en SureSelect<sup>XT</sup>. Brièvement, 50 ng d'ADN ont été fragmentés et indexés par tagmentation (utilisation de transposase) permettant leur séquençage en pool. Les fragments d'intérêt sont capturés en phase liquide à l'aide des sondes Agilent, et séquencées de manière paired-end 2x150pb sur séquenceur NextSeq 500.

Des outils bioinformatiques ont été implémentés ou créés pour répondre aux besoins d'analyses bioinformatiques. Le pipeline bioinformatique employé est basé sur l'utilisation des outils BWA-Picard-GATK pour effectuer les étapes de *base calling* et d'*alignement* comme recommandé par le BroadInstitut (*best practice*). L'étape de *variant calling* est réalisée par un double pipeline : GATK HaplotypeCaller et VarScan2. L'utilisation originelle de VarScan 2 (détection de mutation somatique) a ici été détournée pour détecter des mutations présentes en mosaïques. L'annotation des variants a été réalisée par Alamut Batch et par Annovar. Les variations du nombre de copies (CNV) ont été également interrogées à l'aide de l'outil CANOES. Ainsi, les variations ponctuelles, les délétions/insertions, les CNV et les réarrangements de grandes tailles peuvent être détectés à l'aide de ce pipeline.

## ❖ Processus d'analyse

### **Recherche de mutations à fort effet biologique prédit**

Cette étude connaît plusieurs temps d'analyse. Dans le but d'identifier des altérations potentiellement délétères rapidement, nous avons employé un premier filtre qualifié de DSD (*Disruptive and Strictly Damaging*). Ce filtre permet de façon drastique d'identifier des altérations rares perte de fonction ainsi que des altérations rares prédites délétères par 3/3 logiciels de prédiction. Grâce à ce filtre DSD, les variants conservés dans l'analyse sont les variants rares (MAF < 1% dans ExAC\_all, ESP, 1000 genomes), de bonne qualité (profondeur > 20x et GQ > 80) et possédant l'un des effets prédits suivants : (i) perte de fonction comme par exemple les variants décalant le cadre de lecture, les variants non-sens, les variants affectant les sites canoniques d'épissage ou (ii) prédits délétère (*Strictly Damaging*) par trois logiciels de prédiction de pathogénicité sur trois (SIFT, PolyPhen2 HVAR, Mutation Taster). Nous avons également éliminé les variants dont la balance allélique n'était pas comprise entre 20% et 80%. Cet intervalle permet d'identifier des variations constitutionnelles hétérozygotes (avec une balance allélique autour de 50 %  $\pm$  10 % en fonction de la qualité de séquençage). Néanmoins, l'emploi de cette fenêtre, volontairement large, permet l'identification d'événements en mosaïque dont la balance allélique est supérieure à 20 %.

### **Recherche de second événement, de variants à effet moins drastiques et d'altérations en mosaïque**

Le second temps d'analyse a été d'employer une analyse VarScan2 permettant d'identifier l'ensemble des altérations présentes dans les 201 gènes chez un individu. Cette analyse permet par exemple d'identifier les seconds événements moléculaires dans un même gène ou dans un autre gène lorsqu'un mécanisme d'oligogénisme/polygénisme est évoqué chez le patient mais également, l'identification d'événements rares pour lesquels les logiciels de prédiction n'indiquent pas tous un caractère délétère.

### **Recherche de variations du nombre de copies**

Ces analyses ont été poursuivies par une analyse d'identification des réarrangements de grande taille comme les CNV à l'aide de l'outil CANOES. CANOES est basé sur le comptage de reads et est capable d'estimer une variance des lectures entre les échantillons et un ensemble d'échantillons de référence<sup>151</sup>. Cet outil est capable de travailler à la fois à partir de données exomiques et de données issues de panels.



#### ❖ Confirmation des variations identifiées

Toutes les variations potentiellement délétères identifiées au terme du séquençage NGS et de l'analyse bioinformatique font l'objet d'une confirmation par séquençage Sanger. Pour cela, des amplifications par PCR ont été réalisées à partir de 50 à 100 ng d'ADN génomique extrait à partir du sang périphérique. Le séquençage Sanger de ces fragments a ensuite été réalisé à l'aide du kit BigDye® Terminator v3.1 Cycle sequencing (Life Technologies). Les réactions de séquençage ont été séparées par électrophorèse sur séquenceur 3100xl Genetic Analyzer (Life Technologies) et analysées à l'aide du logiciel Sequencing Analysis Software 5.2.0 (Life Technologies).

### III. Résultats

L'objectif principal de ce projet a été la détermination de la prévalence de variations constitutionnelles délétères et/ou probablement délétères dans un panel de 201 gènes impliqués dans la cancérogénèse, au sein des cohortes constituées de patients atteints de cancers à développement précoce et en particulier pour mon projet, de patients ayant développé un cancer avant l'âge de 18 ans.

#### ❖ Identification des mutations à fort effet biologique prédit.

Dans le but d'identifier des mutations à fort effet biologique prédit, nous avons dans un premier temps appliqué un filtre permettant de ne conserver que les altérations rares perte de fonction ou prédites délétères par 3/3 logiciels de prédiction, et présente sur au moins 20 % des reads. Sur les 82 patients analysés, après cure manuelle des artefacts de séquençage, 69 patients (84 %) analysés présentaient à minima une altération DSD (*Disruptive and Strictly Damaging*). En moyenne, les patients présentaient 1,57 altérations DSD (de 0 à 5 altérations maximum). L'ensemble de ces altérations DSD a été discuté en réunion de concertation pluridisciplinaire entre biologistes et cliniciens pour évaluer leur caractère délétère et leur implication possible dans la pathologie. Etant donné qu'il serait impossible de détailler l'ensemble des variants identifiés, la suite de cet exposé sera consacrée à quelques exemples, illustrant la sensibilité du pipeline d'analyse utilisé mais également les difficultés d'interprétation des résultats.

### Exemple n°1 : Association Leucémie myélomonocytaire juvénile, Rhabdomyosarcome et mutation dans le gène *NF1*

Le patient #11 a développé conjointement une **leucémie myélomonocytaire juvénile (LMMJ) et un rhabdomyosarcome** à l'âge de 4 ans sans contexte familial de cancer. Grâce au panel 201 gènes du cancer, nous avons mis en évidence une mutation faux sens du gène ***NF1*** (NM\_000267; c.3827G>A; p.R1276Q). Ce variant est déjà décrit comme étant délétère dans la neurofibromatose de type 1, maladie causant généralement des tumeurs non cancéreuses le long des trajets nerveux : des *neurofibromes*. (<https://databases.lovd.nl/shared/genes/NF1>) La neurofibromatose peut aussi provoquer des tâches « café au lait » et des anomalies osseuses. De plus, Benessahraoui et ses collaborateurs ont décrit qu'il existe un risque accru pour les patients porteurs de mutations *NF1* de développer des leucémies myélomonocytaires juvéniles <sup>152</sup>. L'association d'un rhabdomyosarcome et d'une LMMJ dûe à une mutation du gène *NF1* est une notion évoquée par Matsui et ses collaborateurs depuis 1993 <sup>153</sup>.

Etonnamment chez ce patient cette mutation est présente non pas à 50 % mais avec une balance allélique de 79 % (1815/2285). Nous avons alors évoqué la possibilité d'une contamination blastique du sang lié à la leucémie de ce jeune patient. Grâce à une analyse des autres gènes présents dans le panel, nous avons pu mettre en évidence qu'il s'agissait d'une duplication **somatique** du bras long du chromosome 17 (17q) liée à une contamination par des cellules tumorales leucémiques. En remontant dans le dossier médical de ce patient, nous avons pu confirmer l'existence d'une contamination blastique du sang estimée à 30 %, ce qui est cohérent avec notre résultat (50 % + 30 %). Ce premier exemple ne constitue pas une découverte mais nous permet de valider la sensibilité de nos pipelines d'analyse associés à ce panel.

### Exemple n°2 : Association d'un rhabdomyosarcome et d'une mutation du gène *SOS1*

Le patient #09 a développé sporadiquement un rhabdomyosarcome à l'âge de 11 ans (Figure 47). A l'aide du panel 201 gènes du cancer, un variant altérant un site accepteur d'épissage de l'exon 21 du gène *SOS1* a été détecté (NM\_005633; exon21; c.3347-1G>A). Aucun domaine protéique n'est décrit dans les derniers exons du gène *SOS1*, hormis le fait que ce domaine est riche en proline (PxxP) <sup>154</sup>. Ce variant a été détecté avec une balance allélique proche de 50 %.

Les variants gain de fonction du gène *SOS1* sont associés à une RASopathie, le

syndrome de Noonan. Les patients atteints du syndrome de Noonan ont des signes caractéristiques comme une petite taille, une dysmorphie faciale caractéristique ou des anomalies cardiaques. De plus, les patients atteints de ce syndrome ont généralement une susceptibilité à développer des rhabdomyosarcomes<sup>155,156</sup>. Or, dans ce cas, ce patient ne présente aucun signe clinique d'un syndrome de Noonan.

La destruction du site accepteur d'épissage de l'exon 2, pourrait avoir plusieurs conséquences : la première serait la création d'un nouveau site d'épissage décalé d'une base, l'activation de ce site résulterait en un décalage du cadre de lecture aboutissant à une protéine tronquée (p.Ser1146Thrfs\*23); la deuxième conséquence possible serait un saut complet de l'exon 21 (délétion de 45 pb), provoquant une délétion en phase et la création d'un nouveau codon à la jonction (Figure 48). Un test d'épissage à l'aide d'un minigène ou une RT-PCR à partir de l'ARN du patient permettrait de trancher entre ces deux hypothèses.

Nous avons pu avoir accès à l'ADN des parents non atteints de ce jeune patient et nous avons réalisé le séquençage Sanger de l'exon 21 du gène *SOS1* chez les parents : la mère de ce jeune patient porte également la variation (Figure 49). La présence de l'altération de la variation chez la mère non atteinte pourrait exclure l'hypothèse de l'implication de ce variant dans la pathologie. Cependant, il est possible comme dans beaucoup de formes héréditaires de cancer que la pénétrance des mutations soit incomplète et que la mère ne présente pas de phénotype cancéreux pour cette raison. De plus, de nombreux facteurs peuvent influencer l'épissage et l'expression de ce gène. Ainsi, l'effet de la mutation peut être différent en fonction du fond génétique de l'individu. Il est donc impossible de conclure en l'état sur la pathogénicité de cette altération.

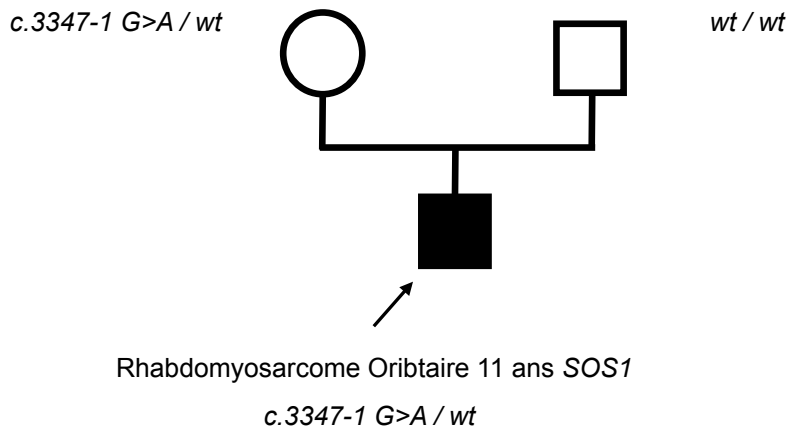
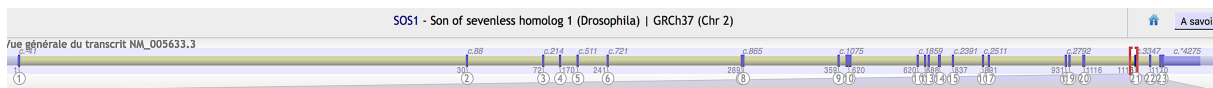


Figure 47 : Arbre génétique du patient #09 porteur de l'altération SOS1 (NM\_005633; exon21; c.3347-1G>A)



Utilisation d'un site cryptique situé  
1 base en avant  
**p.Ser1116Thrfs\*23**

Saut complet de l'exon 21  
**Délétion en phase de 45pb**

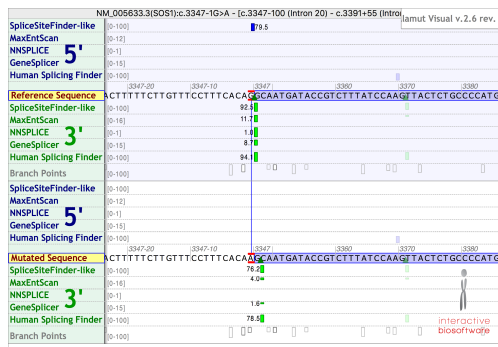
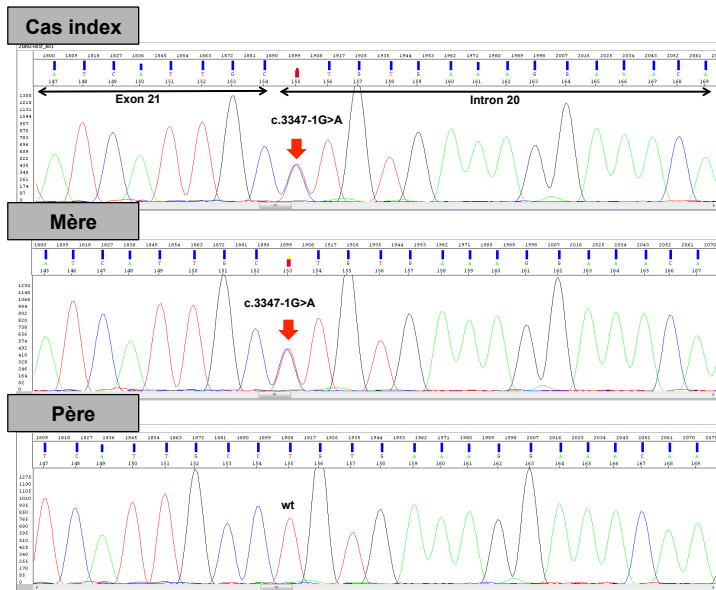


Figure 48 : Illustration des effets probables de l'altération SOS1 identifiée

La variation c.3347-1G>A (NM\_005633) du gène SOS1 pourrait avoir deux effets possibles : soit la création et l'activation d'un nouveau site accepteur d'épissage situé une base en aval du début d'exon entrainant le décalage du cadre de l'ecture aboutissant à la création d'un codon stop 23 acides animés après ou soit un saut complet de l'exon 21 entrainant une délétion en phase de 45pb avec un nouvel acide aminé à la jonction.

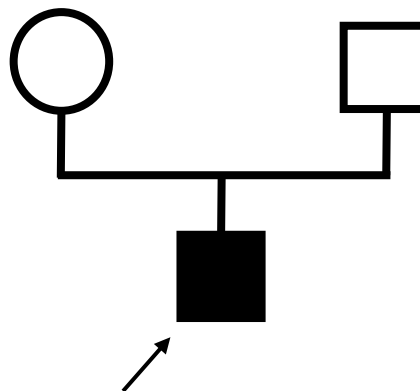


**Figure 49 :** Confirmation de la mutation *SOS1* chez le patient et recherche de cette mutation chez les parents

Capture d'écran des électrogrammes Sanger chez le patient et ses deux parents. La variation *SOS1* c.3347-1G>A (NM\_005633; exon21; c.3347-1G>A) est bien détectée chez l'enfant à l'état hétérozygote par séquençage Sanger. Le séquençage Sanger de cette région a révélé que cette variation a été héritée de sa mère.

### Exemple n°3 : Adénocarcinome colique à 14 ans et mutation du gène *DICER1*

Nous rapportons ici, l'analyse d'un jeune patient ayant développé un adénocarcinome colique sporadique à l'âge de 14 ans (Figure 50). Devant la précocité de cette tumeur, des analyses des gènes *TP53*, *MMR*, *CDH1*, *MUTYH* ont été réalisées. Ces dernières se sont avérées négatives.



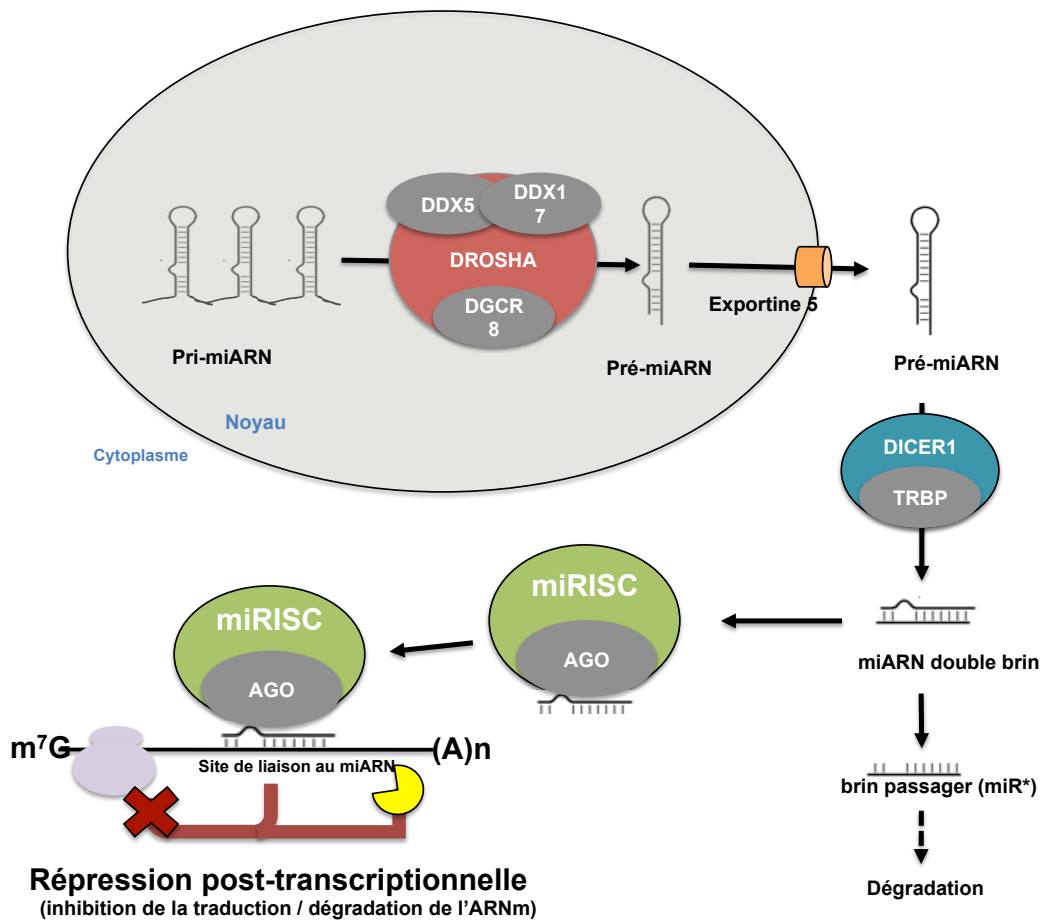
Adénocarcinome colique 14 ans

*DICER1* NM\_001195573 c.2423A>G p.K808R / wt

**Figure 50 :** Arbre génétique du patient analysé

A l'aide du panel, une variation faux-sens à l'état hétérozygote non répertoriée en population générale du gène *DICER1* a été mise en évidence (*DICER1*:NM\_001195573:exon14:c.2423A>G:p.K808R). La protéine DICER1 joue un rôle crucial dans la biogenèse des microARN et est responsable du clivage du précurseur du microARN en microARN mature <sup>157</sup> (figure 51). Les altérations perte de fonction de ce gène sont documentées dans un syndrome de susceptibilité aux blastomes pleuropulmonaires, aux goitres multinodulaires familiaux, aux gynandroblastomes, aux tumeurs malignes de la granulosa de l'ovaire, aux tumeurs malignes de Sertoli-Leydig de l'ovaire, aux kystes rénaux, et plus rarement aux rhabdomyosarcomes, aux sarcomes rénaux, aux blastomes de la glande pituitaire et aux pinéoblastomes <sup>158</sup>.

Bien que cette altération ne soit pas située dans un domaine protéique connu de *DICER1*, cette altération pourrait être à l'origine de la pathologie observée. La variation touche un acide-aminé très conservé jusqu'à la levure. Il est intéressant de noter qu'en 2010, Shaheen et ses collaborateurs rapportent un patient porteur d'une mutation *DICER1* âgé de 3 ans ayant développé des polypes hamartomateux intestinaux ainsi qu'un néphrome kystique bilatéral ainsi qu'un blastome pleuropulmonaire à l'âge de 9 mois <sup>159</sup>. Il est donc possible que ce phénotype extrême puisse résulter d'une mutation de *DICER1*, ce qui pourrait étendre le spectre des phénotypes associés aux mutations de ce gène <sup>157</sup>. L'ADN des parents a été demandé afin de vérifier l'absence de cette variation chez les parents. Nous avons également demandé l'ADN tumoral pour rechercher une perte d'hétérozygotie à ce locus, qui accompagne généralement les mutations constitutionnelles de ce gène.



**Figure 51 :** Biogenèse et maturation des microARN

Les gènes des microARN sont majoritairement transcrits par l'ARN polymérase II sous forme de miARN primaire (pri-ARN). Ils sont ensuite maturés par le complexe enzymatique Drosha en pré-miARN. Le précurseur miARN (pré-miARN) est exporté du noyau vers le cytoplasme via l'Exportine 5. Dans le cytoplasme, il subit une deuxième étape de maturation par le complexe Dicer. Le microARN est alors incorporé dans le complexe effecteur RISC (*RNA-induced Silencing Complex*). Sous cette forme, le microARN est capable de reconnaître des séquences cibles complémentaires, le plus souvent dans le 3'UTR des ARN messagers et d'induire une répression post-transcriptionnelle de ces messagers (répression de la traduction, dégradation des transcrits via déadénylation et décapping suivis d'une dégradation par des exonucléases).

#### Exemple n°4 : Corticosurréalome et kystes desmoïdes de l'ovaire à l'âge de 20 ans et mutation du gène *DICER1*

Nous rapportons, ici, une patiente ayant présenté, en dehors d'un contexte familial de cancer, un corticosurréalome sporadiquement à l'âge de 20 ans ainsi que des kystes desmoïdes de l'ovaire au même âge, sans qu'aucune base moléculaire n'ait pu être mise en évidence. Bien que cette patiente n'ait pas développé une tumeur pédiatrique *stricto sensu*, le phénotype présenté et l'absence d'altération détectable dans le gène *TP53* nous a amené à réaliser l'analyse du panel 201 gènes du cancer chez cette patiente. Cette dernière a révélé la présence d'une variation faux-sens à l'état hétérozygote dans l'exon 16 du gène *DICER1* (*DICER1*:NM\_001195573; c.2660C>A; p.S887Y). Ce variant a déjà été détecté une fois dans la base de données ExAC en population européenne uniquement (1/120694 allèles, soit présent une fois à l'état hétérozygote chez 60,000 personnes). Malheureusement, comme pour le variant précédent, cette altération n'est pas située dans un domaine connu de la protéine *DICER1*. Néanmoins, en plus des prédictions délétères des logiciels de prédiction, il est intéressant de noter que le gène *DICER1* possède un *z-score* de 4,07 (ExAC). Ce score rend compte du nombre de variants attendus comparé au nombre de variants observés chez les 60 000 individus séquencés. En pratique, cela signifie que le gène *DICER1* semble peu tolérant aux changements d'acides aminés. De plus, le nucléotide impacté est très conservé dans l'évolution. Ainsi, l'ensemble de ces éléments plaide en faveur d'une mutation à impact potentiel sur la protéine. En conclusion, ce variant semble extrêmement intéressant dans le cadre également d'une extension phénotypique des mutations *DICER1*.

