



**HAL**  
open science

# Apport du séquençage d'exome pour le diagnostic et le traitement des maladies complexes

Arthur Jacob

► **To cite this version:**

Arthur Jacob. Apport du séquençage d'exome pour le diagnostic et le traitement des maladies complexes. Médecine humaine et pathologie. Université de Lille, 2020. Français. NNT : 2020LILUS011 . tel-03267865

**HAL Id: tel-03267865**

**<https://theses.hal.science/tel-03267865>**

Submitted on 22 Jun 2021

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

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

**UNIVERSITÉ DU DROIT ET DE LA SANTÉ – LILLE**  
**ÉCOLE DOCTORALE BIOLOGIE – SANTÉ**

Année universitaire 2019-2020

**Thèse Soutenue en Vue de l'Obtention du Titre de Docteur de l'Université de Lille**

Discipline : Génétique, Génomique, Bio-informatique  
Spécialité : Génétique humaine

**Apport du séquençage d'exome pour le diagnostic et le  
traitement des maladies complexes**

Par **Arthur Jacob**

Sous la direction des **Professeurs Philippe FROGUEL et Arash RAFII**

Soutenue le 22 septembre 2020 devant le jury composé de :

|                                    |                       |
|------------------------------------|-----------------------|
| <b>Professeur David GENEVIEVE</b>  | Rapporteur            |
| <b>Professeur Jacques BECKMANN</b> | Rapporteur            |
| <b>Docteur Jennifer PASQUIER</b>   | Examineur             |
| <b>Professeur Arash RAFII</b>      | Co-directeur de thèse |
| <b>Professeur Philippe FROGUEL</b> | Directeur de thèse    |

(Epi)génomique Fonctionnelle et Physiologie Moléculaire Du Diabète et Maladies Associées  
UMR1283 (CNRS / Université de Lille / Institut Pasteur de Lille)

## Remerciements

Je tiens en premier lieu à remercier mes directeurs de thèse, les Professeurs Philippe Froguel et Arash Rafii de m'avoir donné l'opportunité de travailler sur ce thème de recherche et d'avoir supervisé celui-ci pendant trois ans.

Je tiens à remercier les membres de mon jury les Professeurs David Genevieve, Jacques Beckmann et Jennifer Pasquier pour avoir accepté d'évaluer mon travail de thèse.

Merci au Pr Pascal Pujol (Université de Montpellier) qui m'a été d'une grande aide pour le recrutement des familles atteintes de cancer du sein.

J'aimerais également remercier toute l'équipe du laboratoire avec qui j'ai beaucoup appris et ai eu beaucoup de plaisir à travailler. En particulier, je remercie Dr Amélie Bonfond, Dr Mehdi Derhourhi, Alaa Badredine, Emmanuel Vaillant, Emmanuelle Durand, Mathilde Boissel, Frédéric Allegaert et Nicolas Larcher qui ont participé à l'accomplissement de ce projet. Sans leur participation et leur contribution, ce projet n'aurait pas pu être mené avec succès.

Enfin, je remercie mes parents, mes grands-parents, mes soeurs, et ma petite amie Estelle pour m'avoir fourni un soutien sans faille tout au long de mes années d'études et de la rédaction de cette thèse. Je vous remercie.

## Résumé

Depuis l'achèvement du projet du génome humain en 2001, le domaine de la génomique a progressé de façon exponentielle, en grande partie grâce à l'introduction du séquençage de nouvelle génération (NGS), une technique de séquençage qui a révolutionné les méthodes d'investigation des maladies génétiques. L'accessibilité croissante de ces technologies permet l'essor de la médecine de précision, basée sur une prise en charge spécifique de chaque patient en fonction de son profil génétique. Le séquençage peut être utilisé pour le diagnostic de maladies, la recherche de prédispositions génétiques, ou encore pour le choix thérapeutique en particulier en oncologie. Le séquençage de l'exome (WES), en particulier, offre une méthode efficace pour étudier les maladies, car les régions exoniques représentent 2% du génome entier, mais peuvent contenir jusqu'à 85% de variants fonctionnels responsables des maladies. Or, l'analyse génétique des patients en milieu clinique est principalement effectuée par le séquençage ciblé de panels de quelques gènes choisis en fonction du contexte clinique. Le travail effectué dans le cadre de cette thèse CIFRE en partenariat avec la société Prenostics a été de développer des analyses WES pertinentes permettant de caractériser le profil génétique de patients atteints de maladies génétiques et de cancers. L'objectif a été d'évaluer leur apport dans le diagnostic et l'établissement de stratégies thérapeutiques personnalisées dans trois contextes distincts de pratique clinique, en le comparant à l'approche conventionnelle de séquençage de panels.