#### ❖ Recherche de mutations en mosaïque

L'utilisation du panel avait également pour objectif de détecter des altérations en mosaïque grâce à une forte profondeur de lecture. En parallèle de l'analyse décrite ci-dessus, nous avons donc systématiquement recherché les altérations en mosaïque avec une attention particulière pour les gènes *TP53* et *DICER1* étant donné la forte contribution de ces gènes dans les cancers pédiatriques. Cette analyse a été réalisée à l'aide du logiciel VarScan2. Nous n'avons pas pu détecter de mosaïcisme dans ces gènes ou d'autres, dans la cohorte pédiatrique. Néanmoins, la capacité de notre pipeline a détecter de telles altérations a cependant pu être démontré pour la mise en évidence d'une altération en mosaïque du gène *TP53* chez une femme ayant développé un cancer du sein avant 31 ans (voir Annexe).



#### ❖ Recherche des variations du nombre de copies (CNV)

A l'aide du logiciel CANOES, nous avons systématiquement recherché la présence de variations du nombre de copies dans les 201 gènes du panel. Treize CNV potentiels ont été détectés dans la cohorte pédiatrique. Seuls trois d'entre eux possédaient un score de confiance suffisamment élevé (SQ > 90) : les CNV impactant les gènes *FAM157A*, *CHEK2* et *FANCA*. Le CNV affectant *FAM157A* semble être un artefact sûrement dû à la présence de paralogues de ce gène comme le gène *FAM157B*. La séquence dupliquée du gène *FAM157A* (chr3:197,894,512-197,896,797) est d'ailleurs homologue à 98 % à une séquence de *FAM157B*. La délétion affectant les exons 9 et 10 du gène *CHEK2* semble bien présente après analyse sur les fichiers BAM. Néanmoins, elle est déjà décrite comme polymorphique dans la population slave et notamment polonaise (0,4 %) <sup>160</sup>. Enfin, la délétion partielle du gène *FANCA* semble être un artefact sur la base de l'analyse visuelle des fichiers BAM. Ainsi, malheureusement, l'analyse des CNV dans la cohorte pédiatrique s'est révélée infructueuse.

## IV. Conclusion, discussion, perspectives

En conclusion, le séquençage de 201 gènes du cancer par panel est un outil sensible pour identifier des variations dans des gènes relatif au cancer. Dans le premier temps d'analyse, seuls les variants avec un impact délétère prédit ont été conservés à l'aide d'un filtre drastique (filtre DSD). Cette analyse a permis de mettre en évidence 129 variants potentiellement causaux chez 69 patients, tous discutés plus finement en réunion de concertation pluridisciplinaire. Dans les faits, nous avons pu démontrer la robustesse de notre pipeline d'analyse en identifiant une altération à une forte balance allélique (80 %) du fait d'une contamination tumorale leucémique. Parmi les 129 variants prédits délétères, on compte :

- (i) des altérations qui ne semblent pas contribuer directement au phénotype observé, et sont finalement le reflet de la variabilité de nos gènes ou encore peuvent-être des facteurs modificateurs dans des modèles plus complexes d'oligogénisme que notre effectif ne permet pas d'évaluer;

- (ii) des variations qui semblent contribuer au phénotype mais dont la démonstration n'est pas évidente à l'instar de l'identification d'une variation à l'état hétérozygote du gène *SOS1* chez un patient ayant développé un rhabdomyosarcome sans signe clinique de syndrome de Noonan,

- (iii) des variants qui laissent penser à un chevauchement phénotypique comme cela est le cas avec l'identification de deux altérations du gène *DICER1* dans des cas de tumeurs précoces.

La première conclusion est qu'au sein de notre cohorte, l'identification de variants clairement délétères à l'aide du panel 201 gènes du cancer n'est pas aussi fréquente que ce qui est décrit dans la littérature, à savoir autour de 10 %<sup>95</sup>. En effet, seuls deux patients de notre cohorte présentent une altération clairement démontrée comme délétère soit 2,5 % (2/82). Cependant, en adjoignant les altérations potentiellement délétères dont l'effet n'est pas encore démontré, le taux de détection d'altération est d'environ 7 %. Ce taux est similaire à celui obtenu par l'équipe de Zhang en 2015<sup>130</sup>, qui pour rappel, avait criblé 1000 patients ayant développé une tumeur pédiatrique à l'aide de panels *in silico* à partir de données de WES et WGS et obtenait un taux de détection autour de 8,5 % .

Parmi les altérations détectées dans ce projet, on compte deux variants faux-sens du gène *DICER1*. Si la démonstration de la causalité des altérations identifiées dans ce projet de *DICER1* est concluante, le spectre phénotypique pourra être étendu. En effet, nous avons à deux reprises identifié des variations dans le gène *DICER1* chez des patients pédiatriques dont le phénotype n'est pas celui décrit dans la littérature (adénocarcinome colique et un corticosurréalome). Cependant, le spectre phénotypique associé aux altérations de *DICER1* est large à l'instar du spectre du gène *TP53*. Il est donc possible que nous soyons en présence d'une extension phénotypique, ce qui rejoint l'idée d'un nombre limité de gènes impliqués dans le cancer comme préalablement décrit et comme nous l'avons démontré dans le contexte du chondrosarcome à l'aide de l'exome. Etant donné que nous avons employé un filtre d'analyse drastique (filtre DSD), nous avons quand même pris soin de rechercher dans la cohorte pédiatrique la présence d'autres événements moléculaires du gène *DICER1* y compris en mosaïque à l'aide d'une analyse VarScan2. Cependant, cette analyse n'a pas révélé de variant supplémentaire. Néanmoins, il serait quand même particulièrement relevant de cribler le gène *DICER1* chez plus de patients ayant développé des tumeurs pédiatriques et pour lesquels aucune base moléculaire n'a pu être mise en évidence pour compléter ce travail.

D'autre part, il est intéressant de souligner qu'une part non négligeable des autres patients présentent un événement moléculaire prédit délétère par trois logiciels sur trois dans un gène de réparation de l'ADN. Il n'est pas exclu qu'il s'agisse d'un biais dans la sélection originelle des gènes du panel. Néanmoins, 31 patients présentent une altération prédite délétère par 3 logiciels sur 3 dans un gène de réparation, 10 présentent une altération dans deux gènes de réparation de l'ADN et deux patients présentent une altération dans trois gènes de réparation de l'ADN. Ainsi, 52 % (43/82) des patients présentent donc à minima

une altération prédite délétère dans un gène de réparation de l'ADN. Or, dans la littérature, seulement 6 % des variants des gènes de réparation de l'ADN décrits dans la base de données ExAC sont prédites délétères par PolyPhen2, un des trois logiciels employés dans notre analyse <sup>161</sup>. Donc, *a priori*, y compris en étant plus discriminant, il semble y avoir un grand nombre de variants rares prédits délétères dans ces gènes dans notre cohorte. Pour appuyer cette hypothèse, nous souhaitons (i) compiler l'ensemble des données de séquençage du panel au sein d'une base de données et (ii) réaliser une analyse statistique cas/contrôle pour détecter des enrichissements en variants rares dans ces gènes. Cette analyse statistique, réalisée par la statisticienne de notre laboratoire Camille Charbonnier, vient de débuter (Juin 2017). Pour réaliser cette étude, nous avons accès à la cohorte française d'exomes témoins (projet FREX) qui devrait rapidement livrer les premiers résultats. A l'heure actuelle, une curation de notre cohorte a été entreprise pour ne pas fausser l'analyse statistique. Cela passe par l'exclusion de prélèvements éloignés génétiquement (ethnies), l'exclusion de variants de mauvaise qualité (exclusion des bases lues moins de 20 fois), mais également l'analyse du nombre de singletons et de variants rares par échantillon, de variants fréquents par échantillon.

En parallèle du filtre DSD, j'ai pu réaliser une analyse de l'ensemble des variants présents dans les 201 gènes du panel, notamment à la recherche de mosaïque grâce à l'outil VarScan2. Cette analyse a été conduite à titre systématique à l'ensemble de la cohorte et qui est malheureusement restée sans succès. J'ai également pu évaluer la présence de CNV dans la cohorte, analyse qui s'est révélée également infructueuse.

En perspective, grâce à la réduction drastique du coût du séquençage à haut-débit, de l'augmentation de leurs qualités et de leurs efficacités, il est maintenant possible de réaliser de nombreux exomes et donc de réaliser des analyses *in silico* d'un panel de gènes sélectionnés à partir des données de WES. Ainsi, ce projet sera étendu à des exomes pour réaliser une analyse *in silico* avec une version V2 du panel. La version V2 contiendra plus de gènes et notamment des gènes centrés autour de la réparation de l'ADN. Ainsi, 450 patients devraient bénéficier d'un séquençage d'exome prochainement.

**DISCUSSION GÉNÉRALE  
ET  
PERSPECTIVES**

## Discussion générale et perspectives

Depuis leur émergence, les nouvelles technologies de séquençage ont permis des révolutions scientifiques conceptuelles considérables, particulièrement en Génétique Humaine. Son utilisation en tandem avec des techniques d'enrichissement en régions génomiques d'intérêt permet aujourd'hui d'analyser facilement et rapidement l'ensemble des régions codantes du génome humain. Ces avancées, à coût sans cesse décroissant, ont permis une entrée massive de ces technologies dans les laboratoires de recherche et de diagnostic, si bien qu'en moins de 10 ans, elles ont bouleversé nos connaissances, métamorphosé le paysage de nos laboratoires, et modifié les pratiques diagnostiques.

Grâce à notre proximité avec le Laboratoire de Génétique Moléculaire du CHU de Rouen, nous disposons d'un accès à un grand nombre d'échantillons biologiques, et notamment d'ADN de patients évocateurs d'un syndrome de Li-Fraumeni (LFS) pour lesquels aucune base moléculaire n'a pu être mise en évidence. Cette collection issue de plus de 20 années de recrutement représente donc une opportunité unique d'identifier de nouveaux événements moléculaires à l'origine de prédispositions génétiques au cancer. Ainsi, l'objectif de ce projet de recherche était d'évaluer la pertinence des nouveaux outils pangénomiques comme le séquençage à haut-débit de l'exome et d'un panel de gènes pour identifier les altérations constitutionnelles responsables de ces histoires personnelles/familiales de cancers à survenue précoce.

### **L'exome, un outil puissant pour l'identification de variants**

Les résultats obtenus pendant ma thèse confirment plusieurs éléments préalablement évoqués dans la partie introductive de ce document. Le premier élément est la **puissance du séquençage d'exome** pour conduire à l'identification de variants causaux y compris à partir d'un nombre restreint de patients pour l'identification de variants à l'origine de pathologies Mendéliennes comme les cancers héréditaires; comme cela a été le cas dans le premier projet de ma thèse consacré à l'identification de nouveaux gènes de prédisposition aux chondrosarcomes familiaux. Alors qu'aucune base moléculaire n'avait pu être mise en évidence par des techniques conventionnelles, en employant une stratégie exomique comparative intrafamiliale, nous avons pu à partir du séquençage exomique de seulement deux apparentés atteints mettre en évidence les bases moléculaires à l'origine de cette présentation atypique. Alors qu'aucune signe clinique d'Ostéochondromes Multiples n'a pu être détectée chez cette famille, nous avons mis en évidence une mutation tronquante du gène *EXT2*, co-ségrégant avec la pathologie. Cette découverte nous a permis de proposer

un conseil génétique adapté à cette famille et un suivi des apparentés porteurs de l'altération. D'autre part, cette découverte nous a permis d'élargir le spectre phénotypique des mutations du gène *EXT2* aux chondrosarcomes familiaux. Ainsi, il semble exister à minima trois phénotypes : (i) la maladie des Ostéochondromes Multiples (OM) en présence d'altérations hétérozygotes perte de fonction du gène *EXT2*, (ii) le syndrome de convulsions-scoliose-macrocéphalie causé par l'inactivation bi-allélique constitutionnelle de *EXT2*, sans signe de OM (iii) et enfin, une prédisposition au chondrosarcomes familiaux liée également à des altérations hétérozygotes perte de fonction du gène *EXT2*.

Étant donné, que nous avons identifié un potentiel facteur modificateur en *cis*, nous émettons l'hypothèse que trois phénotypes distincts existeraient en fonction des niveaux protéiques d'*EXT2*. Dans les cas de mutations bi-allélique du gène *EXT2*, les niveaux protéique d'*EXT2* sont *a priori* résiduels voire nuls et conduiraient au syndrome de convulsions-scoliose-macrocéphalie. Dans le cas des altérations hétérozygotes perte de fonction du gène *EXT2*, les niveaux protéiques sont *a priori* diminués de moitié et conduirait à l'OM. Ici, nous décrivons une situation intermédiaire où malgré une mutation perte de fonction hétérozygote du gène *EXT2*, les patients ne présentent aucun signe clinique d'OM. Il est possible que les variants introniques en *cis* que nous avons identifié modifient l'épissage et donc modifient la quantité résiduelle de la protéine *EXT2* conduisant à un troisième phénotype différent de l'OM.

D'un point de vue conceptuel, ces travaux démontrent la force de l'exome et de l'approche comparative intrafamiliale dans le contexte d'histoires familiales de maladies Mendéliennes jusqu'alors inexplicables et apporte un exemple d'extension phénotypique développée dans la partie introductive de ce document. Néanmoins, cette stratégie d'analyse comparative intrafamiliale est limitée par le côté exceptionnel de ces grandes familles présentant des tumeurs rares.

Ces travaux ont également illustré la notion de **découverte additionnelle**, évoquée dans la partie introductive de ce document. L'identification de la mutation *de novo* du gène *MYO7A* chez le jeune patient décrit dans le projet 2 a permis une prise en charge adaptée de sa surdit , pr alablement attribu e   son traitement par radioth rapie. Il s'agit effectivement d'une d couverte fortuite mais modifiant substantiellement la mise en  uvre de mesures de soin : il peut  tre maintenant propos  un conseil g n tique   ce patient et   sa descendance. L'essor de nouvelles technologies et notamment l'arriv e imminente de l'exome diagnostique soul ve la question de la gestion de ces  v nements et leurs stipulations dans le consentement pr alable   l' tude.

De toute évidence, l'exome et ses différentes applications s'avèrent être un outil puissant et rapide pour identifier de nouvelles bases moléculaires des prédispositions génétiques au cancer. Le séquençage des régions codantes semble répondre à de nombreux projets, cependant, il est intéressant de noter que le nombre de publications employant l'exome dans le contexte des cancers héréditaires est quant à lui paradoxalement limité. Plusieurs éléments pourraient expliquer cette difficulté. Il est possible de citer l'hétérogénéité tumorale, une pénétrance incomplète des mutations et la présence de phénocopies. Au vu de la littérature, la stratégie d'analyse intrafamiliale est sûrement aujourd'hui la méthode produisant le plus de résultats pour l'identification de nouvelles bases moléculaires à partir d'un nombre restreint d'individus, car fournissant à la fois l'identification du variant et une partie de l'analyse de co-ségrégation avec la pathologie.

### **Le grand défi de l'interprétation des variants**

La deuxième conclusion de ces travaux est la très grande difficulté d'interprétation des variants détectés par séquençage à haut-débit. Malgré une facilité d'identification fournie par ces nouveaux outils, l'interprétation demeure le facteur limitant et notamment lorsque le gène n'est pas préalablement impliqué dans le cancer. Ceci est particulièrement notable avec l'emploi de l'exome, qui est une technique sans *a priori* et donc chacun des 20 000 gènes peut être incriminé.

La démonstration de la pathogénicité du variant peut s'avérer ardue lorsque la protéine n'est pas ou très peu caractérisée. En sus, il est parfois délicat d'établir le lien entre l'impact protéique et une pathologie tumorale. Durant ma thèse, nous avons été confronté à cette difficulté lors de l'interprétation des variants et notamment des deux variants identifiés dans le gène *HID-1*. Bien que nous ayons identifié une mutation *de novo* et une autre mutation chez un patient non apparenté présentant le même phénotype, la relevance biologique des variants identifiés dans le gène *HID-1* est encore en cours d'appréciation malgré un nombre important de manipulations, ce qui confirme que le défi de la Génétique Moléculaire de demain n'est plus l'identification des variations mais bien leur interprétation. Étant donné que la moitié des variants identifiés dans des cohortes importantes sont en réalité des *singletons*, les stratégies d'interprétation sont amenées à changer, ne pouvant plus être uniquement basées sur la récurrence. Une solution réside sûrement dans l'emploi de cohortes importantes afin de réaliser des études de cas-contrôles et de quantifier la charge en variants rares dans des gènes de susceptibilité.

## **Les limites du séquençage d'exome**

Bien que l'exome soit un outil puissant pour l'identification de variants, cela n'est pas toujours le cas comme nous l'avons malheureusement expérimenté dans le projet consacré aux corticosurrénales de l'enfant. Cela s'explique en partie, par la nature même de l'exome, focalisée sur les exons. Dans les faits, beaucoup d'autres régions peuvent être responsables de pathologies Mendéliennes comme les régions non-codantes, les régions régulatrices et plus largement certaines régions inter-géniques. De plus, l'exome n'est pas prévu pour la détection d'événements structuraux comme les inversions, les translocations ou les réarrangements équilibrés bien que de nombreux développements aillent dans ce sens comme de nouveaux outils bio-informatiques (CANOES, CNVseq...), de nouveaux kits de capture possédant des sondes réparties le long du génome (ex : kit OneSeq d'Agilent) ou des solutions de séquençage de troisième génération.

Le séquençage de génome (WGS) pourrait répondre en partie à ces problématiques. En effet, l'absence de capture permet de détecter les CNV, inversions et translocations notamment. Cependant, son coût reste encore élevé et les infrastructures informatiques non adaptées. Néanmoins, il y a fort à parier que son utilisation devienne incontournable par la suite.

Une autre limite du WES est la difficulté de détecter les altérations en mosaïque. En effet, les profondeurs de lecture employées dans le WES sont généralement de l'ordre de 100 x. Ainsi, un variant présent à un faible taux dans un échantillon risque d'être éliminé dans la stratégie de filtration au risque d'être confondu avec un variant de mauvaise qualité. Une des solutions consisterait à employer par exemple le tissu sain adjacent lorsque ce dernier est disponible et non le sang.

## **Le séquençage de panels de gènes : une solution alternative au WES**

La sélection de gènes à inclure dans un panel peut être discutable et/ou critiquable en fonction du projet. Néanmoins, en attendant une future disponibilité du WGS et de WES à fortes profondeurs, le séquençage à haut-débit de panels de gènes représente une solution intermédiaire pertinente. En effet, de par des enrichissements plus restreints, il est possible d'obtenir des profondeurs de lecture plus importantes à coûts maîtrisés et donc de détecter des événements comme les mosaïques. De plus, la taille de capture réduite rend possible le séquençage d'un grand nombre de patients au sein du même run de séquençage. Dans le contexte des prédispositions génétiques au cancer, ce nouvel outil possède de nombreux avantages comme la possibilité d'ajouter rapidement un nouveau gène dans un panel. L'interprétation fonctionnelle est généralement facilitée, car les variations identifiées sont



d'emblée dans des gènes d'intérêt. Par cette approche, il est donc possible de réaliser des diagnostics différentiels ainsi que de mettre en évidence des chevauchements phénotypiques. D'ailleurs, nos premiers résultats semblent aller dans ce sens avec l'identification de variants dans le gène *DICER1*.

Comme évoqué dans la partie introductive de ce document, le panel *in silico* à partir des données d'exome est une approche nouvelle et intéressante. En effet, il est possible d'ajouter à façon et de façon rétrospective, un ou plusieurs nouveaux gènes d'intérêt. C'est d'ailleurs, le futur que nous souhaitons donner à notre projet de panel 201 gènes du cancer. Il est important de souligner que cette approche permet la correction de l'un des biais d'un panel de gènes à savoir la sélection des gènes en amont et de façon figé dans le temps.

### **Remarques concluantes**

Le défi de la génomique humaine n'est plus, compte-tenu de l'évolution technologique, la détection des altérations mais leur interprétation. Avant le développement du NGS et le passage à l'ère génomique, l'identification d'un gène impliqué dans une maladie mendélienne était basée sur la liaison génétique conduisant à une cartographie génétique suivie d'une cartographie physique. La possibilité de séquencer de façon massive représente un changement de paradigme puisque le NGS permet d'identifier des mutations rares chez les patients suspectés de présenter des maladies génétiques. Cependant, compte-tenu de la découverte de la variabilité insoupçonnée du génome humain, nous avons appris que la détection d'un variant rare voire privé, même survenu *de novo*, est un argument totalement insuffisant pour conclure à sa causalité dans le phénotype observé chez le patient. L'interprétation de ces variations rares nécessite de disposer d'un faisceau d'arguments incluant, outre la confrontation à la présentation clinique du patient, des prédictions bioinformatiques à la fois de l'impact potentiel de la variation sur la protéine mais également sur l'ARNm et en particulier sur l'épissage, des analyses fonctionnelles, des analyses de ségrégation et des analyses statistiques adaptées. Ce dernier point est crucial car nous savons aujourd'hui qu'entre les maladies monogéniques et les maladies multifactorielles, l'oligogénisme joue un rôle sans doute important dans le déterminisme génétique des maladies humaines, en particulier celles à révélation précoce. La démonstration de l'implication dans une maladie de variants rares, isolément ou en combinaison, nécessite de démontrer avec des tests statistiques appropriés dédiés aux variations rares, tests agrégeant ces variations par gène ou par voie biologique, un enrichissement par rapport à des sujets contrôles appropriés et appariés. Or en 2017, la collectivité internationale ne dispose aujourd'hui que de la collection ExAC ce qui introduit d'emblée un biais. Il est indispensable de disposer dans chaque pays de bases de données

répertorient les variations rares dans les différents groupes de populations. Dans ce contexte, le projet FREX et son développement seront déterminants pour analyser de façon pertinente les variations détectées chez les patients vivants en France.

Le deuxième challenge sera, une fois l'implication de ces variations rares démontrée, de quantifier le risque de maladie associé à ces variations. L'identification d'une variation causale dans le cadre d'une maladie génétique permet de poser le diagnostic de certitude permettant une prise en charge médicale personnalisée du patient, une prévention des complications et du conseil génétique permettant de lever une inquiétude illégitime et une surveillance injustifiée chez sujets non porteurs et à l'inverse d'assurer une prévention voire une détection de la maladie ou de ses complications chez les sujets porteurs. Compte-tenu de ces impacts, la quantification du risque associé à ces variations est donc cruciale. Cette question a sans doute été négligée depuis le développement du NGS qui a conduit de façon trop hâtive à inclure dans des panels qualifiés de « panels diagnostiques » des gènes dont les variations confèrent un risque modéré et qui ne répondent pas un déterminisme monogénique.

Le développement du séquençage de nouvelle génération et la possibilité d'explorer l'exome et demain le génome représentent un saut technologique sans précédent dans l'histoire de la génétique humaine et de la génétique médicale. Pour que cette révolution technologique soit une véritable révolution médicale au service des patients, il sera essentiel, au-delà des développements technologiques, que les enjeux d'interprétation et de connaissance des risques associés à ces variations soient pris en considération à leur juste valeur.



## **REFERENCES**

## Références bibliographiques

1. Garber, J. E. & Offit, K. Hereditary cancer predisposition syndromes. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **23**, 276–292 (2005).
2. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).
3. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
4. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21**, 846–853 (2015).
5. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170–173 (1976).
6. Hofstra, R. M. W. *et al.* A mutation in the RET proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature* **367**, 375–376 (1994).
7. Kihara, M. *et al.* Germline RET mutation carriers in Japanese patients with apparently sporadic medullary thyroid carcinoma: A single institution experience. *Auris. Nasus. Larynx* **43**, 551–555 (2016).
8. Donis-Keller, H. *et al.* Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum. Mol. Genet.* **2**, 851–856 (1993).
9. Mulligan, L. M. *et al.* Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* **363**, 458–460 (1993).
10. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
11. Cavenee, W. K. *et al.* Prediction of familial predisposition to retinoblastoma. *N. Engl. J. Med.* **314**, 1201–1207 (1986).
12. Dunn, J. M., Phillips, R. A., Becker, A. J. & Gallie, B. L. Identification of germline and somatic mutations affecting the retinoblastoma gene. *Science* **241**, 1797–1800 (1988).
13. Dyson, N. J. RB1: a prototype tumor suppressor and an enigma. *Genes Dev.* **30**, 1492–1502 (2016).
14. Malkin, D. *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233–1238 (1990).
15. Srivastava, S., Zou, Z. Q., Pirolo, K., Blattner, W. & Chang, E. H. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* **348**, 747–749 (1990).
16. Friedberg, E. C. DNA damage and repair. *Nature* **421**, 436–440 (2003).
17. Das, R. & Ghosh, S. K. Genetic variants of the DNA repair genes from Exome Aggregation Consortium (EXAC) database: significance in cancer. *DNA Repair* **52**, 92–102 (2017).
18. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer* **15**, 181–194 (2015).
19. Ripperger, T. *et al.* Childhood cancer predisposition syndromes-A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *Am. J. Med. Genet. A.* (2017). doi:10.1002/ajmg.a.38142
20. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
21. Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
22. Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215–1225 (1993).
23. Miyaki, M. *et al.* Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer.

*Nat. Genet.* **17**, 271–272 (1997).

24. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
25. Rahman, N. & Scott, R. H. Cancer genes associated with phenotypes in monoallelic and biallelic mutation carriers: new lessons from old players. *Hum. Mol. Genet.* **16 Spec No 1**, R60–66 (2007).
26. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. *Science* **265**, 2088–2090 (1994).
27. Howlett, N. G. *et al.* Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**, 606–609 (2002).
28. Singh, S. *et al.* Sonic Hedgehog Mutations Identified in Holoprosencephaly Patients Can Act in a Dominant Negative Manner. *Hum. Genet.* **125**, 95–103 (2009).
29. Bresler, S. C., Padwa, B. L. & Granter, S. R. Nevroid Basal Cell Carcinoma Syndrome (Gorlin Syndrome). *Head Neck Pathol.* **10**, 119–124 (2016).
30. Carlomagno, F. *et al.* The different RET-activating capability of mutations of cysteine 620 or cysteine 634 correlates with the multiple endocrine neoplasia type 2 disease phenotype. *Cancer Res.* **57**, 391–395 (1997).
31. Hyndman, B. D., Gujral, T. S., Krieger, J. R., Cockburn, J. G. & Mulligan, L. M. Multiple functional effects of RET kinase domain sequence variants in Hirschsprung disease. *Hum. Mutat.* **34**, 132–142 (2013).
32. Edkins, S. *et al.* Recurrent KRAS codon 146 mutations in human colorectal cancer. *Cancer Biol. Ther.* **5**, 928–932 (2006).
33. Pollock, P. M. & Meltzer, P. S. Lucky draw in the gene raffle. *Nature* **417**, 906–907 (2002).
34. Feinberg, A. P., Vogelstein, B., Droller, M. J., Baylin, S. B. & Nelkin, B. D. Mutation affecting the 12th amino acid of the c-Ha-ras oncogene product occurs infrequently in human cancer. *Science* **220**, 1175–1177 (1983).
35. Tartaglia, M. *et al.* Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat. Genet.* **29**, 465–468 (2001).
36. Tartaglia, M. *et al.* PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. *Am. J. Hum. Genet.* **70**, 1555–1563 (2002).
37. Legius, E. *et al.* PTPN11 mutations in LEOPARD syndrome. *J. Med. Genet.* **39**, 571–574 (2002).
38. Evers, B. & Jonkers, J. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. *Oncogene* **25**, 5885–5897 (2006).
39. Ballinger, M. L. *et al.* Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol.* **17**, 1261–1271 (2016).
40. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
41. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).
42. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463–5467 (1977).
43. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
44. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
45. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
46. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
47. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
48. Mitra, R. D., Shendure, J., Olejnik, J., Edyta-Krzyszowska-Olejnik & Church, G. M. Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **320**, 55–65 (2003).

49. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
50. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. & Nyrén, P. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Anal. Biochem.* **242**, 84–89 (1996).
51. Ronaghi, M., Uhlén, M. & Nyrén, P. A Sequencing Method Based on Real-Time Pyrophosphate. *Science* **281**, 363–365 (1998).
52. Nyrén, P. The history of pyrosequencing. *Methods Mol. Biol. Clifton NJ* **373**, 1–14 (2007).
53. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
54. Chan, M. *et al.* Development of a next-generation sequencing method for BRCA mutation screening: a comparison between a high-throughput and a benchtop platform. *J. Mol. Diagn. JMD* **14**, 602–612 (2012).
55. Wall, J. D. *et al.* Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res.* **24**, 1734–1739 (2014).
56. Loose, M., Malla, S. & Stout, M. Real-time selective sequencing using nanopore technology. *Nat. Methods* (2016). doi:10.1038/nmeth.3930
57. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
58. Chaisson, M. J. P. *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2015).
59. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
60. Mignardi, M. & Nilsson, M. Fourth-generation sequencing in the cell and the clinic. *Genome Med.* **6**, 31 (2014).
61. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
62. Moorthie, S., Mattocks, C. J. & Wright, C. F. Review of massively parallel DNA sequencing technologies. *HUGO J.* **5**, 1–12 (2011).
63. Mankos, M. *et al.* A novel low energy electron microscope for DNA sequencing and surface analysis. *Ultramicroscopy* **145**, 36–49 (2014).
64. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203 (2011).
65. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
66. Schatz, M. C. & Langmead, B. The DNA Data Deluge: Fast, efficient genome sequencing machines are spewing out more data than geneticists can analyze. *IEEE Spectr.* **50**, 26–33 (2013).
67. Bashiardes, S. *et al.* Direct genomic selection. *Nat. Methods* **2**, 63–69 (2005).
68. Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* **21**, 673–678 (2003).
69. Cantsilieris, S., Stessman, H. A., Shendure, J. & Eichler, E. E. Targeted Capture and High-Throughput Sequencing Using Molecular Inversion Probes (MIPs). *Methods Mol. Biol. Clifton NJ* **1492**, 95–106 (2017).
70. Samorodnitsky, E. *et al.* Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum. Mutat.* **36**, 903–914 (2015).
71. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
72. Vissers, L. E. L. M. *et al.* A de novo paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
73. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).

74. Dorschner, M. O. *et al.* Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am. J. Hum. Genet.* **93**, 631–640 (2013).
75. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
76. Yuen, R. K. C. *et al.* Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* **21**, 185–191 (2015).
77. Francioli, L. C. *et al.* Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47**, 822–826 (2015).
78. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
79. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nat. Rev. Genet.* **13**, 565–575 (2012).
80. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
81. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
82. Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **32**, 2001–2009 (2014).
83. Fokstuen, S. *et al.* A DNA resequencing array for pathogenic mutation detection in hypertrophic cardiomyopathy. *Hum. Mutat.* **29**, 879–885 (2008).
84. Hoppman-Chaney, N. *et al.* Evaluation of Oligonucleotide Sequence Capture Arrays and Comparison of Next-Generation Sequencing Platforms for Use in Molecular Diagnostics. *Clin. Chem.* **56**, 1297–1306 (2010).
85. Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18032–18037 (2011).
86. Pritchard, C. C. *et al.* ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J. Mol. Diagn. JMD* **14**, 357–366 (2012).
87. Castéra, L. *et al.* Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur. J. Hum. Genet. EJHG* **22**, 1305–1313 (2014).
88. Judkins, T. *et al.* Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. *BMC Cancer* **15**, 215 (2015).
89. Mannan, A. U. *et al.* Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: implications of embracing a multi-gene panel in molecular diagnosis in India. *J. Hum. Genet.* **61**, 515–522 (2016).
90. Kraus, C. *et al.* Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int. J. Cancer* **140**, 95–102 (2017).
91. Schroeder, C. *et al.* HBOC multi-gene panel testing: comparison of two sequencing centers. *Breast Cancer Res. Treat.* **152**, 129–136 (2015).
92. Susswein, L. R. *et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet. Med.* (2015). doi:10.1038/gim.2015.166
93. Buys, S. S. *et al.* A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer* n/a-n/a (2017). doi:10.1002/cncr.30498
94. Easton, D. F. *et al.* Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *N. Engl. J. Med.* **372**, 2243–2257 (2015).
95. Kamps, R. *et al.* Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* **18**, (2017).



96. Slavin, T. P. *et al.* Clinical Application of Multigene Panels: Challenges of Next-Generation Counseling and Cancer Risk Management. *Front. Oncol.* **5**, 208 (2015).
97. Yurgelun, M. B. *et al.* Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J. Clin. Oncol.* JCO.2016.71.0012 (2017). doi:10.1200/JCO.2016.71.0012
98. Maxwell, K. N. *et al.* Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. *Genet. Med.* **17**, 630–638 (2015).
99. Lincoln, S. E. *et al.* A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. *J. Mol. Diagn. JMD* **17**, 533–544 (2015).
100. Tung, N. *et al.* Frequency of Germline Mutations in 25 Cancer Susceptibility Genes in a Sequential Series of Patients With Breast Cancer. *J. Clin. Oncol.* **34**, 1460–1468 (2016).
101. 1000 Genome Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
102. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).
103. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33 Suppl**, 228–237 (2003).
104. Majewski, J., Schwartzenuber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? *J. Med. Genet.* **48**, 580–589 (2011).
105. Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* **59**, 5–15 (2014).
106. Ng, S. B. *et al.* Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes. *Nature* **461**, 272–276 (2009).
107. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793 (2010).
108. Sokolenko, A. P. *et al.* Identification of novel hereditary cancer genes by whole exome sequencing. *Cancer Lett.* **369**, 274–288 (2015).
109. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* **12**, 228 (2011).
110. Zhang, K. *et al.* A germline missense mutation in COQ6 is associated with susceptibility to familial schwannomatosis. *Genet. Med.* **16**, 787–792 (2014).
111. Shi, J. *et al.* Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat. Genet.* **46**, 482–486 (2014).
112. Robles-Espinoza, C. D. *et al.* POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.* **46**, 478–481 (2014).
113. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
114. Hamdan, F. F. *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).
115. O’Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
116. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
117. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
118. Nakashima, M. *et al.* Causative novel PNKP mutations and concomitant PCDH15 mutations in a patient

- with microcephaly with early-onset seizures and developmental delay syndrome and hearing loss. *J. Hum. Genet.* **59**, 471–474 (2014).
119. Hashimoto, R. *et al.* Whole-exome sequencing and neurite outgrowth analysis in autism spectrum disorder. *J. Hum. Genet.* **61**, 199–206 (2016).
120. Gauthier, J. *et al.* De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7863–7868 (2010).
121. Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
122. Ackerman, J. P. *et al.* Whole Exome Sequencing, Familial Genomic Triangulation, and Systems Biology Converge to Identify a Novel Nonsense Mutation in TAB2-encoded TGF-beta Activated Kinase 1 in a Child with Polyvalvular Syndrome. *Congenit. Heart Dis.* (2016). doi:10.1111/chd.12400
123. Tournier, I. *et al.* Germline mutations of inhibins in early-onset ovarian epithelial tumors. *Hum. Mutat.* **35**, 294–297 (2014).
124. Jones, S. *et al.* Exomic Sequencing Identifies PALB2 as a Pancreatic Cancer Susceptibility Gene. *Science* **324**, 217–217 (2009).
125. Roberts, N. J. *et al.* ATM Mutations in Patients with Hereditary Pancreatic Cancer. *Cancer Discov.* **2**, 41–46 (2012).
126. Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
127. Neveling, K. *et al.* A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum. Mutat.* **34**, 1721–1726 (2013).
128. Lee H, Deignan JL, Dorrani N & *et al.* CLinical exome sequencing for genetic identification of rare mendelian disorders. *JAMA* **312**, 1880–1887 (2014).
129. Kansal, R. *et al.* An infant with MLH3 variants, FOXG1-duplication and multiple, benign cranial and spinal tumors: A clinical exome sequencing study. *Genes. Chromosomes Cancer* **55**, 131–142 (2016).
130. Zhang, J. *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
131. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet. EJHG* **20**, 490–497 (2012).
132. Stenson, P. D. *et al.* The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum. Genomics* **4**, 69–72 (2009).
133. Tan, N. *et al.* Is ‘incidental finding’ the best term?: a study of patients’ preferences. *Genet. Med. Off. J. Am. Coll. Med. Genet.* (2016). doi:10.1038/gim.2016.96
134. Green, R. C. *et al.* Clinical Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of Genomic Medicine. *Am. J. Hum. Genet.* **98**, 1051–1066 (2016).
135. Teer, J. K. & Mullikin, J. C. Exome sequencing: the sweet spot before whole genomes. *Hum. Mol. Genet.* **19**, R145–R151 (2010).
136. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5473–5478 (2015).
137. Bougeard, G. *et al.* Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 2345–2352 (2015).
138. Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* **17**, 241 (2016).
139. Raffel, C. Medulloblastoma: Molecular Genetics and Animal Models. *Neoplasia N. Y. N* **6**, 310–322 (2004).

140. Coutant, S. *et al.* EVA: Exome Variation Analyzer, an efficient and versatile tool for filtering strategies in medical genomics. *BMC Bioinformatics* **13 Suppl 14**, S9 (2012).
141. Bolz, H. *et al.* Impaired calmodulin binding of myosin-7A causes autosomal dominant hearing loss (DFNA11). *Hum. Mutat.* **24**, 274–275 (2004).
142. Harada, H. *et al.* Identification of DMC1, a novel gene in the TOC region on 17q25.1 that shows loss of expression in multiple human cancers. *J. Hum. Genet.* **46**, 90–95 (2001).
143. Ailion, M. & Thomas, J. H. Isolation and characterization of high-temperature-induced Dauer formation mutants in *Caenorhabditis elegans*. *Genetics* **165**, 127–144 (2003).
144. Wang, L. *et al.* HID-1 is a peripheral membrane protein primarily associated with the medial- and trans-Golgi apparatus. *Protein Cell* **2**, 74–85 (2011).
145. Yu, Y. *et al.* HID-1 is a novel player in the regulation of neuropeptide sorting. *Biochem. J.* **434**, 383–390 (2011).
146. Mesa, R. *et al.* HID-1, a new component of the peptidergic signaling pathway. *Genetics* **187**, 467–483 (2011).
147. Gajjar, A. J. & Robinson, G. W. Medulloblastoma—translating discoveries from the bench to the bedside. *Nat. Rev. Clin. Oncol.* **11**, 714–722 (2014).
148. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
149. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
150. Muller, E. *et al.* Genetic profiles of cervical tumors by high-throughput sequencing for personalized medical care. *Cancer Med.* **4**, 1484–1493 (2015).
151. Backenroth, D. *et al.* CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res.* **42**, e97 (2014).
152. Benessahraoui, M., Aubin, F., Paratte, F., Plouvier, E. & Humbert, P. Leucémie myélomonocytaire juvénile, xanthomes et neurofibromatose de type 1. *Arch. Pédiatrie* **10**, 891–894 (2003).
153. Matsui, I. *et al.* Neurofibromatosis type 1 and childhood cancer. *Cancer* **72**, 2746–2754 (1993).
154. Shannon, K. & Bollag, G. Sending out an SOS. *Nat. Genet.* **39**, 8–9 (2007).
155. Jongmans, M. C. J. *et al.* Noonan syndrome, the SOS1 gene and embryonal rhabdomyosarcoma. *Genes. Chromosomes Cancer* **49**, 635–641 (2010).
156. Roberts, A. E. *et al.* Germline gain-of-function mutations in SOS1 cause Noonan syndrome. *Nat. Genet.* **39**, 70–74 (2007).
157. Foulkes, W. D., Priest, J. R. & Duchaine, T. F. DICER1: mutations, microRNAs and mechanisms. *Nat. Rev. Cancer* **14**, 662–672 (2014).
158. Doros, L. *et al.* DICER1-Related Disorders. in *GeneReviews*(®) (eds. Pagon, R. A. *et al.*) (University of Washington, Seattle, 2014).
159. Shaheen, I. S. *et al.* Bilateral progressive cystic nephroma in a 9-month-old male infant requiring renal replacement therapy. *Pediatr. Nephrol.* **25**, 1755–1758 (2010).
160. Cybulski, C. *et al.* A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J. Med. Genet.* **43**, 863–866 (2006).
161. Das, R. & Ghosh, S. K. Genetic variants of the DNA repair genes from Exome Aggregation Consortium (EXAC) database: significance in cancer. *DNA Repair* **52**, 92–102 (2017).

# ANNEXES

## Revisiting Li-Fraumeni Syndrome From *TP53* Mutation Carriers

Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Camille Charbonnier, Pierre Fermey, Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Emilie Consolino, Laurence Brugières, Olivier Caron, Patrick R. Benusiglio, Brigitte Bressac-de Paillerets, Valérie Bonadona, Catherine Bonaïti-Pellié, Julie Tinat, Stéphanie Baert-Desurmont, and Thierry Frebourg

See accompanying editorial on page 2331

Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Camille Charbonnier, Pierre Fermey, Julie Tinat, Stéphanie Baert-Desurmont, Thierry Frebourg, Institut National de la Santé et de la Recherche Médicale (Inserm) U1079, University of Rouen, Institute for Research and Innovation in Biomedicine; Julie Tinat, Stéphanie Baert-Desurmont, Thierry Frebourg, University Hospital, Rouen; Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Curie Institute, Paris; Emilie Consolino, Laurence Brugières, Olivier Caron, Patrick R. Benusiglio, Brigitte Bressac-de Paillerets, Gustave Roussy Institute, Villejuif; Valérie Bonadona, Centre National de la Recherche Scientifique (CNRS) UMR 5558, University of Lyon 1, Leon Berard Cancer Center, Lyon; and Catherine Bonaïti-Pellié, Inserm UMR-S 669, University of Paris-Sud, Villejuif, France.

Published online ahead of print at [www.jco.org](http://www.jco.org) on May 26, 2015.

Supported by the INCa, the French National Cancer Institute, and the ARC Foundation for Cancer Research.

Authors' disclosures of potential conflicts of interest are found in the article online at [www.jco.org](http://www.jco.org). Author contributions are found at the end of this article.

Corresponding author: Thierry Frebourg, Inserm U1079, Faculty of Medicine, 22 boulevard Gambetta, 76183 Rouen Cedex 1, France; e-mail: [frebourg@chu-rouen.fr](mailto:frebourg@chu-rouen.fr).

© 2015 by American Society of Clinical Oncology

0732-183X/15/3321w-2345w/\$20.00

DOI: 10.1200/JCO.2014.59.5728

### A B S T R A C T

#### Purpose

The aim of the study was to update the description of Li-Fraumeni syndrome (LFS), a remarkable cancer predisposition characterized by extensive clinical heterogeneity.

#### Patients and Methods

From 1,730 French patients suggestive of LFS, we identified 415 mutation carriers in 214 families harboring 133 distinct *TP53* alterations and updated their clinical presentation.

#### Results

The 322 affected carriers developed 552 tumors, and 43% had developed multiple malignancies. The mean age of first tumor onset was 24.9 years, 41% having developed a tumor by age 18. In childhood, the LFS tumor spectrum was characterized by osteosarcomas, adrenocortical carcinomas (ACC), CNS tumors, and soft tissue sarcomas (STS) observed in 30%, 27%, 26%, and 23% of the patients, respectively. In adults, the tumor distribution was characterized by the predominance of breast carcinomas observed in 79% of the females, and STS observed in 27% of the patients. The *TP53* mutation detection rate in children presenting with ACC or choroid plexus carcinomas, and in females with breast cancer before age 31 years, without additional features indicative of LFS, was 45%, 42% and 6%, respectively. The mean age of tumor onset was statistically different ( $P < .05$ ) between carriers harboring dominant-negative missense mutations (21.3 years) and those with all types of loss of function mutations (28.5 years) or genomic rearrangements (35.8 years). Affected children, except those with ACC, harbored mostly dominant-negative missense mutations.

#### Conclusion

The clinical gradient of the germline *TP53* mutations, which should be validated by other studies, suggests that it might be appropriate to stratify the clinical management of LFS according to the class of the mutation.

*J Clin Oncol* 33:2345-2352. © 2015 by American Society of Clinical Oncology

### INTRODUCTION

Li-Fraumeni syndrome (LFS; Mendelian Inheritance in Man [MIM] #151623), resulting from germline mutations of the *TP53* tumor suppressor gene (MIM\*191170), represents a remarkable inherited cancer susceptibility disorder not only because of the wide tumor spectrum, but also because the key role of p53 makes LFS a paradigm by which to understand the genetic determinism of cancer in humans. Indeed, the p53 protein is a transcription factor, upregulating the transcription of target genes involved in cell cycle arrest, DNA repair, apoptosis, and senescence, in response to DNA damage.<sup>1</sup> LFS was initially described in 1969 by Frederick Li and

Joseph Fraumeni in four families<sup>2</sup> and the original definition was subsequently established in 1988 from 24 families presenting with an autosomal dominant transmission of early-onset tumors including soft tissue sarcomas (STS), breast cancers, CNS tumors, leukemias and adrenocortical carcinomas (ACC) before the age of 45 years.<sup>3</sup> The second key step in the history of LFS was the identification in 1990 of germline *TP53* mutations in LFS families.<sup>4,5</sup> Twenty-four years later, despite transient controversies, the only gene unequivocally shown to be involved in LFS is *TP53*. Therefore, LFS can be defined as an inherited form of cancers caused by germline mutations of the *TP53* gene.<sup>6-9</sup> Since its discovery, the French LFS working group has updated

the definition criteria for LFS to cover the different clinical presentations associated with germline *TP53* mutations and to facilitate its clinical recognition.<sup>10-12</sup> These criteria, designated “the Chompret criteria,” take into account the three clinical situations suggestive of LFS: (1) familial presentation [a proband with an LFS tumor (breast cancer, STS, osteosarcoma, CNS tumor, ACC, leukemia, bronchoalveolar lung cancer) under 46 years and one first- or second-degree relative with an LFS tumor under 56 years or with multiple tumors], (2) multiple primary tumors (two of which belong to the narrow LFS spectrum, the first being developed before 46 years) or (3) rare cancers [ACC or choroid plexus carcinoma (CPC) irrespective of the family history]. The sensitivity and specificity of these criteria have been estimated to 82% to 95% and 47% to 58%, respectively.<sup>12-14</sup>

Considering the high penetrance of *TP53* mutations and the rapidity of LFS pedigree evolution, we performed, in the context of a national study, an update of the clinical data from the mutation carriers that we identified over the last 20 years.