- Dans le domaine des maladies génétiques pédiatriques, le diagnostic moléculaire par les méthodes conventionnelles n'atteint que 25%, laissant la majorité des familles sans conseil génétique pertinent. Notre analyse WES d'une cohorte de 26 enfants atteints de maladies génétiques non diagnostiqués par les analyses génétiques conventionnelles, notre approche WES a permis d'établir un diagnostic positif dans 35% des cas.

- Environ 5 à 10% des cancers du sein sont héréditaires, mais plus de la moitié d'entre eux ne sont pas élucidés par le séquençage de panel classique incluant les gènes à risque (*BRCA1*, *BRCA2*, *PALB2*, etc) et ne sont pas causés par des gènes de prédisposition connus. En analysant l'exome de quatre familles, nous avons tenté d'identifier les gènes en cause de cas familiaux ne présentant pas d'altération des gènes connus *BRCA1/BRCA2* (familles BRCAx). Après filtration des variants à

risque transmis parmi les membres atteints, nous avons identifié les gènes candidats *HIST1H1C*, *TYRO3*, *TPH1*, *SLC12A3* et *CCNF* comme de possibles gènes de prédisposition au cancer du sein. Cependant, sans études fonctionnelles poussées permettant de valider leur implication, le WES ne semble pas apporter de bénéfice pour la prise en charge de ces familles.

- Enfin, dans le domaine de la cancérologie, la personnalisation du traitement est au centre des enjeux actuels. Notre étude d'une cohorte de 35 tumeurs solides réfractaires a démontré la faisabilité et l'efficacité du WES pour la caractérisation du profil génétique de tumeurs solides et la prise de décision en oncologie. La détermination des pattern mutationnels tumoraux, de la charge tumorale et des signatures mutationnelles à partir des données de WES a permis d'émettre des suggestions thérapeutiques pour la moitié d'entre eux et a contribué à la modification du traitement d'au moins huit patients sur 35.

Cette étude décrit les différentes applications, les limites et les avantages du WES comme outil d'investigation moléculaire des maladies humaines. En démontrant le bénéfice d'utilisation du WES en clinique, nos résultats contribuent aux efforts en cours pour son intégration dans le parcours de soin et le développement de la médecine de précision.

## Abstract

Since the completion of the Human Genome Project in 2001, the field of genomics has grown exponentially, in large part due to the introduction of Next Generation Sequencing (NGS). This technique has revolutionized the investigation methods of genetic diseases, allowing high-throughput genome-wide sequencing to establish the genetic basis of diseases. The increasing accessibility of these technologies allows the development of precision medicine, based on the specific care of each patient according to his genetic profile. Sequencing can be used for the diagnosis of diseases, the search for genetic predispositions to a disease, or for the therapeutic choice, in particular in oncology. Exome sequencing (WES), in particular, offers an effective method for studying diseases, since exonic regions represent 2% of the entire genome, but can contain up to 85% of functional variants responsible for diseases. However, the genetic analysis of patients in a clinical setting is mainly

carried out by the targeted sequencing of panels of a few genes chosen according to the clinical context. The work carried out during this CIFRE thesis in partnership with the company Prenostics was to develop relevant WES analyzes to characterize the genetic profile of patients with rare genetic diseases and cancers. The objective was to assess their contribution to the diagnosis and establishment of personalized treatment strategies in three distinct contexts of clinical practice, by comparing it to the conventional approach of panel sequencing.

- In the field of pediatric genetic diseases, molecular diagnosis by conventional methods only reaches 25%, leaving the majority of families without precise genetic counseling. Our WES analysis of a cohort of 26 children with genetic diseases not diagnosed by conventional genetic analyzes, our WES approach resulted in a positive diagnosis in 35% of cases.

- About 5 to 10% of breast cancers are hereditary, but more than half of them are not elucidated by the genes at risk of breast cancer (*BRCA1*, *BRCA2*, *PALB2*, etc.) that are included in conventional panel sequencing. By analyzing the exome of four families, we attempted to identify the genes involved in familial cases non-mutated for the known *BRCA1* / *BRCA2* genes (*BRCAx* families). After filtering the risk variants transmitted among affected limbs, we identified the candidate genes *HIST1H1C*, *TYRO3*, *TPH1*, *SLC12A3* and *CCNF* as possible genes of predisposition to breast cancer. However, without in-depth functional studies to validate their involvement, WES does not seem to provide any benefit for patient management.

- Finally, in the field of oncology, the personalization of treatment is at the center of current issues. Our study of a cohort of 35 refractory solid tumors aimed at demonstrating the feasibility and efficacy of WES for characterizing the genetic profile of solid tumors and for decision-making in oncology. We were able to make treatment suggestions for half of them and helped modify the treatment of at least eight out of 35 patients.

This study describes the different applications, limits and advantages of WES as a molecular investigation tool for human diseases. By demonstrating the benefit of using WES in the clinic, our results contribute to the ongoing efforts to integrate it into the care pathway and the development of precision medicine.

# Table des matières

|  |    |
|--|----|
| Remerciements .....  | 2  |
| Résumé .....   | 3  |
| Abstract .....   | 4  |
| Table des matières .....   | 6  |
| Liste des Figures .....  | 10 |
| Liste des Tableaux .....   | 12 |
| Liste des abréviations .....   | 13 |
| Avant-propos .....   | 15 |
| Introduction .....   | 17 |
| Cadre de la thèse .....  | 17 |
| Contexte scientifique de la thèse .....  | 17 |
| Objectifs de la thèse .....  | 19 |
| Déroulement de la thèse, contribution et aboutissements .....  | 21 |
| Chapitre 1 : Synthèse bibliographique : le séquençage du génome humain, l'essor du NGS et ses applications en médecine ..... | 23 |
| I - Le séquençage du génome humain : du séquençage à l'interprétation des données .....                                      | 23 |
| 1. Les technologies de séquençage .....  | 23 |
| 1.1 Les différents types de séquençage .....   | 23 |
| 1.2 Technologie Illumina .....   | 25 |
| 1.3 Notion de profondeur de séquençage .....   | 27 |
| 2. Alignement sur le génome de référence .....   | 28 |
| 3. Les différents types de variants .....  | 28 |
| 3.1 Les variants mononucléotidiques (SNVs) et variants à petites insertions / délétions (indels) .....                       | 29 |
| 3.2 Les variants structuraux .....   | 30 |
| 4. Interprétation fonctionnelle des variants .....   | 32 |
| 4.1 Fréquence allélique des variants (MAF) .....   | 33 |
| 4.2 Prédictions informatiques .....  | 35 |
| 4.3 Type de variant et données fonctionnelles .....  | 35 |
| 4.4 Ségrégation des variants .....   | 36 |
| 4.5 Prédictions fonctionnelles .....   | 36 |
| 1.5. Validation fonctionnelle de l'impact des variants .....   | 37 |
| 5.1 Au niveau de l'ARN .....   | 37 |
| 5.2 Au niveau protéique .....  | 38 |

|  |    |
|--|----|
| 5.3 Stratégies de corrélation génotype-phénotype .....   | 38 |
| II - Les applications du NGS en clinique .....   | 39 |
| 1. Stratégie de NGS en clinique : WGS versus WES versus séquençage ciblé.....                                    | 40 |
| 2. Stratégies de NGS d'identification de nouveaux gènes/loci associés aux maladies .....                         | 41 |
| 2.1 Maladies mendéliennes monogéniques .....   | 41 |
| 2.2 Maladies complexes et stratégies « GWAS » .....  | 43 |
| 3. Application du WES pour le diagnostic de maladies rares mendéliennes .....                                    | 44 |
| 3.1 Maladies autosomiques dominantes rares : le rôle de premier plan des variants de novo .....                  | 45 |
| 3.2 Maladies récessives rares: le rôle de premier plan des variants hétérozygotes composites et homozygotes..... | 46 |
| 3.3 Quelques exemples concrets.....  | 46 |
| 3.3.1 Troubles neurodéveloppementaux .....   | 48 |
| 3.3.2 Troubles neuromusculaires.....   | 48 |
| 3.3.3 Troubles moteurs.....  | 49 |
| 4. Application du WES pour le diagnostic et le traitement des cancers .....                                      | 50 |
| 4.1 Variants de la lignée germinale .....  | 52 |
| 4.1.1. Les variants de prédisposition .....  | 52 |
| 4.1.2 Les variants pharmacogénétiques.....   | 54 |
| 4.2 Les variants de la lignée somatique .....  | 55 |
| 4.2.1 Les variants pronostiques .....  | 55 |
| 4.2.2 Les variants prédictifs.....   | 56 |
| 4.2.3 Les variants pharmacothérapeutiques: identifier les cibles thérapeutiques .....                            | 57 |
| 4.3 Essais cliniques prospectifs et randomisés sur différents types de tumeurs..                                 | 61 |
| 5. Limitations du NGS.....   | 62 |
| 5.1 Sensibilité analytique .....   | 63 |
| 5.2 Les régions difficiles à séquencer .....   | 63 |
| 5.3 Limitations des bases de données et des connaissances.....   | 64 |
| 5.4 Variations structurales et variations du nombre de copies .....  | 64 |
| Chapitre 2 : Apports du séquençage d'exome pour le diagnostic de maladies rares .....                            | 66 |
| 1. Introduction.....   | 66 |
| 2. Matériel et méthodes.....   | 67 |
| 2.1. Échantillon et cohortes utilisées .....   | 67 |
| 2.2 Extraction d'ADN .....   | 68 |
| 2.3. Séquençage de l'exome .....   | 68 |
| 2.3.1 Librairie.....   | 68 |