## PATIENTS AND METHODS

### Patients

Between 1993 and 2013, 1,730 patients (individuals and unrelated families) with a history suggestive of LFS and addressed by comprehensive cancer centers or university hospitals were screened for germline *TP53* mutations. The majority ( $n = 981$ , 57%) fulfilled the Chompret criteria revised in 2009.<sup>12</sup> For each patient, genetic analyses were performed after informed consent was obtained.

### Molecular Analyses

The 11 exons of *TP53* and intron-exon boundaries were polymerase chain reaction (PCR) amplified from blood genomic DNA and analyzed using Sanger sequencing performed on an automated sequencer (PE Applied Biosystems, Foster City, CA). In families without mutation, we performed quantitative multiplex PCR of short fluorescent fragments (QMPSF) to detect genomic rearrangements.<sup>15</sup> Each alteration was systematically confirmed on an independent blood sample. Dominant-negative mutations were defined on the basis of the ability of the mutant protein to inhibit the activity of the wild-type p53 protein,<sup>16</sup> as indicated in the International Agency for Research on Cancer *TP53* Database.<sup>17</sup>

### Clinical Data Update

For each *TP53* mutation carrier, using a specific case report form sent to all French medical geneticists ensuring medical follow-up of patients with LFS, clinical data were systematically updated between 2008 and 2013: number, type, and localization of tumors, treatment, date of the last news or the age at death, and family history. Data were censored on October 31, 2013. This study was approved by the national ethical committee.

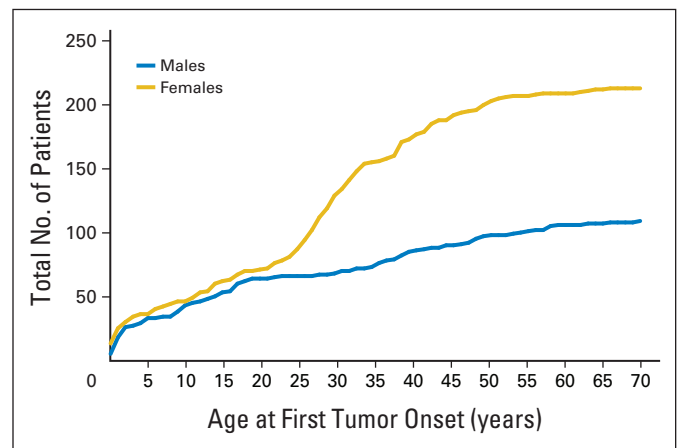
### Statistical Analyses

Statistical analyses were performed using the R statistical software ([www.R-project.org/](http://www.R-project.org/)). Ages of first tumor onset distributions were compared using Mann-Whitney *U* tests. Statistical significance was set at  $P < .05$ .

## RESULTS

### Clinical Presentation of the Families

Among the 214 families carrying a germline *TP53* mutation, 177 (83%) fulfilled the 2009 Chompret criteria for LFS.<sup>12</sup> The sensitivity and specificity of these criteria were estimated as 83% and 47%, respectively, and the mutation detection rate was 18%. Most of the other



**Fig 1.** Age (in years) at first tumor onset in 322 affected germline *TP53* mutation carriers.

families were characterized by early-onset breast cancers occurring before 36 years of age (22 of 37; 60%). Testing of 430 relatives from the 214 LFS families led us to identify a total of 415 *TP53* mutation carriers (257 females and 158 males).

### Tumor Incidence and Age of Onset

Among the *TP53* mutation carriers, 322 (78%) had developed at least one malignancy and a total of 552 tumors; 213 (83%) females and 109 (69%) males were affected, respectively (Fig 1). At age 5 years, 22% of the mutation carriers—and at age 18, 41%—had developed a tumor. The mean age of first tumor onset was 24.9 years (25.9 years in females v 22.7 years in males). The median age of tumor onset was 27.0 years, with a difference of 11 years between females (28.0 years) and males (17.0 years). After excluding breast cancers in female mutation carriers, the median age of tumor onset was lower in females than in males (13.0 v 17.0). We observed 18 cases of tumors occurring during the first year of life.

### LFS Tumor Distribution

The most frequent tumors observed in at least 10% of the affected mutation carriers were breast carcinomas, STS, osteosarcomas, CNS tumors and ACC (Table 1). In children, the most frequently observed tumors were osteosarcomas, ACC, CNS tumors and STS. Appendix Figure A1 presents their distribution according to age classes. In adults, the tumor distribution was characterized by the predominance of breast carcinomas, observed in 79% of affected female mutation carriers, and by a high frequency of STS, observed in 27% of affected carriers, the other tumors being observed in less than 10% of the affected mutation carriers (Table 1). In the 333 females who developed early breast cancer before 31 years of age, the *TP53* mutation rate was 14%. When we considered only patients who did not fulfill the Chompret criteria,<sup>12</sup> the mutation detection rate was 6%. We observed no breast carcinoma among the 59 affected adult male *TP53* mutation carriers. Among the 127 female mutation carriers who had developed a breast carcinoma, 40 (31%) also developed a contralateral breast cancer. Among the 49 breast carcinomas well characterized in terms of *ERBB2* amplification, progesterone and estrogen receptors expression, 27 (55%) exhibited *ERBB2* amplification and 18 (37%) were classified as triple-positive.

Revisiting Li-Fraumeni Syndrome

**Table 1.** Distribution of Tumors in Affected Germline *TP53* Mutation Carriers

Tumor	No. of Tumors	No. of Patients	% of Affected Mutation Carriers	Mean/Median Age at Tumor Onset, Years (range)*
In All Patients (N = 322; 213†, 109‡)				
Breast carcinoma	172 (172†; 0‡)	127 (127†)	<b>60†</b>	35/33 (20-69†)
Soft tissue sarcoma	104 (54†; 50‡)	86 (46†; 40‡)	<b>27</b>	29/31 (0.5-67†; 0.5-70‡)
Osteosarcoma	53 (24†; 29‡)	50 (23†; 27‡)	<b>16</b>	18/16 (8-55†; 5-54‡)
CNS tumor	43 (25†; 18‡)	42 (25†; 17‡)	<b>13</b>	15/11 (0.3-67†; 1-33‡)
Adrenocortical carcinoma	43 (30†; 13‡)	42 (29†; 13‡)	<b>13</b>	6/1 (0.5-41†; 0.7-19‡)
Lung cancer	18 (8†; 10‡)	18 (8†; 10‡)	6	42/44 (14-58†; 37-54‡)
Leukemia	16 (9†; 7‡)	14 (7†; 7‡)	4	14/12 (6-35†; 2-16‡)
Prostate cancer	4 (4‡)	4 (4‡)	4‡	63/62 (57-71‡)
Colorectal cancer	12 (5†; 7‡)	11 (5†; 6‡)	3	40/40 (21-74†; 27-52‡)
Renal cancer	11 (4†; 7‡)	9 (3†; 6‡)	3	51/49 (34-68†; 41-70‡)
Melanoma	11 (9†; 2‡)	8 (6†; 2‡)	2	40/43 (25-65†; 15-55‡)
Lymphoma	7 (3†; 4‡)	7 (3†; 4‡)	2	18/13 (11-42†; 2-32‡)
Stomach carcinoma	7 (3†; 4‡)	7 (3†; 4‡)	2	41/44 (44-58†; 17-54‡)
Pancreas carcinoma	6 (2†; 4‡)	6 (2†; 4‡)	2	46/39 (32-64†; 37-66‡)
Head and neck cancer	6 (1†; 5‡)	6 (1†; 5‡)	2	41/42 (21†; 30-59‡)
Skin cancer	6 (4†; 2‡)	5 (3†; 2‡)	2	31/31 (29-39†; 19-ND‡)
Chondrosarcoma	5 (2†; 3‡)	5 (2†; 3‡)	2	32/31 (21-57†; 19-33‡)
Ovary carcinoma	3 (3†)	3 (3†)	1†	43/35 (26-69†)
Nephroblastoma	3 (0†; 3‡)	3 (0†; 3‡)	0.9	3/2 (1-6‡)
Thyroid carcinoma	3 (3†; 0‡)	3 (3†; 0‡)	0.9	42/40 (35-50†)
Endometrial carcinoma	2 (2†)	2 (2†)	0.9†	54/54 (31-77†)
Testis choriocarcinoma	1 (1‡)	1 (1‡)	0.9‡	17/17 (17‡)
Esophageal cancer	2 (0†; 2‡)	2 (0†; 2‡)	0.6	63/63 (59-67‡)
Bladder carcinoma	2 (1†; 1‡)	2 (1†; 1‡)	0.6	49/49 (66†; 31‡)
Cervical carcinoma	1 (1†)	1 (1†)	0.5†	ND†
Gestational choriocarcinoma	1 (1†)	1 (1†)	0.5†	30/30 (30†)
Neuroblastoma	1 (1†; 0‡)	1 (1†; 0‡)	0.3	0.5/0.5 (0.5†)
Mesothelioma	1 (1†; 0‡)	1 (1†; 0‡)	0.3	52/52 (52†)
Myeloma	1 (0†; 1‡)	1 (0†; 1‡)	0.3	43/43 (43‡)
Thymoma	1 (1†; 0‡)	1 (1†; 0‡)	0.3	33/33 (33†)
Appendix carcinoma	1 (1†; 0‡)	1 (1†; 0‡)	0.3	45/45 (45†)
Anal cancer	1 (0†; 1‡)	1 (0†; 1‡)	0.3	50/50 (50‡)
Not determined	4 (1†; 3‡)	4 (1†; 3‡)		(46†; 33-49‡)
In Children (n = 132; 70†, 62‡)				
Osteosarcoma	40 (20†; 20‡)	39 (19†; 20‡)	<b>30</b>	14/14 (8-18†; 5-18‡)
Adrenocortical carcinoma	37 (25†; 12‡)	36 (24†; 12‡)	<b>27</b>	2/1 (0.5-11†; 0.7-17‡)
CNS tumor	34 (19†; 15‡)	34 (19†; 15‡)	<b>26</b>	8/9 (0.3-18†; 1-18‡)
Soft tissue sarcoma	33 (13†; 21‡)	31 (12†; 19‡)	<b>23</b>	5/2 (0.5-17†; 0.5-16‡)
Leukemia	13 (7†; 6‡)	12 (6†; 6‡)	9	11/12 (6-17†; 2-16‡)
Lymphoma	4 (2†; 2‡)	4 (2†; 2‡)	3	9/10 (11-13†; 2-9‡)
Nephroblastoma	3 (0†; 3‡)	3 (0†; 3‡)	2	3/2 (1-6‡)
Testis choriocarcinoma	1 (1‡)	1 (1‡)	2‡	17/17 (17‡)
Neuroblastoma	1 (1†; 0‡)	1 (1†; 0‡)	0.8	0.5/0.5 (0.5†)
Lung cancer	1 (1†; 0‡)	1 (1†; 0‡)	0.8	14/14 (14†)
Stomach carcinoma	1 (0†; 1‡)	1 (0†; 1‡)	0.8	17/17 (17‡)
In Adults (n = 219; 160†, 59‡)				
Breast carcinoma	172 (172†; 0‡)	127 (127†)	<b>79†</b>	35/33 (20-69†)
Soft tissue sarcoma	70 (41†; 29‡)	59 (34†; 25‡)	<b>27</b>	41/42 (21-67†; 22-70‡)
Lung cancer	17 (7†; 10‡)	17 (7†; 10‡)	8	44/45 (22-58†; 37-54‡)
Prostate cancer	4 (4‡)	4 (4‡)	7‡	63/62 (57-71‡)
Osteosarcoma	13 (4†; 9‡)	13 (4†; 9‡)	6	31/28 (33-55†; 19-54‡)
Colorectal cancer	12 (5†; 7‡)	11 (5†; 6‡)	5	40/40 (21-74†; 27-52‡)
Renal cancer	11 (4†; 7‡)	9 (3†; 6‡)	4	51/49 (34-68†; 41-70‡)
CNS tumor	9 (6†; 3‡)	8 (6†; 2‡)	4	39/33 (20-67†; 30-33‡)

(continued on following page)

**Table 1.** Distribution of Tumors in Affected Germline *TP53* Mutation Carriers (continued)

Tumor	No. of Tumors	No. of Patients	% of Affected Mutation Carriers	Mean/Median Age at Tumor Onset, Years (range)*
Melanoma	10 (9†; 1‡)	7 (6†; 1‡)	3	42/43 (25-65†; 55‡)
Adrenocortical carcinoma	6 (5†; 1‡)	6 (5†; 1‡)	3	27/28 (20-41†; 19‡)
Stomach carcinoma	6 (3†; 3‡)	6 (3†; 3‡)	3	45/46 (44-58†; 29-54‡)
Pancreas carcinoma	6 (2†; 4‡)	6 (2†; 4‡)	3	46/39 (32-64†; 37-66‡)
Head and neck cancer	6 (1†; 5‡)	6 (1†; 5‡)	3	41/42 (21†; 30-59‡)
Skin cancer	6 (4†; 2‡)	5 (3†; 2‡)	2	31/31 (29-39†; 19-ND‡)
Chondrosarcoma	5 (2†; 3‡)	5 (2†; 3‡)	2	32/31 (21-57†; 19-33‡)
Ovary carcinoma	3 (3†)	3 (3†)	2†	43/35 (26-69†)
Leukemia	3 (2†; 1‡)	3 (2†; 1‡)	1	32/32 (28-35†; ND‡)
Lymphoma	3 (1†; 2‡)	3 (1†; 2‡)	1	31/32 (42†; 19-32‡)
Thyroid carcinoma	3 (3†; 0‡)	3 (3†; 0‡)	1	42/40 (35-50†)
Endometrial carcinoma	2 (2†)	2 (2†)	1†	54/54 (31-77†)
Esophageal cancer	2 (0†; 2‡)	2 (0†; 2‡)	0.9	63/63 (59-67‡)
Bladder carcinoma	2 (1†; 1‡)	2 (1†; 1‡)	0.9	49/49 (66†; 31‡)
Cervical carcinoma	1 (1†)	1 (1†)	0.6†	ND†
Gestational choriocarcinoma	1 (1†)	1 (1†)	0.6†	30/30 (30†)
Mesothelioma	1 (1†; 0‡)	1 (1†; 0‡)	0.5	52/52 (52†)
Myeloma	1 (0†; 1‡)	1 (0†; 1‡)	0.5	43/43 (43‡)
Thymoma	1 (1†; 0‡)	1 (1†; 0‡)	0.5	33/33 (33†)
Appendix carcinoma	1 (1†; 0‡)	1 (1†; 0‡)	0.5	45/45 (45†)
Anal cancer	1 (0†; 1‡)	1 (0†; 1‡)	0.5	50/50 (50‡)

\*Numbers in bold indicate values above 10%.  
†In females.  
‡In males.

STS was the second most frequent type of malignancy with a wide range of tumor onset age (6 months to 70 years). The most frequent STS types were rhabdomyosarcomas (RMS), leiomyosarcomas, liposarcomas, and fibrohistiocytic tumors corresponding to 29%, 25%, 12% and 11% of the cases, respectively. Among the 31 children with RMS but without personal or familial history suggestive of LFS,<sup>12</sup> the mutation detection rate was 52%; they were tested because they developed subsequent tumors or because relatives also developed tumors. In the 13 cases of RMS for which pathologic data were available, the diagnosis was compatible with embryonal RMS of anaplastic subtype. Osteosarcoma was the third most commonly observed tumor in affected mutation carriers and the most frequent tumor in children (Table 1).

We observed 43 cases of CNS tumors in 42 patients, occurring mostly in children (n = 34, 81%). These CNS tumors corresponded in 19 cases to glioblastoma and other types of glioma (44%, ages 3 to 67 years), in 13 cases to CPC (30%, ages 4 months to 33 years), in eight cases to medulloblastoma (19%, ages 7 to 14 years) and in three cases to ependymoma (7%, ages 4 months to 13 years). Among the 24 children with CPC, 12 (50%) were found to carry a germline *TP53* mutation, and among the 19 children who were tested only on the basis of the development of CPC, eight (42%) carried a germline *TP53* mutation.

Among 106 patients with ACC tested for *TP53*, 42 carried a germline *TP53* mutation. In 50% of children with ACC (36 of 72), we found a germline *TP53* mutation, whereas this mutation detection rate was only 18% (6 of 34) in adults. In the 53 children with ACC and without familial history suggestive of LFS, we detected a germline mutation in 24 (45%).

### Frequency of Multiple Primary Tumors

Multiple primary tumors (n = 2 to 6; Fig 2A) were observed in 43% (139 of 322) of the mutation carriers. Most of the multiple primary tumors (116 of 139, 83%) were metachronous (Fig 2B and Appendix Fig A2, online only). Among the *TP53* mutation carriers who received radiotherapy for the treatment of their first tumor, we could obtain complete information for 64 of them: 19 (30%) developed 26 secondary tumors within a radiation field, the second tumor occurring between 2 and 26 years after radiotherapy (mean timing: 10.7 years).

### Germline *TP53* Mutation Spectrum

Among the 214 LFS families, we identified 133 distinct germline *TP53* alterations corresponding to 124 point mutations, eight deletions, and one duplication. Most of the families (205 to 214; 96%) carried point mutations, 67% (143 of 214) harbor missense mutations, and 35% (75 of 214) dominant-negative mutations (Appendix Table A1, online only). All the point mutations were distributed among exons 3 to 11 (Appendix Fig A3A, online only). The top mutated codons were codons 125, 158, 175, 248, 273, 282, and 37, which were altogether mutated in 83 (39%) of the families (Appendix Fig A3B). We found in nine families nine distinct *TP53* genomic rearrangements corresponding to two whole gene deletions, six partial deletions (promoter region to exon 1; exons 2 to 6; exons 2 to 10; exons 2 to 11), and one partial duplication (intron 1 to exon 4).

### Genotype-Phenotype Correlations

We did not observe a significant difference in the age of tumor onset between missense mutations and the other alterations (Table 2). Because the biologic effect on the protein of splicing mutations is



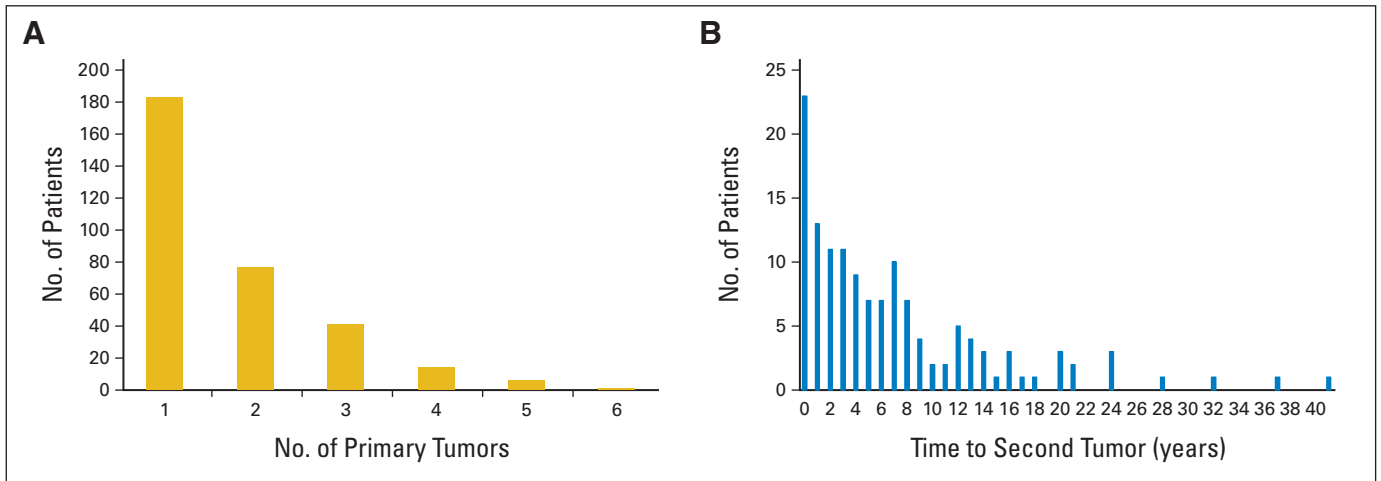


Fig 2. Occurrence of multiple primary tumors in 322 affected germline *TP53* mutation carriers. (A) Number; (B) second tumor.

difficult to predict, we compared the age of tumor onset between missense mutations versus frameshift or in frame deletions/insertions, nonsense mutations and genomic rearrangements, all these alterations being predicted to result only into a loss of function. The difference was statistically significant (mean age of tumor onset: 23.8 v 28.5), and the difference was even higher when we compared missense mutations to genomic rearrangements corresponding to extreme forms of null alterations (mean age of tumor onset: 35.8 years). We then compared the age of tumor onset between missense mutations classified within the *TP53* database as dominant-negative mutations, and the other types of alterations. We found a statistically significant difference (mean age of tumor onset: 21.3 v 27 years) that was higher when we restricted the comparison to the loss of function mutations and to genomic rearrangements (Table 2). These differences are also

illustrated by the differences in the curves representing, for the different classes of *TP53* alterations, the cumulative number of affected patients according to age (Fig 3A). No difference was observed between missense mutations located within and outside the DNA binding domain of p53 corresponding to the residues 103 to 293 (Table 2).

We then specifically compared the age of tumor onset between the most frequently detected point mutations affecting codons 125, 158, 175, 248, 273, 282, and 337, and the *TP53* complete deletion, all these alterations having been detected in at least six patients (Appendix Table A2, online only). The results obtained on these eight distinct *TP53* alterations, representative of the different types of mutations, confirmed a gradient of clinical severity (Fig 3B).

In children presenting with osteosarcomas, CNS tumors, or RMS as first tumor, the most frequent germline *TP53* mutations (Appendix Table A3, online only) were dominant-negative missense mutations, respectively detected in 40% (8 of 20), 62% (18 of 29) and 36% (8 of 22). In contrast, in children with ACC, we found that the majority of cases (26 of 34, 76%) harbored other types of alterations than dominant-negative missense mutations (Appendix Tables A3 and A4, online only). The predominant mutation in childhood ACC was the p.(Arg158His) mutation detected in six families with ACC case(s), and the majority of the families (six of eight) harboring this mutation were characterized by childhood ACC and adult tumors.

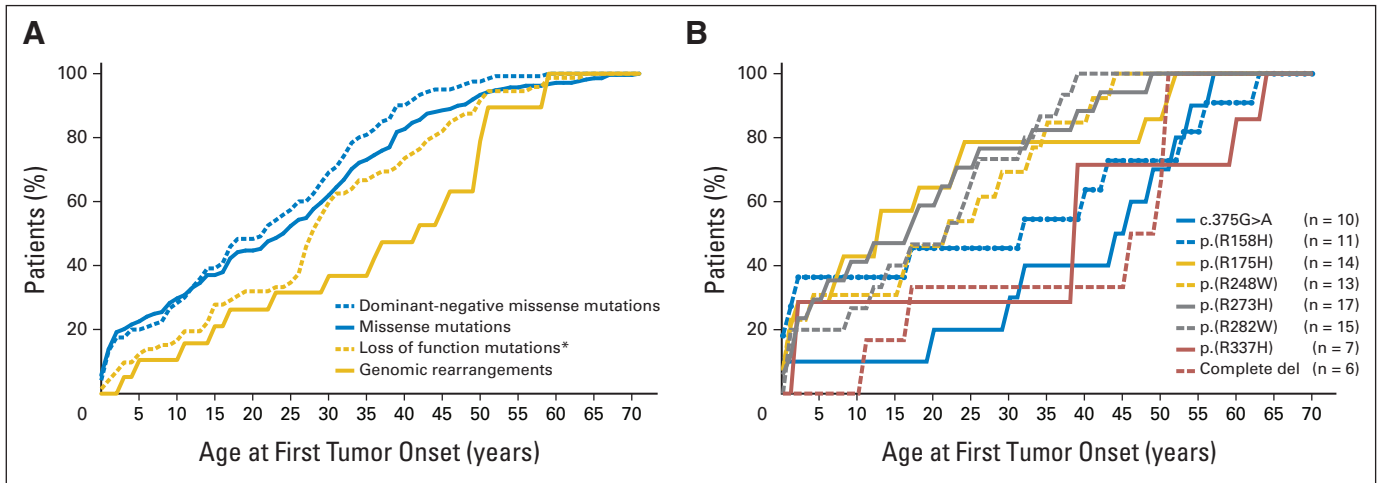
**Table 2.** Comparison of Age of Tumor Onset in LFS According to Type of Germline *TP53* Alteration (N = 322 affected mutation carriers)

Type of Alteration*	<i>P</i> , Mann-Whitney test
Missense mutations (n = 208): 23.8 ± 17.7 yr	
Other alterations (n = 114): 26.9 ± 17.8 yr	.1248
Loss of function mutations† (n = 72): 28.5 ± 16.4 yr	<b>.0354</b>
Genomic rearrangements (n = 19): 35.8 ± 18.3	<b>.0042</b>
Dominant-negative missense mutations (n = 120): 21.3 ± 14.8 yr	
Other alterations‡ (n = 202): 27.0 ± 19.1 yr	<b>.0121</b>
Loss of function mutations† (n = 72): 28.5 ± 16.4 yr	<b>.0042</b>
Genomic rearrangements (n = 19): 35.8 ± 18.3	<b>.0008</b>
Missense mutations within p53 DBD (n = 189): 23.0 ± 17.5 yr	
Missense mutations outside p53 DBD (n = 19): 31.3 ± 19.1 yr	.0561

NOTE. Boldface indicates *P* < .05.  
 Abbreviation: DBD, DNA binding domain; LFS, Li-Fraumeni syndrome; SD, standard deviation.  
 \*For each type of alteration, the number of patients and mean age ± SD at tumor onset are indicated.  
 †Frameshift or in frame deletions/insertions, nonsense mutations, and genomic rearrangements.  
 ‡Note that, among these mutations, some may correspond to bona fide dominant-negative mutations because not all mutations have been tested.

## DISCUSSION

The aim of the study was to provide an updated description of LFS based on a large series of germline *TP53* mutation carriers. Although there was an ascertainment bias in this study, as subjects had been tested on the basis of clinical criteria, this series, which is to our knowledge the largest worldwide, provides a global picture of LFS and offers novel insights into the clinical presentation of the syndrome. Because the patients had been selected on the basis of extended clinical criteria corresponding in most cases to the 2009 version of the Chompret criteria,<sup>12</sup> we think that the relative distribution of the tumors (Table 1) is representative of what should be observed in unselected germline *TP53* mutation carriers. We provide evidence that the tumor



**Fig 3.** Age at first tumor onset according to the type of *TP53* alteration. (A) According to the different classes of *TP53* alterations. (\*) Loss of function mutations correspond to frameshift or in-frame deletions/insertions, nonsense mutations, and genomic rearrangements. (B) According to specific *TP53* alteration. del, deletion.

spectrum can be dichotomized between children and adults. In childhood, LFS is characterized by a wide spectrum including mostly osteosarcomas, ACC, CNS tumors, and STS. In adults, the tumor spectrum is more restricted and characterized by female breast cancers and STS (Table 1).

In terms of tumor subtype, we did not observe that breast carcinomas exhibited predominantly a triple-positive profile, as previously reported,<sup>18</sup> but the majority of tumors were ERBB2 positive. We found that a germline *TP53* mutation could be detected in 6% of patients with early-onset breast cancer (occurring before 31 years of age) who did not fulfill the Chompret criteria. This is in agreement with previously published studies reporting a mutation detection rate between 3% and 8% in females with sporadic early-onset breast cancers.<sup>19-21</sup> We confirmed that RMS was the most frequent STS observed in *TP53* mutation carriers and that the diagnosis was compatible, as previously reported, with embryonal anaplastic RMS.<sup>22</sup> The mutation detection rate observed in children with RMS without personal or familial history suggestive of LFS (52%) was biased, as these children had been selected on the basis of other clinical criteria. These results led us to propose that childhood anaplastic RMS and breast cancer before 31 years of age may be integrated, like CPC and ACC, in a new version of the Chompret criteria (Table 3).

One of the striking observations was the frequency of multiple primary tumors in LFS (40%), already highlighted by several reports.<sup>23-25</sup> This exceptionally high incidence, the development of tumors within radiotherapy fields in 30% of mutation carriers—an observation previously reported<sup>26,27</sup>—and the sequential development of multiple primary tumors in carriers who develop late onset tumors (data not shown) provide strong clinical arguments suggesting that chemotherapy and radiotherapy contribute to the development of secondary tumors in LFS. This hypothesis could easily be explained by the key role of p53 in response to DNA damage.<sup>1</sup>

This updated series of *TP53* mutation carriers also showed that it is possible to distinguish different classes of alterations according to their clinical severity. The most severe mutations are the dominant-negative missense mutations: they are significantly associated with earlier tumor onset (Table 2; Appendix Table A2; Fig 3), and they represent the predominant germline alterations in carriers who de-

velop childhood cancers except ACC (Appendix Table A3). The less severe alterations correspond to loss of function mutations, such as nonsense mutations, frameshift mutations, or genomic rearrangements; these alterations are associated with later tumor onset (Table 2; and Appendix Table A2; Fig 3). These inactivating mutations were predominantly detected in pedigrees characterized by cancers occurring in adults. The nondominant-negative missense mutations define an intermediate class, in terms of clinical severity (Table 2; Appendix Table A2; Fig 3). We recently demonstrated the biologic severity of the germline dominant-negative missense mutations which drastically alter the mediated transcriptional response to DNA damage.<sup>28</sup> This clinical gradient of germline *TP53* mutations had already been suggested by mouse models of LFS, which had revealed the phenotypic severity of dominant-negative mutations, as compared with null mutations.<sup>29-34</sup> Interestingly, in pedigrees including childhood ACC, we detected, as previously reported,<sup>35</sup> mostly other types of alterations than dominant-negative missense mutations (Appendix Tables A3 and A4), and the predominant one was the p.(Arg158His) mutation associated with adult tumors. The phenotypic expression of this mutation is reminiscent of that of the p.(Arg337His) mutation, associated

**Table 3.** 2015 Version of Chompret Criteria

Familial presentation	Proband with tumor belonging to LFS tumor spectrum (eg, premenopausal breast cancer, soft tissue sarcoma, osteosarcoma, CNS tumor, adrenocortical carcinoma) before age 46 yr, AND at least one first- or second-degree relative with LFS tumor (except breast cancer if proband has breast cancer) before age 56 yr or with multiple tumors
Multiple primitive tumors	Proband with multiple tumors (except multiple breast tumors), two of which belong to LFS tumor spectrum and first of which occurred before age 46 yr
Rare tumors	Patient with adrenocortical carcinoma, choroid plexus tumor, or rhabdomyosarcoma of embryonal anaplastic subtype, irrespective of family history
Early-onset breast cancer	Breast cancer before age 31 yr

Abbreviation: LFS, Li-Fraumeni syndrome.

with a founder effect in Brazil.<sup>36-38</sup> We confirmed here the association of low penetrant *TP53* mutations with childhood ACC, an observation first made in 1999.<sup>39</sup> Adrenocortical glands undergo, during the last phases of embryonic development, an important remodeling involving apoptosis.<sup>40,41</sup> It is therefore possible that p53 is particularly critical for the development of adrenal glands and that even a subtle alteration of p53 may have a specific effect within this tissue.

This study has potentially two important clinical implications. First, the high incidence of multiple primary tumors in *TP53* mutation carriers suggests the contribution of chemotherapy and radiotherapy used in the treatment of the first cancer to the development of secondary tumors. Therefore, one of the priorities of LFS clinical research in the future should be the development of treatment protocols that are not genotoxic to reduce the risk of secondary tumors in patients with LFS. Second, the clinical gradient of germline *TP53* mutations leads us to suggest that it might be appropriate to stratify the clinical management of carriers according to the class of the mutation; our results should prompt other investigators to evaluate the validity of this clinical gradient. When considering the diagnosis of LFS in a patient, the first concern of the physician is the benefit-risk ratio, in terms of both medical and psychological consequences. In families harboring dominant-negative missense mutations, their severity may constitute an argument for presymptomatic testing in children and annual screening protocols, such as those including total body magnetic resonance imaging (MRI),<sup>8,42</sup> which is under evaluation in different countries. In contrast, in families harboring other classes of mutations, the lower severity of the mutations, which are mostly associated with adult cancers, could justify to restrict presymptomatic testing to adults, with the aim in particular to offer to female mutation carriers annual breast MRI from 20 years of age, as recommended by the National Comprehensive Cancer Network in US and also performed

in several countries.<sup>9</sup> In families with mutations associated with ACC, such as p.(Arg158His) and p.(Arg337His), presymptomatic testing might be offered to children, but with the aim to offer a detection program restricted to ACC and based on biochemical analyses and/or abdominal ultrasound. This stratification of clinical management according to the clinical severity of the mutations—if this clinical gradient is confirmed by other studies—would allow one to take into account, for the patients' benefit, the remarkable heterogeneity of this syndrome.

#### AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at [www.jco.org](http://www.jco.org).

#### AUTHOR CONTRIBUTIONS

**Conception and design:** Gaëlle Bougeard, Thierry Frebourg

**Administrative support:** Gaëlle Bougeard, Catherine Bonaiti-Pellié, Thierry Frebourg

**Provision of study materials or patients:** Gaëlle Bougeard, Muriel Belotti, Emilie Consolino, Laurence Brugières, Olivier Caron, Brigitte Bressac-de Paillerets, Valérie Bonadona, Julie Tinat, Thierry Frebourg

**Collection and assembly of data:** Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Pierre Fermey, Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Olivier Caron, Patrick R. Benusiglio, Brigitte Bressac-de Paillerets, Valérie Bonadona, Catherine Bonaiti-Pellié, Stéphanie Baert-Desurmont, Thierry Frebourg

**Data analysis and interpretation:** Gaëlle Bougeard, Camille Charbonnier, Thierry Frebourg

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

#### REFERENCES

- Reinhardt HC, Schumacher B: The p53 network: Cellular and systemic DNA damage responses in aging and cancer. *Trends Genet* 28:128-136, 2012
- Li FP, Fraumeni JF Jr: Soft tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med* 71:747-752, 1969
- Li FP, Fraumeni JF Jr, Mulvihill JJ, et al: A cancer family syndrome in twenty-four kindreds. *Cancer Res* 48:5358-5362, 1988
- Malkin D, Li FP, Strong LC, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250:1233-1238, 1990
- Srivastava S, Zou ZQ, Pirolo K, et al: Germline transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* 348:747-749, 1990
- Malkin D: Li-fraumeni syndrome. *Genes Cancer* 2:475-484, 2011
- Mai PL, Malkin D, Garber JE, et al: Li-Fraumeni syndrome: Report of a clinical research workshop and creation of a research consortium. *Cancer Genet* 205:479-487, 2012
- McBride KA, Ballinger ML, Killick E, et al: Li-Fraumeni syndrome: Cancer risk assessment and clinical management. *Nat Rev Clin Oncol* 11:260-271, 2014
- Kamihara J, Rana HQ, Garber JE: Germline TP53 Mutations and the Changing Landscape of Li-Fraumeni Syndrome. *Hum Mutat* 35:654-662, 2014
- Chompret A, Abel A, Stoppa-Lyonnet D, et al: Sensitivity and predictive value of criteria for p53 germline mutation screening. *J Med Genet* 38:43-47, 2001
- Bougeard G, Sesboué R, Baert-Desurmont S, et al: Molecular basis of the Li-Fraumeni syndrome: An update from the French LFS families. *J Med Genet* 45:535-538, 2008
- Tinat J, Bougeard G, Baert-Desurmont S, et al: Version of the Chompret criteria for Li Fraumeni syndrome. *J Clin Oncol* 27:e108-e109; author reply e110, 2009
- Gonzalez KD, Noltner KA, Buzin CH, et al: Beyond Li Fraumeni syndrome: Clinical characteristics of families with p53 germline mutations. *J Clin Oncol* 27:1250-1256, 2009
- Ruijs MW, Verhoef S, Rookus MA, et al: TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: Mutation detection rate and relative frequency of cancers in different familial phenotypes. *J Med Genet* 47:421-428, 2010
- Bougeard G, Brugières L, Chompret A, et al: Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. *Oncogene* 22:840-846, 2003
- Guimaraes DP, Hainaut P: TP53: A key gene in human cancer. *Biochimie* 84:83-93, 2002
- Petitjean A, Mathe E, Kato S, et al: Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: Lessons from recent developments in the IARC TP53 database. *Hum Mutat* 28:622-629, 2007
- Wilson JR, Bateman AC, Hanson H, et al: A novel HER2-positive breast cancer phenotype arising from germline TP53 mutations. *J Med Genet* 47:771-774, 2010
- Lalloo F, Varley J, Moran A, et al: BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives. *Eur J Cancer* 42:1143-1150, 2006
- Mouchawar J, Korch C, Byers T, et al: Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res* 70:4795-4800, 2010
- McCuaig JM, Armel SR, Novokmet A, et al: Routine TP53 testing for breast cancer under age 30: Ready for prime time? *Fam Cancer* 11:607-613, 2012
- Hettmer S, Archer NM, Somers GR, et al: Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. *Cancer* 120:1068-1075, 2014
- Hisada M, Garber JE, Fung CY, et al: Multiple primary cancers in families with Li-Fraumeni syndrome. *J Natl Cancer Inst* 90:606-611, 1998
- Nutting C, Camplejohn RS, Gilchrist R, et al: A patient with 17 primary tumours and a germ line mutation in TP53: Tumour induction by adjuvant therapy? *Clin Oncol (R Coll Radiol)* 12:300-304, 2000
- Izawa N, Matsumoto S, Manabe J, et al: A Japanese patient with Li-Fraumeni syndrome who

had nine primary malignancies associated with a germline mutation of the p53 tumor-suppressor gene. *Int J Clin Oncol* 13:78-82, 2008

26. Heymann S, Delaloge S, Rahal A, et al: Radio-induced malignancies after breast cancer postoperative radiotherapy in patients with Li-Fraumeni syndrome. *Radiat Oncol* 5:104, 2010

27. Limacher JM, Frebourg T, Natarajan-Ame S, et al: Two metachronous tumors in the radiotherapy fields of a patient with Li-Fraumeni syndrome. *Int J Cancer* 96:238-242, 2001

28. Zerdomi Y, Aury-Landas J, Bonaiti-Pellié C, et al: Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. *Hum Mutat* 34:453-461, 2013

29. Olive KP, Tuveson DA, Ruhe ZC, et al: Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* 119:847-860, 2004

30. Lang GA, Iwakuma T, Suh YA, et al: Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* 119:861-872, 2004

31. Song H, Hollstein M, Xu Y: p53 gain-of-function cancer mutants induce genetic instability

by inactivating ATM. *Nat Cell Biol* 9:573-580, 2007

32. Liu DP, Song H, Xu Y: A common gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene* 29:949-956, 2010

33. Hanel W, Marchenko N, Xu S, et al: Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis. *Cell Death Differ* 20:898-909, 2013

34. Muller PA, Vousden KH: Mutant p53 in cancer: New functions and therapeutic opportunities. *Cancer Cell* 25:304-317, 2014

35. Palmero EI, Achatz MI, Ashton-Prolla P, et al: Tumor protein 53 mutations and inherited cancer: Beyond Li-Fraumeni syndrome. *Curr Opin Oncol* 22:64-69, 2010

36. Custódio G, Parise GA, Kiesel Filho N, et al: Impact of neonatal screening and surveillance for the TP53 R337H mutation on early detection of childhood adrenocortical tumors. *J Clin Oncol* 31:2619-2626, 2013

37. Achatz MI, Olivier M, Le Calvez F, et al: The TP53 mutation, R337H, is associated with Li-

Fraumeni and Li-Fraumeni-like syndromes in Brazilian families. *Cancer Lett* 245:96-102, 2007

38. Achatz MI, Hainaut P, Ashton-Prolla P: Highly prevalent TP53 mutation predisposing to many cancers in the Brazilian population: A case for newborn screening? *Lancet Oncol* 10:920-925, 2009

39. Varley JM, McGown G, Thorncroft M, et al: Are there low-penetrance TP53 alleles? Evidence from childhood adrenocortical tumors. *Am J Hum Genet* 65:995-1006, 1999

40. Spencer SJ, Mesiano S, Lee JY, et al: Proliferation and apoptosis in the human adrenal cortex during the fetal and perinatal periods: Implications for growth and remodeling. *J Clin Endocrinol Metab* 84:1110-1115, 1999

41. Coulter CL: Fetal adrenal development: Insight gained from adrenal tumors. *Trends Endocrinol Metab* 16:235-242, 2005

42. Villani A, Tabori U, Schiffman J, et al: Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: A prospective observational study. *Lancet Oncol* 12:559-567, 2011

## Innovative Content From JCO: Oncology Grand Rounds



- Case-based descriptions of treatment or diagnostic approaches
- Places report's findings into practical, clinical, real-world context

Subscribe or activate your online access today.

Phone: 888-273-3508 or 703-519-1430; e-mail: [jcoservice@asco.org](mailto:jcoservice@asco.org); online: [jco.org/subscriptions](http://jco.org/subscriptions).



American Society of Clinical Oncology

**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

**Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers**

*The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [jco.ascopubs.org/site/ifc](http://jco.ascopubs.org/site/ifc).*

**Gaëlle Bougeard**

No relationship to disclose

**Mariette Renaux-Petel**

No relationship to disclose

**Jean-Michel Flaman**

No relationship to disclose

**Camille Charbonnier**

No relationship to disclose

**Pierre Fermey**

No relationship to disclose

**Muriel Belotti**

No relationship to disclose

**Marion Gauthier-Villars**

No relationship to disclose

**Dominique Stoppa-Lyonnet**

No relationship to disclose

**Emilie Consolino**

No relationship to disclose

**Laurence Brugières**

No relationship to disclose

**Olivier Caron**

**Consulting or Advisory Role:** AstraZeneca

**Patrick R. Benusiglio**

No relationship to disclose

**Brigitte Bressac-de Pailleters**

**Patents, Royalties, Other Intellectual Property:** Patent "Marqueur de predisposition à un cancer" No. PTC/FR2010/052853

**Valérie Bonadona**

No relationship to disclose

**Catherine Bonaiti-Pellié**

No relationship to disclose

**Julie Tinat**

No relationship to disclose

**Stéphanie Baert-Desurmont**

No relationship to disclose

**Thierry Frebourg**

No relationship to disclose

### Acknowledgment

We gratefully acknowledge the LFS families for their helpful contribution to this study. We thank Mario Tosi for critical review of the manuscript; Emilie Angot for her pathologic expertise; and Emilie Bouvignies, Françoise Charbonnier, Steeve Fourneaux, Gwendoline Lienard, Cosette Martin, Stéphanie Vasseur, and Emeline Yhuel for their technical assistance. We are grateful to the French LFS network: Caroline Abadie, Anne Abel, Claude Adenis, Souria Aissaoui, Florence Amram, Hélène Antoine-Poirel, Séverine Audebert-Bellanger, Emmanuelle Barouk, Amandine Baurand, Marion Belléguic, Odile Béra, Pascaline Berthet, Yves-Jean Bignon, Aurélie Bisacchi, Claire Bouvattier, Myriam Bronner, Virginie Bubien, Bruno Buecher, Véronique Byrde, Cécile Cassini, Estelle Cauchin, Edith Chipoulet, Jean Chiesa, Odile Cohen-Haguenaer, Chrystelle Colas, Marie-Agnès Collonge-Rame, Christophe Cordier, François Cornelis, Carole Corsini, Isabelle Coupier, Olivier Delattre, Hélène Delhomelle, Capucine Delnatte, Liliane Demange, Bénédicte Demeer, Antoine De Pauw, Marina Di Maria, Hélène Dreyfus, Catherine Dugast, Patrick Edery, François Eisinger, Salima El Chehadeh, Natacha Entz-Werlé, Viviane Feillel, Claude Férec, Sandra Fert-Ferrer, Mélanie Fradin, Didier Frappaz, Marc Frenay, Jean-Pierre Fricker, Dominique Gaillard, Virginie Gandemer, Marion Gérard, Paul Gesta, Brigitte Gilbert-Dussardier, Anne-Paule GimenezRoqueplo, Sophie Giraud, Fabienne Giuliano, Laurence Gladieff, Philippe Gorry, Sophie Grandjouan, Marine Guillaud-Bataille, Rosine Guimbaud, Georges Haddad, Laëtitia Huiart, Olivier Ingster, Bertrand Isidor, Philippe Jonveaux, Hubert Journal, Sophie Julia, Tatiana Kogut-Kubiak, Marilyn Lackmy-Port-Lis, Khadija Laforêt, Hakima Lallaoui, Claire Laquet, Christine Lasset, Marine Lebrun, Bruno Leheup, Sophie Lejeune, Dominique Leroux, JeanMarc Limacher, Michel Longy, Alain Lortholary, Elisabeth Luporsi, Sylvie Manouvrier-Hanu, Ludovic Mansuy, Véronique Mari, Michèle Mathieu-Dramard, Christine Maugard, JeanMichaël Mazzella, Marie-Emmanuelle Morin-Meschin, Claude Moraine, Isabelle Mortemousque, Catherine Noguès, Sylvie Odent, Laurence Olivier-Faivre, Nelly Pasz, JeanPhilippe Peyrat, Magalie Peysse, Christophe Piguet, Pascal Pujol, Audrey Remenieras, Annick Rossi, Cécile Rouzier, Marc Sanson, Hagay Sobol, Joanna Sokolowska, Florent Soubrier, Marie-Domini Tabone Monnet, Isabelle Tennevet, Christel Thauvin-Robinet, Nicolas Taris, Annick Toutain, Jean-Pierre Vannier, Laurence Venat, Philippe Vennin, Alice Yvard, Mathilde Warcoïn, Marjolaine Willems, and Hélène Zattara-Cannoni.