|   |     |
|---|-----|
| 2.3.2. Enrichissement .....   | 69  |
| 2.3.3. Séquençage .....   | 71  |
| 2.4. Analyse bioinformatique .....  | 71  |
| 2.5. Analyse et interprétation des variants .....   | 71  |
| 2.6 Isolement d'ARN et RT-PCR .....   | 72  |
| 2.7. Séquençage Sanger.....   | 73  |
| 3. Résultats.....   | 74  |
| 3.1. Conception du test et séquençage d'une cohorte de patients .....   | 74  |
| 3.2 Troubles neuro-développementaux.....  | 75  |
| 3.2.1 Analyse fonctionnelle de l'impact du variant synonyme de <i>EFTUD2</i> identifié chez l'individu A.....                                     | 78  |
| 3.2.1.1 Prédiction in silico .....  | 80  |
| 3.2.1.2 Validation fonctionnelle .....  | 81  |
| 3.3. Autisme .....  | 83  |
| 3.4. Malformations congénitales.....  | 84  |
| 3.5. Troubles du mouvement .....  | 87  |
| 3.5.1 Le cas du patient V : l'identification d'une altération bi-allélique du gène <i>ASCC3</i> révèle un nouveau syndrome neuro-musculaire ..... | 87  |
| 3.6. Vue d'ensemble des résultats de la cohorte .....   | 92  |
| 3.7 Intérêt du séquençage en trio .....   | 94  |
| 3.8. Implication pour l'entreprise et mise en place d'un pipeline.....  | 95  |
| 4. Discussion.....  | 97  |
| 5. Conclusions .....  | 101 |
| Chapitre 3 : Intérêt du WES appliqué aux familles BRCAx et caractérisation de nouveaux facteurs de prédisposition au CS héréditaire.....          | 102 |
| 1. Introduction.....  | 102 |
| 2. Matériel et Méthodes .....   | 103 |
| 2.1. Patients.....  | 103 |
| 2.2. Extraction, préparation des librairies et séquençage d'exome.....  | 104 |
| 2.3. Analyse des données .....  | 104 |
| 2.4. Séquençage Sanger.....   | 104 |
| 2.5. Prédiction <i>in-silico</i> .....  | 105 |
| 3. Résultats.....   | 105 |
| 3.1. Séquençage d'exome des familles BRCAx .....  | 105 |
| 3.2. Analyse des données de séquençage et stratégie de filtrage des variants.....   | 106 |
| 3.3. Analyse des gènes connus de prédisposition au cancer du sein héréditaire....   | 108 |
| 3.4. Identification de nouveaux gènes candidats.....  | 108 |