### Appendix

**Table A1.** Different Types of Germline *TP53* Alterations Detected in 214 LFS Families

Type of Alteration	No. of Families	%
Missense mutations	143	67
Dominant-negative mutations	75	35
Splicing mutations*	23	11
Frameshift deletions/insertions	21	10
Nonsense mutations	12	6
Genomic rearrangements	9	4
In-frame deletions/insertions	6	3

Abbreviation: LFS, Li-Fraumeni syndrome.  
\*According to bioinformatic predictions and/or analyses.

**Table A2.** Age of Onset in Patients With LFS According to Specific *TP53* Alteration

Germline <i>TP53</i> Alteration	Type of Alteration	No. of Mutation Carriers (No. of families)	Age at Tumor Onset (years)			Type and Age at Onset in Childhood Tumors (years)
			Mean	Median	Range	
Complete genomic deletion	Null mutation	6 (1)	37.7	48.0	11-51	Ly 11; Os 17
c.375G>A	Splicing mutation	10 (4)	38.4	45.0	0.8-57	ACC 0.8
p.(R337H)	Non-dominant-negative missense mutation outside the p53 DBD	7 (5)	35.0	39.0	2-64	ACC 2; ACC 2
p.(R158H)	Non-dominant-negative missense mutation within the p53 DBD	11 (8)	27.9	32.0	0.5-63	ACC 0.5; ACC 0.7; ACC 1; CPC 2; ACC 17
p.(R248W)	Dominant-negative missense mutation within the p53 DBD	13 (7)	20.7	22.0	0.9-44	RMS 0.9; RMS 1; RMS 1, ACC 4; Os 16; Os 17
p.(R282W)	Dominant-negative missense mutation within the p53 DBD	15 (9)	19.6	22.0	1-39	ACC 1; ACC 1; RMS 1; CPC 9; Os 12; CNS 14; Os 17
p.(R273H)	Dominant-negative missense mutation within the p53 DBD	17 (9)	17.9	17.0	0.8-49	ACC 0.8; ACC 1; CPC 2; STS 2; ACC 4; CNS 6; CNS 9; Leu 12; T chor 17; Os 18
p.(R175H)	Dominant-negative missense mutation within the p53 DBD	14 (9)	18.6	13.0	0.4-52	CPC 0.4; CPC 1; ACC 1; RMS 2; Leu 7; CNS 8; Leu 13; RMS 13; CNS 18

Abbreviations: ACC, adrenocortical carcinoma; CPC, choroid plexus carcinoma; DBD, DNA binding domain; Leu, leukemia; Ly, lymphoma; Os, osteosarcoma; RMS, rhabdomyosarcoma; STS, soft tissue sarcoma; T chor, testis choriocarcinoma.

## Revisiting Li-Fraumeni Syndrome

**Table A3.** Distribution of the Germline Alterations in Patients With Childhood Tumors

Type of Mutation	ACC		RMS		Osteosarcomas		CNS	
	n	%	n	%	n	%	n	%
All types	34	100	22	100	20	100	29	100
Dominant-negative missense mutations	8	24	8	<b>36</b>	8	<b>40</b>	18	<b>62</b>
Non-dominant-negative missense mutations*	16	<b>47</b>	5	23	4	20	6	21
Splicing mutations	5	15	7	32	1	5	3	10
Frameshift deletions/insertions	2	6	2	9	2	10	1	3
Nonsense mutations	1	3	0	0	2	10	1	3
In frame deletions/insertions	0	0	0	0	2	10	0	0
Genomic rearrangements	2	6	0	0	1	5	0	0

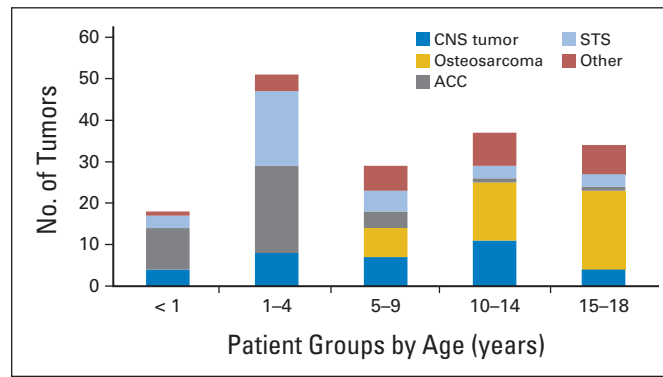
NOTE. n indicates number of patients with this tumor as first malignancy. Bold font indicates, for each tumor type, the most frequent types of mutation. Abbreviations: ACC, adrenocortical carcinoma; RMS, rhabdomyosarcoma.

\*Note that, among these mutations, some may correspond to bona fide dominant-negative mutations because not all mutations have been tested.

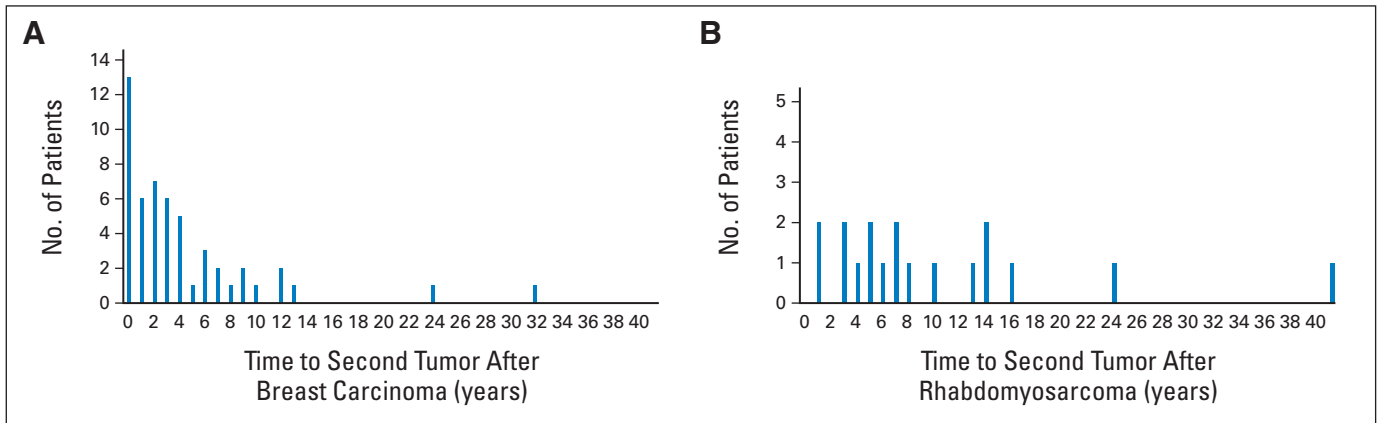
**Table A4.** Germline *TP53* Alterations Detected in Children With Adrenocortical Carcinoma

Mutation	No. of Patients
p.(Gln52HisfsTer71)	1
p.(Trp53Ter)	1
p.(Thr125Met)	2
c.375G>A, p.?	1
p.(Asn131Ser)	1
p.(Pro152Leu)	1
p.(Arg158His)	4
p.(Arg175His) <sup>DN</sup>	2
p.(His179Tyr) <sup>DN</sup>	1
c.560-2A>C, p.?	1
p.(Val216Met)	1
c.671A>C, p.?	1
c.673-2A>G	1
p.(Ile254Thr)	1
c.783-1G>A, p.?	1
p.(Arg273His) <sup>DN</sup>	3
p.(Arg273Leu) <sup>DN</sup>	1
p.(Val274PhefsTer71)	1
p.(Arg282Gly)	1
p.(Arg282Trp) <sup>DN</sup>	2
c.994-2A>T, p.?	1
p.(Arg337Cys)	2
p.(Arg337His)	2
p.(Ser378ProfsTer44)	1
Deletion exons 2-6	1
Deletion exons 2-11	1

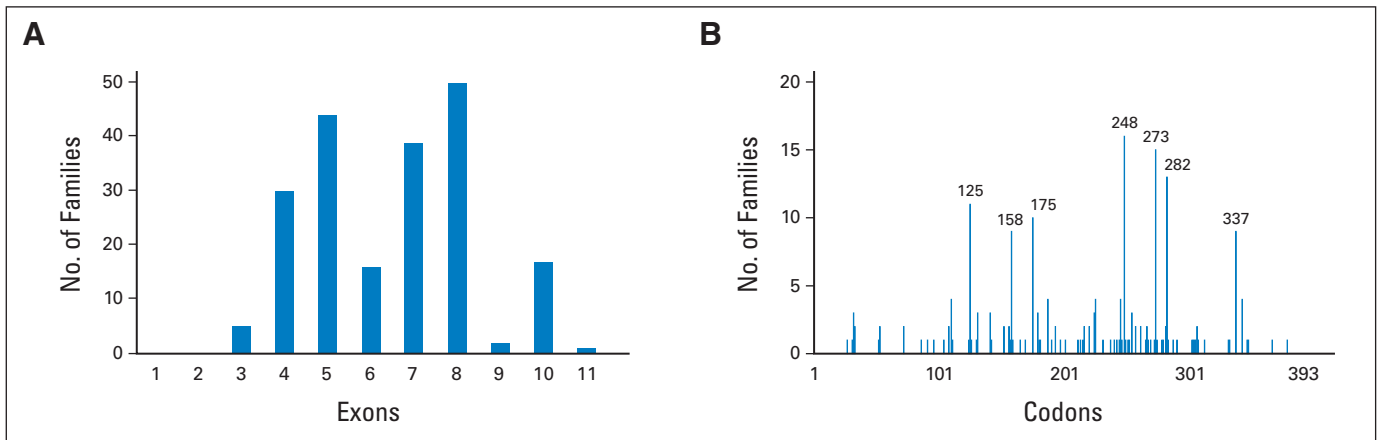
NOTE. <sup>DN</sup> indicates mutations with a dominant-negative effect according to the IARC p53 database.



**Fig A1.** Distribution of the childhood tumors according to age group. ACC, adrenocortical carcinoma; STS, soft tissue sarcoma.



**Fig A2.** Time to second tumor. (A) After breast carcinoma; (B) after rhabdomyosarcoma.



**Fig A3.** Distribution of germline *TP53* mutations. (A) According to the exons; (B) according to the codons.





## ORIGINAL ARTICLE

Contribution of de novo and mosaic *TP53* mutations to Li-Fraumeni syndrome

Mariette Renaux-Petel,<sup>1,2</sup> Françoise Charbonnier,<sup>1</sup> Jean-Christophe Théry,<sup>1,3</sup> Pierre Fermei,<sup>1</sup> Gwendoline Lienard,<sup>1</sup> Jacqueline Bou,<sup>1</sup> Sophie Coutant,<sup>1</sup> Myriam Vezain,<sup>1</sup> Edwige Kasper,<sup>1</sup> Steeve Fourneaux,<sup>1</sup> Sandrine Manase,<sup>1</sup> Maud Blanluet,<sup>1</sup> Bruno Leheup,<sup>4</sup> Ludovic Mansuy,<sup>5</sup> Jacqueline Champigneulle,<sup>6</sup> Céline Chappé,<sup>7</sup> Michel Longy,<sup>8</sup> Nicolas Sévenet,<sup>8</sup> Brigitte Bressac-de Paillerets,<sup>9</sup> Léa Guerrini-Rousseau,<sup>10</sup> Laurence Brugières,<sup>10</sup> Olivier Caron,<sup>11</sup> Jean-Christophe Sabourin,<sup>1,12</sup> Isabelle Tournier,<sup>1</sup> Stéphanie Baert-Desurmont,<sup>1</sup> Thierry Frébourg,<sup>1</sup> Gaëlle Bougeard<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2017-104976>).

For numbered affiliations see end of article.

**Correspondence to**

Professor Thierry Frébourg, Department of Genetics, Rouen University Hospital, Inserm U1245, IRIB, Normandy Centre for Genomic and Personalized Medicine, Faculty of Medicine, 22 boulevard Gambetta, 76183 Rouen CEDEX 1, France; [thierry.frebourg@chu-rouen.fr](mailto:thierry.frebourg@chu-rouen.fr)

MR-P and FC contributed equally.

Received 11 August 2017  
Revised 28 September 2017  
Accepted 9 October 2017

**ABSTRACT**

**Background** Development of tumours such as adrenocortical carcinomas (ACC), choroid plexus tumours (CPT) or female breast cancers before age 31 or multiple primary cancers belonging to the Li-Fraumeni (LFS) spectrum is, independently of the familial history, highly suggestive of a germline *TP53* mutation. The aim of this study was to determine the contribution of de novo and mosaic mutations to LFS.

**Methods and results** Among 328 unrelated patients harbouring a germline *TP53* mutation identified by Sanger sequencing and/or QMPSF, we could show that the mutations had occurred de novo in 40 cases, without detectable parental age effect. Sanger sequencing revealed two mosaic mutations in a child with ACC and in an unaffected father of a child with medulloblastoma. Re-analysis of blood DNA by next-generation sequencing, performed at a depth above 500X, from 108 patients suggestive of LFS without detectable *TP53* mutations, allowed us to identify 6 additional cases of mosaic *TP53* mutations, in 2/49 children with ACC, 2/21 children with CPT, in 1/31 women with breast cancer before age 31 and in a patient who developed an osteosarcoma at age 12, a breast carcinoma and a breast sarcoma at age 35.

**Conclusions** This study performed on a large series of *TP53* mutation carriers allows estimating the contribution to LFS of de novo mutations to at least 14% (48/336) and suggests that approximately one-fifth of these de novo mutations occur during embryonic development. Considering the medical impact of *TP53* mutation identification, medical laboratories in charge of *TP53* testing should ensure the detection of mosaic mutations.

historical definition in 1988,<sup>2</sup> we and others have sequentially updated the Chompret criteria for LFS to cover the different situations suggestive of LFS and to facilitate the recognition of this syndrome: (i) ‘familial’ criterion: proband with tumour belonging to LFS tumour spectrum (eg, premenopausal breast cancer, soft tissue sarcoma, osteosarcoma, CNS tumour, adrenocortical carcinoma (ACC)) before age 46, and at least one first or second-degree relative with LFS tumour (except breast cancer if proband has breast cancer) before age 56 or with multiple tumours; (ii) ‘multiple primitive tumours’ criterion: proband with multiple tumours (except multiple breast tumours), two of which belong to LFS tumour spectrum and first of which occurred before age 46; (iii) ‘rare tumours’ criterion: patient with ACC, choroid plexus tumour (CPT) or rhabdomyosarcoma of embryonal anaplastic subtype (anRMS) or (iv) ‘early-onset breast cancer’ criterion: female breast cancer before age 31.<sup>3–10</sup>

Indeed, certain individual presentations are highly suggestive of the presence of a germline *TP53* mutation. Independently of the familial history, the mutation detection rate in children presenting with ACC has been estimated to be 45%–66%,<sup>5 9–12</sup> with CPT to be 42%–100%,<sup>5 9</sup> with anRMS up to 73%,<sup>8</sup> and in women with breast carcinoma before age 31 to be 6%.<sup>9 13 14</sup>

The development of next-generation sequencing (NGS) technologies has recently uncovered in humans the high frequency of de novo mutations estimated to be 1.58 *per exome*<sup>15</sup> and has revealed, thanks to the reading depth, the contribution of mosaic mutations in an increased number of genetic conditions.<sup>16 17</sup> We report in this study the contribution of de novo mutations and mosaic *TP53* mutations in sporadic clinical situations strongly suggestive of LFS.

**METHODS*****TP53* molecular analysis by Sanger sequencing and QMPSF**

The 11 exons of *TP53* were PCR amplified from blood genomic DNA and analysed using Sanger sequencing on an automated sequencer 3130xl

**INTRODUCTION**

Li-Fraumeni syndrome (LFS; Mendelian Inheritance in Man (MIM) #151623) results from heterozygous germline mutations of the *TP53* tumour suppressor gene (MIM\*191170) and is characterised by a wide tumour spectrum and a diversity of age of tumour onset. Since its original recognition in 1969 by Li and Fraumeni<sup>1</sup> and its



CrossMark

**To cite:** Renaux-Petel M, Charbonnier F, Théry J-C, et al. *J Med Genet* Published Online First: [please include Day Month Year].  
doi:10.1136/jmedgenet-2017-104976

## Cancer genetics

Genetic Analyzer (Applied Biosystems), as previously described,<sup>4</sup> and variant detection was performed using the Variant Reporter software V.1.1 (Applied Biosystems). The search for genomic rearrangements was performed using quantitative multiplex PCR of short fluorescent fragments (QMPSF), as previously described.<sup>18</sup> For each patient, informed consent for genetic analyses was obtained.

### Microsatellite analysis for fatherhood and motherhood determination

To ascertain fatherhood and motherhood, blood genomic DNA samples from child and both parents were systematically PCR-amplified using primers surrounding microsatellite loci D1S439 (AFM225xe11), D9S1784 (AFMa136xa5), D14S986 (AFMa184xa5), D19S913 (AFMb301xc9), D17S960, D17S1353 and D17S1844, whose sequences are described in the online supplementary table S1, and then subjected to capillary electrophoresis on an automated sequencer 3130xl Genetic Analyzer. Data were analysed with the GeneMapper software V.4.1 (Applied Biosystems). Protocol details are available on request.

### Libraries preparation, next-generation sequencing and bioinformatic analyses

SureSelect capture for the *TP53* gene was designed using Agilent eArray (Agilent, Santa Clara, California, USA). Library preparation was performed using the QXT SureSelect enrichment kit (Agilent) from 50 ng of peripheral blood genomic DNA. DNA fragmentation was enzymatically performed by a transposase. After the enrichment step, independent indexed libraries were pooled. Libraries were sequenced on a MiSeq or a NextSeq 500 platform (Illumina, San Diego, California, USA) using 2×150 bp paired-end sequencing. Paired-end reads were mapped on the Human hg19 genome using the Burrows-Wheeler Alignment - Maximal Exact Matches (BWA-MEM) algorithm.<sup>19</sup> Postprocessing and quality score recalibration were performed using Picard and GATK software programmes (Broad Institute) according to the Broad Institute best practices. The variant caller VarScan2 V.2.3.9 (<http://dkoboldt.github.io/varscan/>) was used for the detection of single-nucleotide variants (SNV) and indels to be able to detect mosaic germline variants.<sup>20</sup> Annotation of the generated vcf files was performed using Alamut Batch (Interactive Biosoftware, Rouen, France).

### Primer extension (SNaPshot)

Mosaic *TP53* mutations were confirmed using the ABI PRISM SNaPshot™ Multiplex kit (Applied Biosystems). Each putative mutated exon of *TP53* was PCR-amplified from blood genomic DNA, as previously described. Five microlitres of the PCR product were incubated for 1 hour at 37°C with 1 U of Shrimp alkaline phosphatase (SAP) and 0.8 U of exonuclease I (ExoI) in a final volume of 10 µL, to remove excess primers and unincorporated deoxynucleotide triphosphates (dNTPs). Enzymes were deactivated at 75°C for 15 min. A multiplex primer extension was performed for each mutation on both strand, from 2 µL of purified PCR, in a final volume of 10 µL using the SNaPshot™ Multiplex Ready Reaction Mix and 0.1–0.3 µM of pooled primers (see the online supplementary table S2). Extension consisted of 25 cycles at 96°C for 10s, 50°C for 5s and 60°C for 30s. Removal of the 5' phosphoryl groups from unincorporated fluorescently labelled dideoxynucleotide triphosphates ([F]ddNTPs) was performed by phosphatase treatment using 1 U SAP on 10 µL of extended product, in a final volume of 15 µL, for 1 hour at 37°C. SAP was deactivated at 75°C for 15 min. Capillary electrophoresis of SNaPshot™ products was

performed on a 3130xl Genetic Analyzer using a 120 LIZ size standard. Data were analysed with the GeneMapper software V.4.1.

### Cloning of *TP53* mutant allele

In family F337, the genomic region surrounding *TP53* exon 3 was PCR-amplified using the following primers: forward 5' CCCCTAGCAGAGACCTGTGGG3', and reverse 5' CCCAG-CCCAACCCTTGTCCT3', and subjected to agarose electrophoresis. The aberrant band was cut from the gel, purified, phosphorylated and ligated to the pcDNA3 plasmid, previously *EcoRV* blunt digested and dephosphorylated. Bacterial clones were screened by PCR, using the T7 (5'TAATACGACTCACTATAGGG3') and Sp6 (5'ATTTAGGTGACACTATAG3') universal primers. Plasmids were isolated from positive clones and Sanger sequenced using primers T7 and Sp6.

## RESULTS

Using Sanger sequencing of the 11 exons completed by QMPSF, we have identified a deleterious heterozygous *TP53* alteration in 328 unrelated French index cases. Among these, we found that *TP53* mutations had occurred de novo in 40 cases (25 women and 15 men), including 4 cases previously reported (table 1).<sup>21</sup> This represents 12% of this series of *TP53* mutation carriers. In each case, fatherhood and motherhood were confirmed by microsatellite analyses (data not shown). For 28 index case-parent trios, we could perform NGS analysis with a 400X–1000X read depth after *TP53*-enriched capture and confirm, in each case, the absence in parent blood DNA of the *TP53* mutation detected in the index case.

Most of the patients harbouring a de novo heterozygous *TP53* mutation (34/40) had been screened for *TP53* because they fulfilled Chompret criteria. The phenotype of the patients is presented in table 1. The majority of the de novo *TP53* mutations corresponded to substitutions (18 transitions, 15 transversions) and 7 to small deletions or duplication (table 1).

As mutation rate in gonads increases with parental age and, in particular, with paternal age,<sup>22 23</sup> we evaluated the age of the parents for all index cases harbouring a de novo *TP53* mutation. At birth of the index cases, we observed a mean and median age for fathers of 31 years (range 19–41 years) and for mothers of 28 years (range 19–47 years). These values are similar to the mean ages of childbirth observed in the general population in France, corresponding to 33 and 30 years old, in men and women, respectively (data for 2015, from Insee, the French National Institute of Statistics and Economic Studies).

We detected two cases of mosaic mutation thanks to Sanger sequencing. In family F281, we had previously identified in the index case, who presented with a sporadic medulloblastoma at 3 years of age, a germline *TP53* mutation within exon 8 (NM\_000546.5: c.814G>A, p.(Val272Met)). Targeted Sanger sequencing in both parents initially suggested that the mutation had occurred de novo in the germline. However, careful examination of the Sanger electropherograms generated from the healthy father blood DNA revealed a very weak signal corresponding to the mutant peak, which height was in the same magnitude than the sequencing background noise (figure 1). This suggested the existence of a mosaic *TP53* mutation in the father. This was confirmed both by SNaPshot primer extension (figure 2) and NGS analysis, showing in father blood DNA the presence of the mutation in 43 among 866 reads (5%) (table 2, see the online supplementary figure S1). The transmission of the mutation to his daughter indicated that this mosaic mutation was also present in the gonads. In another case (F315), corresponding to a boy who presented at 4 months of age with a non-secreting

**Table 1** Patients with heterozygous de novo *TP53* mutations detected by Sanger sequencing

Family	Sex*	Phenotype, age at diagnosis†	Mutation‡	Predicted effect on the protein§	Percentage of mutant reads¶
F192	M	RMS, 1 / OS, 11+19	c.844C>T	p.(Arg282Trp)	55% (424/767)
F70	F	LMS, 39 / BC, 42	c.476C>A	p.(Ala159Asp)	52% (489/932)
F194	M	BL, 2 / OS, 16 / GA, 17	c.761T>A	p.(Ile254Asn)	52% (552/1059)
F304	F	ACC, 21	c.451C>G	p.(Pro151Ala)	51% (411/799)
F256	F	ALL, 16	c.743G>A	p.(Arg248Gln)	49% (389/788)
F267	F	OS, 16 / BC, 28	c.743G>A	p.(Arg248Gln)	49% (517/1048)
F234	M	RMS, 2 / OS, 12	c.518T>C	p.(Val173Ala)	49% (524/1067)
F126	M	WT, 2 / MB, 7	c.378C>G	p.(Tyr126*)	49% (354/724)
F279	F	ACC, 5	c.527G>A	p.(Cys176Tyr)	49% (475/974)
F215	F	ACC, 25 / BC, 28	c.844C>T	p.(Arg282Trp)	48% (327/678)
F66	F	OS, 18 / BilatBC, 34+35	c.310C>T	p.(Gln104*)	48% (360/750)
F270	M	BCC, 33 / UPS, 40	c.613T>G	p.(Tyr205Asp)	48% (307/643)
F60	F	Bilat BC, 25+35	c.1009C>T	p.(Arg337Cys)	47% (482/1015)
F50	F	RMS, 1 / OS, 2	c.814G>A	p.(Val272Met)	47% (243/517)
F143	F	RMS, 1 / RMS, 2	c.672+1G>T	p.?	47% (337/717)
F235	M	CPC, 0.5	c.365_366del	p.(Val122Aspfs*26)	46% (311/672)
F174	F	OS, 11	c.919+1G>T	p.?	43% (284/667)
F47	F	BC, 26 / GB, 27	c.393_395del	p.(Asn131del)	42% (321/761)
F83	M	STS, 38+44	c.329G>C	p.(Arg110Pro)	42% (352/839)
F171	F	CPC, 1	c.524G>A	p.(Arg175His)	41% (354/865)
F38**	M	OS, 19 / CS, 31	c.842A>T	p.(Asp281Val)	41% (266/655)
F148	F	CPC, 17 / BS, 24	c.845G>C	p.(Arg282Pro)	38% (225/585)
F199	F	Bilat BC, 30	c.782+1G>A	p.?	38% (452/1200)
F287	F	HD, 16 / AML, 24 / BC, 25	c.428T>C	p.(Val143Ala)	34% (302/877)
F266	M	GB, 3	c.731G>A	p.(Gly244Asp)	34% (363/1077)
F127	F	Bilat BC, 29	c.913_916del	p.(Lys305Glu fs*39)	30% (175/575)
F186	F	Bilat BC, 25	c.632_641del	p.(Thr211Ile fs*33)	28% (202/733)
F277	M	MB, 9 / OS, 13	c.96+31_97-32del	p.?	14% (76/550)
F7	F	STS, 2 / ACC, 6 / OS, 14 / Bilat BC, 24+25	c.535C>T	p.(His179Tyr)	ND
F62**	F	ACC, 1 / RMS, 1 / OS, 17	c.818G>T	p.(Arg273Leu)	ND
F96**	M	CPC, 4	c.733G>A	p.(Gly245Ser)	ND
F115	F	Bilat BC, 27	c.390_426del	p.(Asn131Cys fs*27)	ND
F118**	M	MB, 10	c.376-2A>G	p.?	ND
F155	F	BC, 52	c.524G>A	p.(Arg175His)	ND
F180	M	Liposarcoma, 46 / AML, 48	c.743G>A	p.(Arg248Gln)	ND
F245	F	CRC, 24 / Bilat BC, 31+34	c.374C>A	p.(Thr125Lys)	ND
F321	F	BC, 29 / OS, 35	c.681dup	p.(Asp228*)	ND
F327	M	No tumour, 33	c.535C>G	p.(His179Asp)	ND
F342	F	Bilat BC, 26	c.817C>G	p.(Arg273Gly)	ND
F343	M	ACC, 1	c.844C>T	p.(Arg282Trp)	ND

\*F, female; M, male.

†Age of tumour onset in years is indicated.

‡cDNA numbering with the first nucleotide corresponding to the A of the ATG translation initiation codon in the reference sequence (GenBank RefSeq-file accession number NM\_000546.5).

§Protein numbering with the initiation codon numbered as codon 1 (accession number NP\_000537.3).

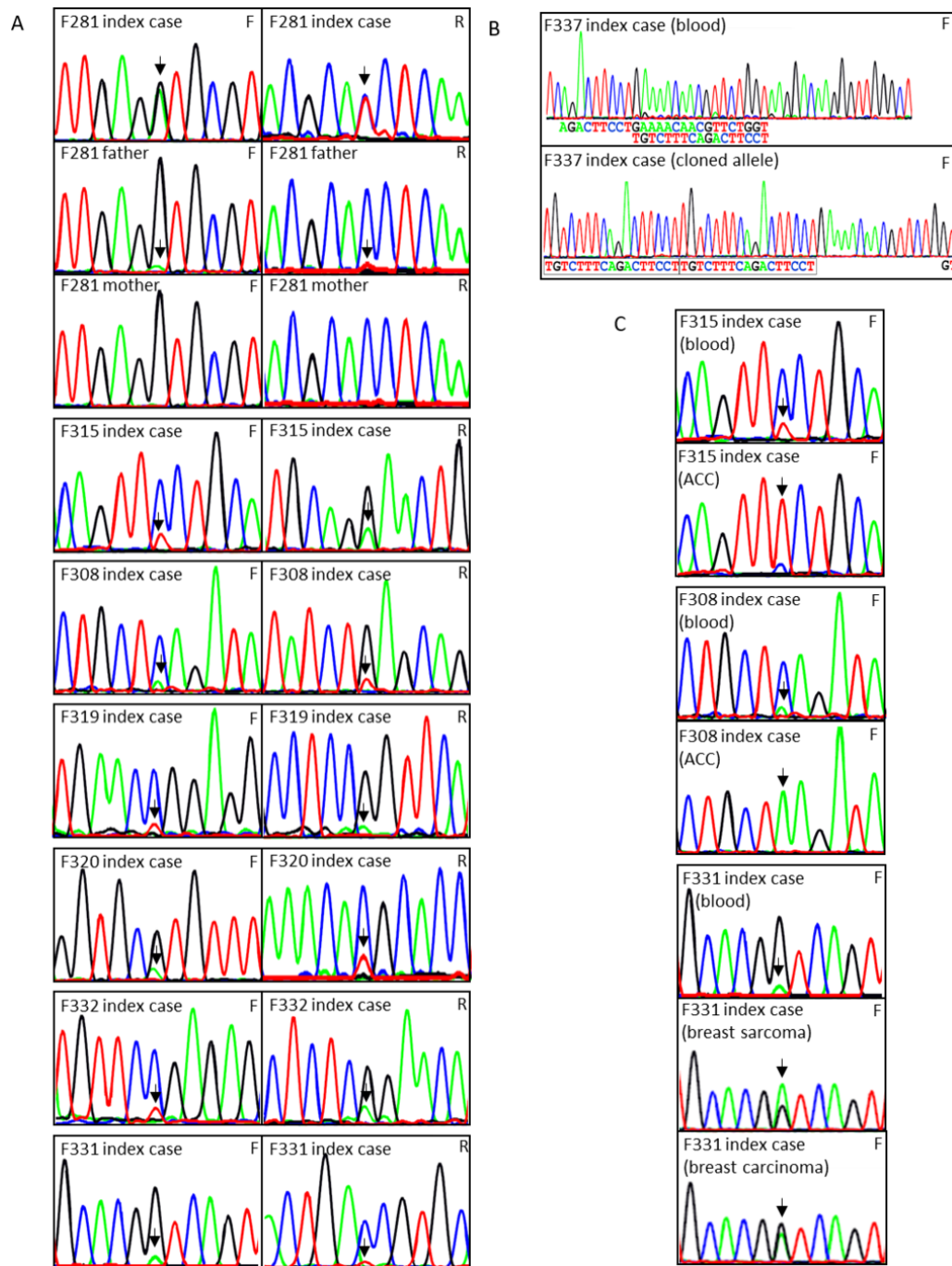
¶ND, not determined.

\*\*Previously described in Chompret *et al.* (Br J Cancer 2000).<sup>21</sup>

ACC, adrenocortical carcinoma; ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; BC, breast carcinoma; BCC, basal cell carcinoma; Bilat, bilateral; BL, Burkitt's lymphoma; BS, breast sarcoma; CPC, choroid plexus carcinoma; CRC, colorectal carcinoma; CS, chondrosarcoma; GB, glioblastoma; GA, gastric adenocarcinoma; HD, Hodgkin's disease; LMS, leiomyosarcoma; MB, medulloblastoma; OS, osteosarcoma; RMS, rhabdomyosarcoma; STS, soft tissue sarcoma; UPS, undifferentiated pleomorphic sarcoma; WT, Wilm's tumour.

voluminous ACC, Sanger sequencing also revealed a mosaic missense *TP53* mutation in exon 7c.[722C=>T], p.[Ser241=>Phe] (Ser241Phe) (figure 1). This mutation was confirmed by SNaPshot analysis (figure 2) and by NGS (table 2, see the online supplementary figure S1) showing 17% of mutant alleles (181/1060).

Then, in order to detect more extensively mosaic *TP53* mutations, we reanalysed by NGS performed with a reading depth above 500X, 108 other patients with clinical presentations strongly suggestive of LFS, corresponding either to ACC (n=55), CPT (n=21), breast carcinoma before age 31 (n=31) or multiple primary

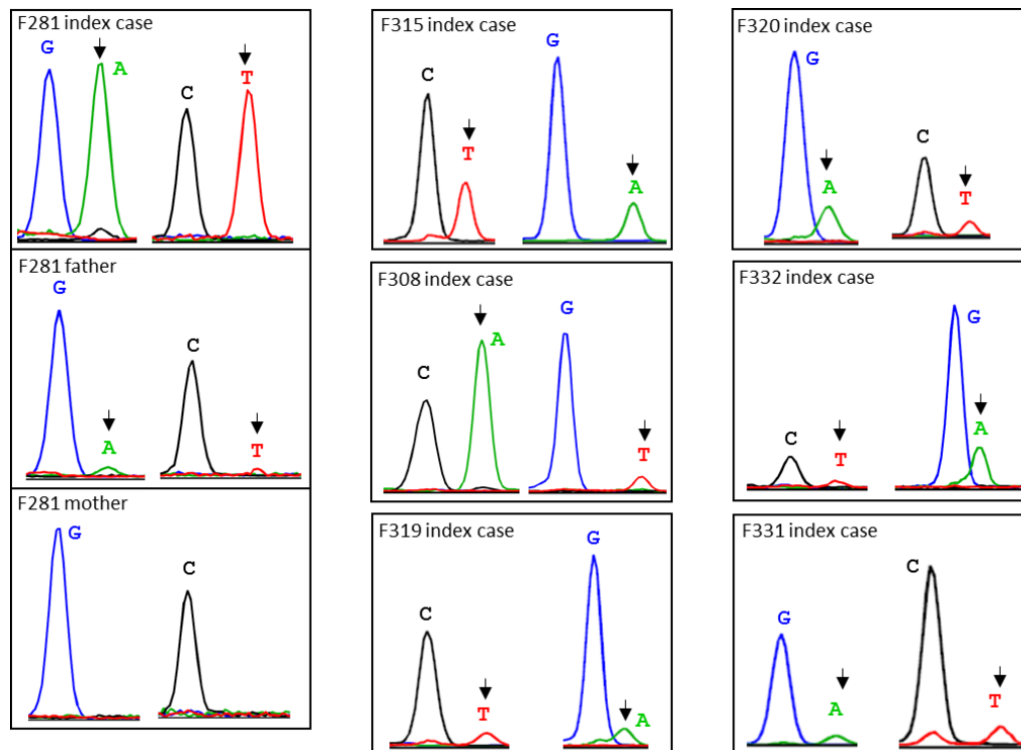


**Figure 1** Sanger sequencing of the *TP53* gene. (A) Mosaic *TP53* mutations detected in blood samples. Each panel corresponds to a family. The arrows indicate the *TP53* mutations: c.814G>A in F281, c.722C>T in F315, c.548C>A in F308, c.742C>T in F319, c.818G>A in F320, c.1024C>T in F332, and c.375+1G>A in F331. (B) Complex mosaic *TP53* duplication, encompassing intron 2 and exon 3, detected in F337: sequencing from blood (upper panel); sequencing of the mutant allele cloned into pcDNA3 (lower panel). (C) Detection of *TP53* mutations in tumour samples. c.722C>T in F315, c.548C>A in F308, and c.375+1G>A in F331. Concomitant detection of the wild-type peaks is explained by the presence of non-malignant cells within the tumour samples. ACC, adrenocortical carcinoma; F, forward sequence; R, reverse sequence.

tumours (n=1) but without detectable *TP53* mutations using Sanger and QMPFSF.

First, we analysed children with ACC. Sanger sequencing of *TP53* performed on an updated ACC cohort to 146 patients (97 children, 49 adults) had revealed a *TP53* mutation in a total of 45 children (46%), including the mosaic mutation described above, and 7 adults (14%) (1 man and 6 women) in agreement with our previous estimates obtained on a smaller series.<sup>9</sup> We reanalysed, using NGS, blood DNA from 49 children (31

women (0.1–18 years), 18 men (0.1–17 years)) and 6 adults (4 women (20–58 years), 2 men (24–50 years)) with ACC but without detectable *TP53* mutations. This revealed two additional cases of mosaic mutation in early-onset ACC: (i) in a girl (F308) who at 8 months of age presented with an ACC revealed by hyperandrogenism and became metastatic at age 3, a nonsense mutation c.[548C=>A], p.[Ser183=>(Ser183\*)] in exon 5 was detected in 17% of the NGS reads (145/832) (table 2, figure 3); (ii) in a girl (F337) who presented with



**Figure 2** Confirmation of mosaic *TP53* mutations using primer extension. *TP53* mutations were detected using the ABI PRISM SNaPshot™ Multiplex kit (Applied Biosystems), in non-quantitative conditions. Each panel corresponds to a family. For each mutation, the two left and two right peaks show the extension of the sense and antisense strands, respectively. Arrows indicate the mutant peaks. The ddNTPs are dye-labelled with dR6G (ddATP, green), dTAMRA (ddCTP, yellow=black), dR110 (ddGTP, blue) and dROX (ddUTP, red). The arrows indicate the *TP53* mutations: c.814G>A in F281, c.722C>T in F315, c.548C>A in F308, c.742C>T in F319, c.818G>A in F320, c.1024C>T in F332, and c.375+1G>A in F331. ddNTPs, dideoxynucleotide triphosphates.

a 5 cm ACC at 14 months of age, a partial *TP53* duplication extending from intron 2 to exon 3, was detected in 4% of the reads (25/571) (table 2). A specific PCR amplification and Sanger sequencing (figure 1) of the cloned aberrant fragment confirmed the 17bp duplication, annotated by Varscan2 as c.[75-10\_81=dup] and predicted to result into a mutant truncated protein p.[Glu28=/(Glu28Cysfs\*22)].

In an extended series of patients with CPT composed of 43 patients, including only one adult case, Sanger sequencing had revealed a heterozygous mutation in 17 patients (40%), in agreement with our previous estimate.<sup>9</sup> Twenty-one children with

CPT (14 boys, 7 girls, range 0.1–16 years), without detectable mutations, were then reanalysed by NGS. This allowed us to identify two other cases of mosaic *TP53* mutations affecting hot spots: (i) c.[742C=/>T], p.[Arg248=/(Arg248Trp)] within exon 7 detected in 14% of the reads (263/1933) in a boy (F319) who presented with a sporadic choroid plexus carcinoma at 2 years of age, and (ii) c.[818G=/>A], p.[Arg273=/(Arg273His)], within exon 8 in 19% of the reads (359/1908) in a boy (F320) who presented with an atypical sporadic choroid plexus papilloma at 6 months of age (table 2, see the online supplementary figure S1).

**Table 2** Patients with mosaic *TP53* mutations

Family	Sex*	Phenotype, age at diagnosis†	Mutation‡	Predicted effect on the protein§	Percentage of mutant reads	Tumour analysis¶
F281	M	No tumour, 42	c.[814G=/>A]	p.[Val272=/(Val272Met)]	5% (43/866)	/
F315	M	ACC, 0.3	c.[722C=/>T]	p.[Ser241=/(Ser241Phe)]	17% (181/1060)	LOH
F308	F	ACC, 0.7	c.[548C=/>A]	p.[Ser183=/(Ser183*)]	17% (145/832)	LOH
F337	F	ACC, 1	c.[75-10_81=dup]	p.[Glu28=/(Glu28Cysfs*22)]	4% (25/571)	ND
F319	M	CPC, 2	c.[742C=/>T]	p.[Arg248=/(Arg248Trp)]	14% (263/1928)	ND
F320	M	Atypical CPP, 0.5	c.[818G=/>A]	p.[Arg273=/(Arg273His)]	19% (359/1902)	ND
F332	F	Bilat BC, 27+34	c.[1024C=/>T]	p.[Arg342=/(Arg342*)]	17% (699/4120)	ND
F331	F	OS, 12 / L. BC, 35 / R. BS, 35	c.[375+1G=/>A]	p.[=?]	7% (40/551)	LOH

\*F, female; M, male.

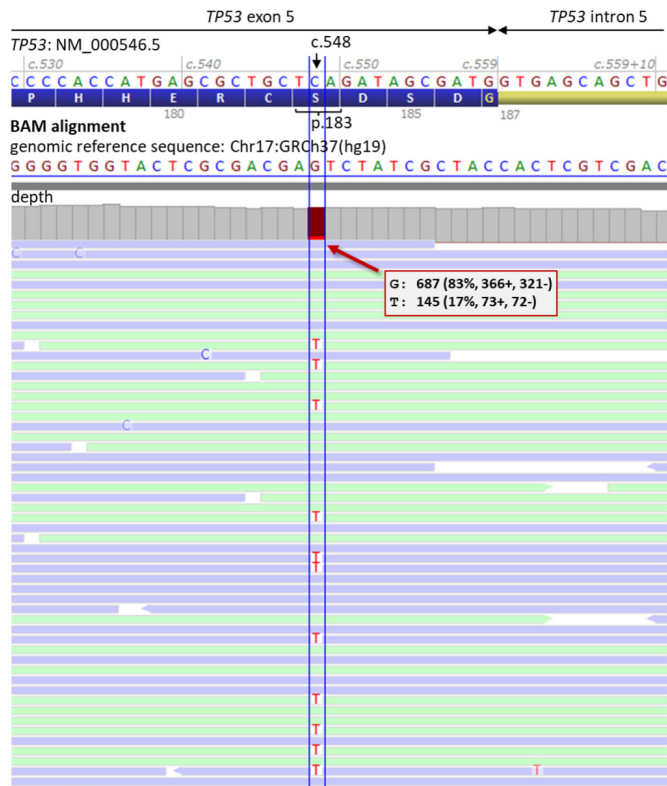
†Age of tumour onset in years is indicated.

‡cDNA numbering with the first nucleotide corresponding to the A of the ATG translation initiation codon in the reference sequence (GenBank RefSeq-file accession number NM\_000546.5).

§Protein numbering with the initiation codon numbered as codon 1 (accession number NP\_000537.3).

¶LOH, loss of heterozygosity; ND, not determined.

ACC, adrenocortical carcinoma; BC, breast carcinoma; Bilat, bilateral; BS, breast sarcoma; CPC, choroid plexus carcinoma; CPP, choroid plexus papilloma; L, left; OS, osteosarcoma; R, right.



**Figure 3** BAM visualisation of the mosaic *TP53* mutation in patient F308, using Alamut visual software (Interactive Biosoftware). Partial sequence of *TP53* NM\_000546.5 is shown. The arrow indicates the *TP53* mutation c.548C>A, p.(Ser183\*), into exon 5. As *TP53* is located on chromosome 17 on reverse strand, the non-reference base detected is a thymine (T). Paired-end DNA sequencing reads are visualised in blue (strands +) and green (strands -). The total number of reads, the percentage and the number of reads on each strand are indicated for the wild-type and the mutant alleles.

Considering that the occurrence of breast cancer in women at a very early age (<31 years) is also suggestive of a germline *TP53* mutation, with a mutation detection rate estimated to be 4%–6%,<sup>9 13 14</sup> blood genomic DNA from 31 women with very early-onset breast cancer (mean 26 years, range 20–30 years) but without *TP53* mutations detectable by Sanger sequencing and QMPSPF were submitted to NGS analysis. A mosaic *TP53* nonsense mutation was detected in 17% of the reads (699/4120) in a woman who presented with a bilateral breast cancer at 27 and 34 years of age (F332): c.[1024C=>T], p.[Arg342=>(Arg342\*)], exon 10 (table 2, see the online supplementary figure S1).

Finally, we reanalysed by NGS a patient (F331), who had developed multiple early primary tumours strongly suggestive of a *TP53* mutation. This female patient had developed an osteosarcoma of the skull diagnosed at 12 years of age and treated by surgery and chemotherapy, then at 35 years of age, she presented with a left breast ductal carcinoma in situ (DCIS) and a right breast sarcoma. While Sanger sequencing had initially failed to detect *TP53* mutation, NGS performed on blood DNA revealed on 7% of the reads (40/551) a splicing mutation affecting the donor site of intron 4: c.[375+1G=>A], p.[=?] (table 2, see the online supplementary figure S1).

In these six additional mosaic *TP53* mutations revealed by NGS, subsequent re-examination of the Sanger electropherograms revealed a trace of the mutant allele (figure 1).

Interestingly, the presence of the mosaic mutation was detectable using SNaPshot analysis (figure 2). Among the eight cases of mosaic *TP53* mutations, tumour DNA was available only in cases F315, F308 and F331 and Sanger sequencing showed, in each case, the presence of the mutation and LOH within the tumour (figure 1). In particular, in patient F331, who had developed multiple primary tumours, the mutation was detected in both tested tumours (figure 1C), confirming the mosaicism.