|   |     |
|---|-----|
| 3.4.1 Famille I.....  | 108 |
| 3.4.2 Famille II.....   | 111 |
| 3.4.3 Famille III.....  | 113 |
| 3.4.4 Famille IV.....   | 115 |
| 3.5. Implications pour l'entreprise .....   | 118 |
| 4. Discussion.....  | 118 |
| 5. Conclusion.....  | 125 |
| Chapitre 4 : Apport du séquençage de l'exome pour la personnalisation des traitements en oncologie .....  | 126 |
| 1. Introduction.....  | 126 |
| 2. Matériel et Méthodes .....   | 127 |
| 2.1 Cohorte cancer.....   | 127 |
| 2.2. Analyse des variants somatiques.....   | 127 |
| 2.3. Analyse des variants constitutionnels.....   | 128 |
| 2.5. Analyse des signatures mutationnelles tumorales .....  | 128 |
| 3. Résultats.....   | 129 |
| 3.1. Mise en œuvre du test génétique.....   | 129 |
| 3.2 Profilage moléculaire des tumeurs par séquençage ciblé de panels et étude des CNV.....  | 131 |
| 3.2 Analyse de pattern mutationnels somatiques et constitutionnels.....   | 131 |
| 3.3. Le calcul de la charge mutationnelle permet de prédire la sensibilité aux immunothérapies.....   | 135 |
| 3.4. Identification des signatures mutationnelles et du profil BRCAness.....  | 136 |
| 3.5 Impact sur les modifications de traitement et devenir des patients.....   | 138 |
| 3.6. Mise en place du pipeline d'analyse en entreprise et construction d'un rapport .....   | 141 |
| 4. Discussion.....  | 144 |
| 5. Conclusion.....  | 148 |
| Conclusions générales et perspectives.....  | 149 |
| Références.....   | 153 |
| Figures supplémentaires .....   | 170 |
| Annexe 1: Exemple de rapport d'analyse des résultats de WES en trio pour le diagnostic de maladies génétiques rares .....                         | 171 |
| Annexe 2 : Exemple de rapport d'analyse des résultats de WES du génome somatique et constitutionnel pour le profilage moléculaire de tumeurs..... | 176 |

## Liste des Figures

|   |     |
|---|-----|
| Figure 1: Dogme central de la biologie moléculaire.....   | 16  |
| Figure 2: Les étapes du processus de séquençage ADN par NGS.....  | 23  |
| Figure 3: Principe du séquençage Illumina.....  | 26  |
| Figure 4: Alignement des lectures <i>paired-end</i> .....   | 27  |
| Figure 5: Types d'altérations génomiques détectées par séquençage haut débit.....   | 30  |
| Figure 6: Évolution du nombre d'associations gène-phénotype recensées dans la base de données OMIM au cours des 30 dernière années..... | 43  |
| Figure 7: Les différents types de biomarqueurs génétiques.....  | 51  |
| Figure 8: Les principaux gènes de prédisposition au cancer du sein, regroupés dans un diagramme risque / fréquence.....                 | 54  |
| Figure 9: Schéma des étapes de la préparation d'une librairie d'ADN pour séquençage de nouvelle génération.....                         | 69  |
| Figure 10: Schéma des différentes étapes du procédé de capture.....   | 70  |
| Figure 11 : Série de filtres appliqués à l'aide du logiciel IVA (Qiagen).....   | 75  |
| Figure 12: Identification du variant de novo chez l'individu A.....   | 79  |
| Figure 13: Prévisions in silico de l'impact du variant EFTUD2 c.702G> T sur l'épissage.....   | 80  |
| Figure 14: Le variant c.702G> T induit un saut de l'exon 9.....   | 82  |
| Figure 15: Séquence d'ARNm de l'allèle WT par rapport à l'allèle mutant.....  | 81  |
| Figure 16: Identification de deux nouveaux variants pathogénique en trans dans le gène ASCC3.....                                       | 88  |
| Figure 17: Description des six cas atteints du nouveau syndrome lié à ASCC3.....  | 89  |
| Figure 18: Localisation et impact des nouveaux variants pathogéniques identifiés dans cette étude.....                                  | 93  |
| Figure 19: Diagramme montrant le taux de diagnostic du WES en trio.....   | 94  |
| Figure 20: Exemple de rapport de séquençage d'exome en trio pour le diagnostic génétique d'une maladie rare non diagnostiquée.....      | 97  |
| Figure 21: Étapes de filtration des variants à haut-risque.....   | 107 |
| Figure 22: Arbre généalogique de la famille I.....  | 109 |
| Figure 23: Arbre généalogique de la famille II.....   | 112 |
| Figure 24: Arbre généalogique de la famille III.....  | 114 |
| Figure 25: Arbre généalogique de la famille IV.....   | 115 |
| Figure 26: Séquençage Sanger de la région d'intérêt du gène <i>CCNF</i> .....   | 117 |

|   |     |
|---|-----|
| Figure 27: Dynamique de la régulation de ATM par l'histone H1.2.....  | 121 |
| Figure 28: Enrichissement fonctionnel des protéines formant le réseau d'interaction de l'histone H1.2.....  | 121 |
| Figure 29: Les voies de signalisation <i>TYRO3</i> .....  | 123 |
| Figure 30: Schéma des étapes du test de profilage moléculaire des tumeurs solides.....  | 130 |
| Figure 31: Histogramme montrant les variants de la lignée somatique et germinale par séquençage d'exome de l'ADN tumoral de 35 tumeurs.....           | 132 |
| Figure 32: Analyse des variants actionnables et potentiellement ciblables, et identification des options thérapeutiques envisageables.....            | 134 |
| Figure 33: Charge mutationnelle des 35 cas en nombre de mutations par mégabase (mut/MB).....  | 135 |
| Figure 34: Les signatures mutationnelles peuvent être associées à une sensibilité ou une résistance à certains traitements anticancéreux.....         | 137 |
| Figure 35: Contenu du rapport rédigé à l'issue de l'analyse moléculaire tumorale, incluant une analyse des altérations actionnables de la tumeur..... | 143 |

## Liste des Tableaux

|   |     |
|---|-----|
| Tableau 1: Directives ACMG pour la classification des variants.....                                   | 33  |
| Tableau 2: Thérapies ciblées approuvées par la FDA.....   | 59  |
| Tableau 3: Règle de combinaison des critères pour la classification des variants.....                 | 72  |
| Tableau 4: Liste des primers utilisés pour la validation du variant EFTUD2 par séquençage Sanger..... | 73  |
| Tableau 5: Description des cas de cas atteints de troubles neuro-développementaux.....                | 77  |
| Tableau 6: Description des patients atteints d'autisme séquencés dans notre étude.....                | 83  |
| Tableau 7: Description des cas de patients atteints de malformations congénitales.....                | 86  |
| Tableau 8: Description des cas de patients atteints de troubles du mouvement.....                     | 91  |
| Tableau 9: Amorces utilisées pour l'amplification des régions comportant les variants candidats.....  | 105 |
| Tableau 10: SNVs délétères identifiés chez les individus atteints de la famille I.....                | 111 |
| Tableau 11: SNVs délétères identifiés chez les individus atteints de la famille II.....               | 113 |
| Tableau 12: SNVs délétères identifiés chez les individus atteints de la famille IV.....               | 116 |
| Tableau 13: Résultats du profilage génétique des tumeurs des 35 individus de notre cohorte.....       | 141 |

## Liste des abréviations

ACMG : American College of Medical Genetics  
ADN : Acide DésoxyriboNucléique  
ADNc : ADN complémentaire  
ARN : Acide ribonucléique  
ARNm : ARN messenger  
BAM : Binary Alignment Map  
BRCAX : Cancer du sein familial non muté pour les gènes BRCA1/BRCA2  
BWA : BurrowWheeler Aligner  
CADD : Combined Annotation-Dependent Depletion  
CAP : College of American Pathologist  
CGH Array : Comparative Genomic Hybridization Array  
CLIA : Clinical Laboratory Improvement Amendments  
CNVs : variations du nombre de copies  
CS : Cancer du Sein  
CSE : Amplificateur d'épissage exonique  
ER : Récepteur à l'Estrogen  
ESP : Exome Sequencing Project  
ExaC : Exome Aggregation Consortium  
FDA : Food and Drug Administration  
FGFR : facteur de croissance des fibroblastes  
GATK : Genome Analysis ToolKit  
GnomAD : Genome Aggregation Database  
GWAS : Genome Wide Association Study  
HSF : Human Splicing Finder  
FISH : Fluorescent In Situ Hybridization  
FFPE : fixé au formol et inclu dans la paraffine (Formalin-fixed Paraffin-Embedded)  
ICGC : Consortium International sur le Genome du Cancer  
IGV : Integrative Genomics Viewer  
IMPACT : Initiative for Molecular Profiling and Advanced Cancer Therapy  
IVA : Ingenuity Pathway Analysis  
MAF : Minor Allele Frequency  
MATCH : Molecular Analysis for Therapy Choice  
MB : MégaBase  
MES : MaxEntScore  
MFDM : Dysostose Mandibulo-Faciale avec Microcéphalie  
NER : Réparation par excision de nucléotide  
NIH : National Institute of Health  
PCR : chaîne de réaction par polymérase  
PD-1 : anti-programmed cell death protein 1  
PolyPhen : Polymorphism Phenotyping  
PR : Récepteur à la Progestérone  
PROVEAN : Protein Variation Effect Analyzer  
PTV : Protein Truncated Variant  
R&D : Recherche et Développement  
RH : Recombinaison Homologue  
SAM : Sequence Alignment Map