## DISCUSSION

In this study, we report the detection of a total of 48 de novo *TP53* mutations including 8 cases of mosaic mutations (tables 1 and 2). The contribution of de novo *TP53* mutations to LFS had already been established by several reports, but mostly focused on single cases.<sup>24–39</sup> From a series of 75 patients with a germline *TP53* mutation, for whom clinical presentation was not detailed, Gonzalez *et al* reported that in at least 5 cases (7%), the mutation had occurred de novo.<sup>31</sup> In our study, de novo *TP53* mutations represented 14% (48/336) of the mutations identified in the index cases. As DNA was not systematically available from both parents of each index case, it is likely that this percentage underestimates the real contribution of de novo *TP53* mutations to LFS. In this series of 336 LFS cases, we could establish that the mutation was inherited in 110 cases and for 43 cases, a family history very suggestive of LFS allowed us to assume that the mutation was also inherited, indicating that the contribution of inherited mutations to LFS can be estimated to be at least 46%. For 55 cases (16%), the familial history suggested a possible inheritance and, in the 80 remaining other cases (24%), the absence of suggestive familial history could be explained either by incomplete penetrance of some mutations or additional cases of de novo mutations. For the 47 childhood ACC and 19 CPT cases associated with *TP53* mutations, this repartition between proven or presumed inherited mutations, possibly inherited mutations, presumed incomplete penetrance or de novo mutations and proven de novo mutations was, respectively, 62% and 53%; 6% and 0%; 19% and 16% and 13% and 32%. The contribution of de novo mutations to LFS, in complement to the incomplete penetrance of *TP53* mutations, reinforces the message that the diagnosis of LFS should be considered, independently of the familial history, in patients with strongly suggestive clinical presentations such as ACC, CPT, early-onset breast cancers or multiple primary cancers belonging to the LFS spectrum.

Only few *TP53* mosaic mutations have been reported so far.<sup>40–42</sup> Thanks to the sensitivity of NGS performed at a high depth on blood, we found, among the 48 cases with *TP53* de novo mutations, a total of 8 cases of mosaic mutations, indicating that in 17% (8/48) of these cases the *TP53* mutation had occurred during embryonic development (table 2). It has been suggested that chemotherapy and, in particular extended chemotherapy, may induce somatic *TP53* mutations in haematopoietic cells, mimicking therefore the presence of mosaic mutations.<sup>43</sup> Among the eight cases with a *TP53* mutation detected in a small fraction of the reads generated from blood DNA, we could show in the three cases (F308, F315, F331) for whom tumour DNA was available, the presence of the mutation and LOH within the tumour (figure 1), supporting the involvement of the *TP53* mutation in tumour development. In an additional case (F337), the patient had not received chemotherapy, refuting this hypothesis.

An alternative hypothesis would be that the mosaic mutations detected in blood from the six affected individuals (F315, F308, F337, F319, F320, F332 and F331) corresponded to somatic mutations and reflected the presence of circulating tumour DNA. We think that we can exclude this hypothesis, at least in four of

six affected patients: (i) in patient F331 who developed multiple primary tumours, the mutation was detected in two different primary tumours, (figure 1C) confirming the true mosaicism; (ii) as highlighted by numerous studies, the presence of circulating tumour DNA is correlated with tumour burden.<sup>44</sup> In patients F315 and F308 with ACC, NGS was performed in both cases on blood sampling 3 weeks after the complete surgical resection of the tumours; in patient F319 with CPC who was treated by chemotherapy and surgery, brain MRI performed at the time of blood sampling had revealed no brain tumour recurrence; and in patient F332 who developed bilateral breast cancer at 27 and 34 years of age, blood sampling was performed 18 years after the second tumour.

In contrast, we cannot exclude that we have underestimated the contribution of *TP53* mosaic mutations to LFS. First, despite the remarkable sensitivity of NGS sequencing performed at a high depth, the detection of mosaic mutations from blood DNA requires that they are present within the haematopoietic lineage. As recently shown in adenomatous polyposis,<sup>45</sup> only NGS performed on non-malignant tissues from tumours would allow providing an accurate estimate. Second, as shown in table 1, among the 40 cases with de novo mutations initially detected by Sanger, we analysed by NGS 28 trios and observed, for some index cases, an allelic imbalance up to 28%. Although we cannot exclude that this may be explained by a technical artefact during the capture or sequencing process, it might suggest that, at least for some of these patients, the mutation was mosaic and had occurred not at the prezygotic but postzygotic stage. Our study illustrates how NGS technologies, as previously highlighted,<sup>16 17</sup> provide new opportunities to access mosaic mutations undetectable or hardly detectable by Sanger sequencing.

Including mosaic mutations, we detected in blood from 97 children with ACC and 42 with CPT, a *TP53* mutation in 47 and 19 patients, respectively, corresponding to a mutation detection rate of 48% and 45%. This confirms that inactivation of *TP53* is the main genetic cause of the development of ACC and CPT, which should be considered as emblematic tumours of LFS.<sup>5 9–12</sup>

Among a limited number of patients who developed breast cancer before 31 years of age or multiple primary cancers including early-breast cancers, we identified a mosaic mutation in a patient with bilateral breast cancer and in another who developed three cancers including a bilateral breast cancer at 35 years of age. This confirms that these both clinical presentations are also suggestive of the presence of a *TP53* mutation.

Identification of a *TP53* mosaic mutation has major consequences for genetic counselling in families. Indeed, the occurrence of a malignancy in a child or a young adult creates a legitimate anxiety in the parents and siblings concerned by the risk of recurrence. Demonstration that the *TP53* mutation detected in the index case has occurred at the postzygotic stage allows reassuring parents and siblings. In contrast, as illustrated by family F281, mosaic mutations detected in blood may also be present in germline cells and may expose the offspring to the same risk as non-mosaic mutations. Identification of a mosaic *TP53* mutation has also drastic impacts for the patient. Indeed, it has now been established that in *TP53* mutation carriers, radiotherapy contributes to the development of secondary tumours and should therefore be, whenever possible, avoided.<sup>9</sup> This rule should also probably be applied to patients with mosaic *TP53* mutations as it is not possible, at the present time, to determine on a routine basis which tissues carry the mutation. Patients with mosaic mutations should also benefit from surveillance protocols, such as those which have recently been elaborated for germline *TP53* mutation carriers. These protocols are based, from the first year of age, on abdominal ultrasound every 6 months, annual total body MRI, annual brain MRI and in women from 20 years on annual breast

MRI.<sup>46</sup> Nevertheless, the benefits of such heavy protocols regardless to the psychological impacts might be questionable in these patients with mosaic *TP53* mutations, as one cannot exclude that the mosaic mutation might be restricted to certain tissues.

In conclusion, we have shown from a large series of *TP53* mutation carriers that de novo mutations represent 14% (48/336) of the mutations, confirming that familial history of cancer is not mandatory to consider the presence of a *TP53* mutation, that certain sporadic cancers such as ACC, CPT in children or female breast cancers occurring before age 31 or multiple primary cancers belonging to the LFS spectrum are highly suggestive of the presence of a *TP53* mutation. Considering the medical impacts for the patients and the families of a *TP53* mutation, we think that medical laboratories in charge of *TP53* testing should ensure in patients with these clinical presentations detection of mosaic mutations, by performing NGS on blood at a high depth or on non-malignant cells from tumour specimens.

#### Author affiliations

<sup>1</sup>Department of Genetics, Normandy Centre for Genomic and Personalized Medicine, Normandie University, UNIROUEN, Inserm U1245 and Rouen University Hospital, Rouen, France

<sup>2</sup>Department of Paediatric Surgery, Rouen University Hospital, Rouen, France

<sup>3</sup>Department of Medical Oncology, Henri Becquerel Centre, Rouen, France

<sup>4</sup>Department of Clinical Genetics, Nancy University Hospital, Nancy, France

<sup>5</sup>Department of Paediatric Oncology, Nancy University Hospital, Nancy, France

<sup>6</sup>Department of Pathology, Nancy University Hospital, Nancy, France

<sup>7</sup>Department of Paediatric Oncology, Rennes University Hospital, Rennes, France

<sup>8</sup>Department of Molecular Genetics, Bergonié Institute, Bordeaux, France

<sup>9</sup>Department of Medical Biology and Pathology, and Inserm U1186, Gustave Roussy, University of Paris-Saclay, Villejuif, France

<sup>10</sup>Child and Adolescent Cancer Department, Gustave Roussy Cancer Campus, Villejuif, France

<sup>11</sup>Department of Oncology, Gustave Roussy Cancer Campus, Villejuif, France

<sup>12</sup>Department of Pathology, Rouen University Hospital, Rouen, France

**Acknowledgements** The authors gratefully acknowledge the LFS families for their helpful contribution to this study. The authors are grateful to the French LFS network, with special thanks to C Abadie, S Audebert, E Barouk, P Benusiglio, P Berthet, Y J Bignon, B Buecher, V Bonadona, C Colas, MA Collonge-Rame, I Couplier, O Delattre, P Denizeau, H Dreyfus, M Gauthier-Villars, P Gesta, S Giraud, O Ingster, S Lejeune, D Leroux, J M Limacher, E Luporsi, E Morin-Meschin, I Mortemousque, L Olivier-Faivre, M F Petit, D Stoppa-Lyonnet, A Toutain, L Venat-Bouvet, P Vennin and Hélène Zattara. The authors acknowledge the NGS core facility located at the Faculty of Medicine of Rouen.

**Contributors** GB and TF: conceived the project. BBP, LB, OC, CC, JC, LG, BL, ML, LM, J-CS and NS: contributed to patient recruitment and/or provided materials. MB, JB, FC, PF, SF, EK, GL, SM, MR-P and J-CT: contributed to molecular analyses. MV and SC: performed bioinformatics analyses. SB-D, JB, GB, FC, PF, TF, GL, M-RP, NS, J-CT and IT: performed data analysis and interpretation. MB, JB, GB, FC, SC, TF, M-RP and IT: wrote the manuscript. All authors: approved the manuscript.

**Funding** This work was supported by the INCa, the French National Cancer Institute and the ARC Foundation for Cancer Research.

**Competing interests** None declared.

**Patient consent** Obtained.

**Provenance and peer review** Not commissioned; externally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

#### REFERENCES

- Li FP, Fraumeni JF. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med* 1969;71:747–52.
- Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, Miller RW. A cancer family syndrome in twenty-four kindreds. *Cancer Res* 1988;48:5358–62.
- Chompret A, Abel A, Stoppa-Lyonnet D, Brugières L, Pagès S, Feunteun J, Bonaïti-Pellié C. Sensitivity and predictive value of criteria for p53 germline mutation screening. *J Med Genet* 2001;38:43–7.
- Bougeard G, Sesboué R, Baert-Desurmont S, Vasseur S, Martin C, Tinat J, Brugières L, Chompret A, de Paillerets BB, Stoppa-Lyonnet D, Bonaïti-Pellié C, Frébourg T. French



- LFS working group. Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *J Med Genet* 2008;45:535–8.
- 5 Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, Han JH, Lowstuter K, Longmate J, Sommer SS, Weitzel JN. Beyond Li Fraumeni syndrome: clinical characteristics of families with p53 germline mutations. *J Clin Oncol* 2009;27:1250–6.
  - 6 Tinat J, Bougeard G, Baert-Desurmont S, Vasseur S, Martin C, Bouvignies E, Caron O, Bressac-de Paillerets B, Berthet P, Dugast C, Bonaïti-Pellié C, Stoppa-Lyonnet D, Frébourg T. 2009 version of the Chompret criteria for Li Fraumeni syndrome. *J Clin Oncol* 2009;27:e108–9.
  - 7 Ruijs MW, Verhoef S, Rookus MA, Pruntel R, van der Hout AH, Hogervorst FB, Kluijft I, Sijmons RH, Aalfs CM, Wagner A, Ausems MG, Hoogerbrugge N, van Asperen CJ, Gomez Garcia EB, Meijers-Heijboer H, Ten Kate LP, Menko FH, van't Veer LJ. TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. *J Med Genet* 2010;47:421–8.
  - 8 Hettner S, Archer NM, Somers GR, Novokmet A, Wagers AJ, Diller L, Rodriguez-Galindo C, Teot LA, Malkin D. Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. *Cancer* 2014;120:1068–75.
  - 9 Bougeard G, Renaux-Petel M, Flaman JM, Charbonnier C, Fermey P, Belotti M, Gauthier-Villars M, Stoppa-Lyonnet D, Consolino E, Brugières L, Caron O, Benusiglio PR, Bressac-de Paillerets B, Bonadona V, Bonaïti-Pellié C, Tinat J, Baert-Desurmont S, Frébourg T. Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. *J Clin Oncol* 2015;33:2345–52.
  - 10 Wasserman JD, Novokmet A, Eichler-Jonsson C, Ribeiro RC, Rodriguez-Galindo C, Zambetti GP, Malkin D. Prevalence and functional consequence of TP53 mutations in pediatric adrenocortical carcinoma: a children's oncology group study. *J Clin Oncol* 2015;33:602–9.
  - 11 Wagner J, Portwine C, Rabin K, Leclerc JM, Narod SA, Malkin D. High frequency of germline p53 mutations in childhood adrenocortical cancer. *J Natl Cancer Inst* 1994;86:1707–10.
  - 12 Varley JM, McGown G, Thorncroft M, James LA, Margison GP, Forster G, Evans DG, Harris M, Kelsey AM, Birch JM. Are there low-penetrance TP53 Alleles? evidence from childhood adrenocortical tumors. *Am J Hum Genet* 1999;65:995–1006.
  - 13 Mouchawar J, Korch C, Byers T, Pitts TM, Li E, McCredie MR, Giles GG, Hopper JL, Southey MC. Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res* 2010;70:4795–800.
  - 14 Evans DG, Moran A, Hartley R, Dawson J, Bulman B, Knox F, Howell A, Laloo F. Long-term outcomes of breast cancer in women aged 30 years or younger, based on family history, pathology and BRCA1/BRCA2/TP53 status. *Br J Cancer* 2010;102:1091–8.
  - 15 Gillissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, Tearle R, Bo T, Pfundt R, Yntema HG, de Vries BB, Kleefstra T, Brunner HG, Vissers LE, Veltman JA. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344–7.
  - 16 Gajeka M. Unrevealed mosaicism in the next-generation sequencing era. *Mol Genet Genomics* 2016;291:513–30.
  - 17 Qin L, Wang J, Tian X, Yu H, Truong C, Mitchell JJ, Wierenga KJ, Craigen WJ, Zhang VW, Wong LC. Detection and quantification of mosaic mutations in disease genes by next-generation sequencing. *J Mol Diagn* 2016;18:446–53.
  - 18 Bougeard G, Brugières L, Chompret A, Gesta P, Charbonnier F, Valent A, Martin C, Raux G, Feunteun J, Bressac-de Paillerets B, Frébourg T. Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. *Oncogene* 2003;22:840–6.
  - 19 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. [q-bio.GN] <https://arxiv.org/abs/1303.3997>
  - 20 Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–76.
  - 21 Chompret A, Brugières L, Ronsin M, Gardes M, Dessarps-Freichay F, Abel A, Hua D, Ligot L, Dondon MG, Bressac-de Paillerets B, Frébourg T, Lemerle J, Bonaïti-Pellié C, Feunteun J. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *Br J Cancer* 2000;82:1932–7.
  - 22 Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, Dominiczak A, Morris A, Porteous D, Smith B, Stratton MR, Hurles ME. UK10K Consortium. Timing, rates and spectra of human germline mutation. *Nat Genet* 2016;48:126–33.
  - 23 Goldmann JM, Wong WS, Pinelli M, Farrar T, Bodian D, Stittrich AB, Glusman G, Vissers LE, Hoischen A, Roach JC, Vockley JG, Veltman JA, Solomon BD, Gillissen C, Niederhuber JE. Parent-of-origin-specific signatures of de novo mutations. *Nat Genet* 2016;48:935–9.
  - 24 Gutiérrez MI, Bhatia KG, Barreiro C, Spangler G, Schwartzmann E, Muriel FS, Magrath IT. A de novo p53 germline mutation affecting codon 151 in a six year old child with multiple tumors. *Hum Mol Genet* 1994;3:2247–8.
  - 25 Ayan I, Luca J, Jaffe N, Strong L, Hansen M. De novo germline mutations of the p53 gene in young children with sarcomas. *Oncol Rep* 1997;4:679–83.
  - 26 Patrikidou A, Bennett J, Abou-Sleiman P, Delhanty JD, Harris M. A novel, de novo germline TP53 mutation in a rare presentation of the Li-Fraumeni syndrome in the maxilla. *Oral Oncol* 2002;38:383–90.
  - 27 Bendig I, Mohr N, Kramer F, Weber BH. Identification of novel TP53 mutations in familial and sporadic cancer cases of German and Swiss origin. *Cancer Genet Cytogenet* 2004;154:22–6.
  - 28 Nogales FF, Musto ML, Sáez AI, Robledo M, Palacios J, Aneiros J. Multifocal intrafollicular granulosa cell tumor of the ovary associated with an unusual germline p53 mutation. *Mod Pathol* 2004;17:868–73.
  - 29 Salmon A, Amikam D, Sodha N, Davidson S, Basel-Vanagaita L, Eeles RA, Abeliovich D, Peretz T. Rapid development of Post-radiotherapy sarcoma and breast cancer in a patient with a novel germline 'de-novo' TP53 mutation. *Clin Oncol* 2007;19:490–3.
  - 30 Russell-Swetek A, West AN, Mintern JE, Jenkins J, Rodriguez-Galindo C, Ribeiro R, Zambetti GP. Identification of a novel TP53 germline mutation E285V in a rare case of paediatric adrenocortical carcinoma and choroid plexus carcinoma. *J Med Genet* 2008;45:603–6.
  - 31 Gonzalez KD, Buzin CH, Noltner KA, Gu D, Li W, Malkin D, Sommer SS. High frequency of de novo mutations in Li-Fraumeni syndrome. *J Med Genet* 2009;46:689–93.
  - 32 Pinto C, Veiga I, Pinheiro M, Peixoto A, Pinto A, Lopes JM, Reis RM, Oliveira C, Baptista M, Roque L, Regateiro F, Cirnes L, Hofstra RM, Seruca R, Castedo S, Teixeira MR. TP53 germline mutations in Portugal and genetic modifiers of age at cancer onset. *Fam Cancer* 2009;8:383–90.
  - 33 Schniederjan MJ, Shehata B, Brat DJ, Esiashvili N, Janss AJ. De novo germline TP53 mutation presenting with synchronous malignancies of the central nervous system. *Pediatr Blood Cancer* 2009;53:1352–4.
  - 34 Schwarzbraun T, Obenauf AC, Langmann A, Gruber-Sedlmayr U, Wagner K, Speicher MR, Kroisel PM. Predictive diagnosis of the cancer prone Li-Fraumeni syndrome by accident: new challenges through whole genome array testing. *J Med Genet* 2009;46:341–4.
  - 35 Choong SS, Latiff ZA, Mohamed M, Lim LL, Chen KS, Vengidasan L, Razali H, Abdul Rahman EJ, Ariffin H. Malaysian Society of Paediatric Haematology-Oncology. Childhood adrenocortical carcinoma as a sentinel cancer for detecting families with germline TP53 mutations. *Clin Genet* 2012;82:564–8.
  - 36 Jhaveri AP, Bale A, Lovick N, Zuckerman K, Deshpande H, Rath K, Schwartz P, Hofstatter EW. The benefit and burden of cancer screening in Li-Fraumeni syndrome: a case report. *Yale J Biol Med* 2015;88:181–5.
  - 37 Schlegelberger B, Kreipe H, Lehmann U, Steinemann D, Ripberger T, Göhring G, Thomay K, Rump A, Di Donato N, Suttorp M. A child with Li-Fraumeni syndrome: Modes to inactivate the second allele of TP53 in three different malignancies. *Pediatr Blood Cancer* 2015;62:1481–4.
  - 38 Austin F, Oyarbide U, Massey G, Grimes M, Corey SJ. Synonymous mutation in TP53 results in a cryptic splice site affecting its DNA-binding site in an adolescent with two primary sarcomas. *Pediatr Blood Cancer* 2017;64:e26584.
  - 39 O'Shea R, Clarke R, Berkley E, Giffney C, Farrell M, O'Donovan E, Gallagher DJ. Next generation sequencing is informing phenotype: a TP53 example. *Fam Cancer*. In Press. 2017.
  - 40 Behjati S, Maschietto M, Williams RD, Side L, Hubank M, West R, Pearson K, Sebire N, Tarpey P, Futreal A, Brooks T, Stratton MR, Anderson J. A pathogenic mosaic TP53 mutation in two germ layers detected by next generation sequencing. *PLoS One* 2014;9:e96531.
  - 41 Prochazkova K, Pavlikova K, Minarik M, Sumerauer D, Kodet R, Sedlacek Z. Somatic TP53 mutation mosaicism in a patient with Li-Fraumeni syndrome. *Am J Med Genet A* 2009;149A:206–11.
  - 42 Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR, Vadodaria B, Chen X, McGee RB, Hines-Dowell S, Nuccio R, Quinn E, Shurtleff SA, Rusch M, Patel A, Becksfors JB, Wang S, Weaver MS, Ding L, Mardis ER, Wilson RK, Gajjar A, Ellison DW, Pappo AS, Pui CH, Nichols KE, Downing JR. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med* 2015;373:2336–46.
  - 43 Swisher EM, Harrell MI, Norquist BM, Walsh T, Brady M, Lee M, Hershberg R, Kalli KR, Lankes H, Konnick EQ, Pritchard CC, Monk BJ, Chan JK, Burger R, Kaufmann SH, Birrer MJ. Somatic mosaic mutations in PPM1D and TP53 in the blood of women with ovarian carcinoma. *JAMA Oncol* 2016;2:370–2.
  - 44 Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* 2017;14:531–48.
  - 45 Spier I, Drichel D, Kerick M, Kirfel J, Horpaopan S, Laner A, Holzapfel S, Peters S, Adam R, Zhao B, Becker T, Lifton RP, Perner S, Hoffmann P, Kristiansen G, Timmermann B, Nöthen MM, Holinski-Feder E, Schweiger MR, Aretz S. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. *J Med Genet* 2016;53:172–9.
  - 46 Kratz CP, Achatz MI, Brugières L, Frébourg T, Garber JE, Greer MC, Hansford JR, Janeway KA, Kohlmann WK, McGee R, Mullighan CG, Onel K, Pajtler KW, Pfister SM, Savage SA, Schiffrman JD, Schneider KA, Strong LC, Evans DGR, Wasserman JD, Villani A, Malkin D. Cancer screening recommendations for individuals with li-fraumeni syndrome. *Clin Cancer Res* 2017;23:e38–45.



## Contribution of de novo and mosaic *TP53* mutations to Li-Fraumeni syndrome

Mariette Renaux-Petel, Françoise Charbonnier, Jean-Christophe Théry, Pierre Fermey, Gwendoline Lienard, Jacqueline Bou, Sophie Coutant, Myriam Vezain, Edwige Kasper, Steeve Fourneaux, Sandrine Manase, Maud Blanluet, Bruno Leheup, Ludovic Mansuy, Jacqueline Champigneulle, Céline Chappé, Michel Longy, Nicolas Sévenet, Brigitte Bressac-de Paillerets, Léa Guerrini-Rousseau, Laurence Brugières, Olivier Caron, Jean-Christophe Sabourin, Isabelle Tournier, Stéphanie Baert-Desurmont, Thierry Frébourg and Gaëlle Bougeard

*J Med Genet* published online October 25, 2017

---

Updated information and services can be found at:

<http://jmg.bmj.com/content/early/2017/10/25/jmedgenet-2017-104976>

---

*These include:*

### References

This article cites 44 articles, 15 of which you can access for free at: <http://jmg.bmj.com/content/early/2017/10/25/jmedgenet-2017-104976#BIBL>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>