SFMPP : Société française de médecine prédictive et personnalisée  
SIFT : Sorting Intolerant From Tolerant  
SNV : Single Nucleotide Variant  
SSF : SpliceSiteFinder-like  
SV : Structural Variant  
TCGA : The Cancer Genome Atlas  
TSA : Trouble du Spectre Autistique  
UTR : Untranslated Transcribed Region  
VCF : Variant Call Format  
VUS : Variant of unknown significance  
WT : Wild-Type

## Avant-propos

L'acide désoxyribonucléique ou ADN est une macromolécule biologique présente dans toutes les cellules animales. L'ADN contient toute l'information génétique, appelée génome, permettant le développement, le fonctionnement et la reproduction des êtres vivants. Dans les cellules, elle est composée d'un double brin antiparallèle prenant la forme d'une double hélice. Chacun des brins est un polymère, dans lequel chaque monomère qui le constitue est un nucléotide. Il existe 4 nucléotides différents : adénine (A), cytosine (C), guanine (G) et thymine (T) qui se suivent dans un ordre précis formant la séquence ADN, définissant le code génétique. Dans les cellules, l'ADN est condensé en structures appelées chromosomes, au nombre de 23 chez l'Homme, qui contiennent l'ensemble de l'information génétique. Les cellules somatiques contiennent une paire de chaque chromosome (diploïdie), tandis que les cellules germinales n'en contiennent qu'un exemplaire de chaque et sont dites haploïdes.

La majeure partie de l'information génétique du génome est codée dans les gènes, dont la composition allélique est appelée génotype. Un gène est une fraction de l'ADN contenant les informations pour la production d'une protéine qui influence une caractéristique particulière de l'organisme. Chaque gène contient un cadre de lecture ouvert allant de l'extrémité 5' à l'extrémité 3' qui peut être transcrit en acide ribonucléique (ARN), ainsi que des séquences de régulation de l'expression génétique telles que les promoteurs et les amplificateurs qui en contrôlent la transcription. Les gènes ne représentent qu'environ 2% de la taille du génome et sont entourés de régions non codantes qui peuvent avoir des rôles de régulation mais dont la fonction reste en grande partie inconnue.

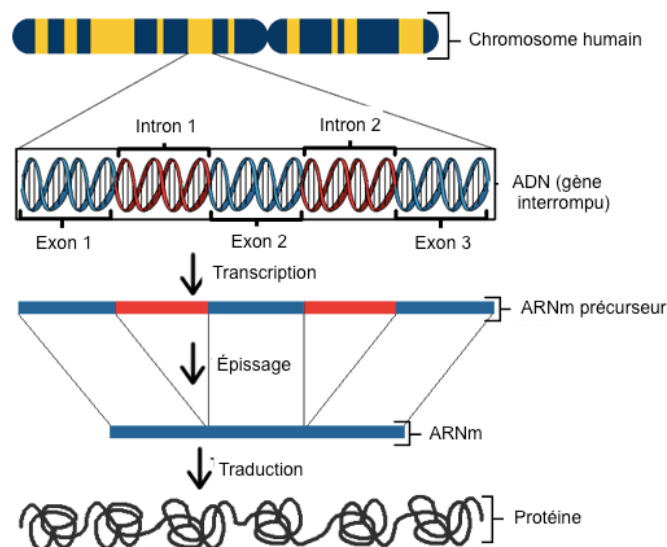
La structure générale d'un gène est la suivante: à l'extrémité 5' se trouve une région 5'UTR qui contient le promoteur et des régions de régulation de la transcription, suivi d'une succession d'exons et d'introns en alternance, jusqu'à la fin de la région codante. Celle-ci est suivie de la région 3'UTR. Les exons et les introns sont transcrits en ARN, dont la séquence remplace les thymines (T) par des uraciles (U). Seuls les exons seront traduits en protéine, à la suite d'une étape de maturation de l'ARN au cours de laquelle les régions introniques sont épissées. Les exons sont donc la partie codante des gènes, et l'ensemble des exons du génome est appelé exome. Lors de la traduction de l'ARN en protéines, les nucléotides sont lus trois par



trois si bien qu'une succession de 3 nucléotides, appelée « codon » code pour l'un des 20 acides aminés dont la séquence constituera à terme une protéine, ou d'un codon stop marquant la fin de la séquence codante (Figure 1).

Chaque individu a un génome unique, comportant de nombreuses variations ou polymorphismes dans son génome par rapport au génome de référence (génome assemblé dans une base de donnée par la communauté scientifique, et représentatif du génome d'une espèce) appelées variants. Ces derniers peuvent être de différents types : les variants d'un seul nucléotide (SNV), les délétions et les insertions. Lorsque les variants sont rares (<1%) dans la population, on peut les appeler mutations.

Ils sont en grande majorité bénins et déterminent les traits phénotypiques propres à chaque individu (couleur des yeux) mais peuvent parfois modifier la séquence de façon délétère pour la fonction de la protéine et résulter en un phénotype pathologique. De ce fait la séquence ADN peut être extrêmement informative sur l'origine d'un phénotype, apporter des connaissances indispensables pour comprendre les maladies au niveau moléculaire et pourrait donc avoir des applications médicales. Dans cette thèse, nous chercherons à montrer comment le séquençage du génome peut améliorer la prise en charge de patients atteints de maladies génétiques.



**Figure 1:** Dogme central de la biologie moléculaire : l'ADN génomique contient les gènes porteurs de l'information génétique, qui sont transcrit en pré-ARNm, transformés en ARNm matures via l'épissage, et traduits en protéines. Adapté de Bioinformatics and functional genomics, Jonathan Pevsner, © Wiley Blackwell.

# Introduction

## Cadre de la thèse

Cette thèse a bénéficié d'une bourse de Convention Industrielle de Formation par la Recherche (CIFRE) délivrée par l'Agence Nationale de la Recherche et de la Technologie (ANRT). Une convention a été réalisée entre l'ANRT, l'Université de Lille et la société Prenostics. Le travail de thèse a été réalisé en cotutelle au sein de l'équipe CNRS UMR8199 dirigée par Philippe Froguel (PU-PH) et dans l'entreprise Prenostics SAS fondée en 2015 par Arash Rafii (Professeur de médecine à l'Université Cornell au Qatar).

Prenostics est une jeune entreprise de biotechnologie française spécialisée dans le domaine de la médecine génomique et des tests génomiques. Désireuse dès son lancement de se développer à travers la recherche dans le but de proposer des technologies objectives et pertinentes, elle s'est rapprochée du domaine universitaire. La stratégie était double car elle permettait de bénéficier de l'expertise de la plateforme de génomique LIGAN (UMR1283) pour les prestations des séquençages et de mettre en place de la Recherche & Développement de l'entreprise. C'est dans ce cadre que la thèse a été mise en place.

## Contexte scientifique de la thèse

Depuis l'année 2001 et le séquençage du génome humain, les technologies de séquençage n'ont cessées de s'améliorer, notamment grâce à l'essor du « séquençage de nouvelle génération » (dit NGS pour *next generation sequencing*) qui a permis au fil des années une diminution spectaculaire du temps et du coût de séquençage. Cette nouvelle possibilité de détecter toute sorte de modifications génétiques à grande échelle a ouvert la voie à de nombreuses applications en médecine qui sont en train de révolutionner la prise en charge des patients.

L'accessibilité croissante du NGS a permis en une dizaine d'années de séquencer des centaines de milliers de génomes dans le monde entier. Le stockage des informations dans d'immenses bases de données a permis de grandes avancées dans la compréhension génétique des maladies humaines et a révélé la grande complexité des mécanismes moléculaires qui les sous-tendent.

Ainsi, en l'espace d'une dizaine d'années, le NGS a permis l'identification de plus de 6000 gènes causant autant de maladies génétiques « rares », rendant possible le diagnostic extrêmement précis de pathologies autrefois non déterminées. Plus le diagnostic est précis, plus les causes de la maladie sont identifiées et plus le patient pourra bénéficier d'une prise en charge adaptée.

Il en est de même en cancérologie, où le séquençage de l'ADN tumoral a révélé que toutes les tumeurs, même si elles sont issues d'un même organe, comportent un profil génétique unique. Par exemple, chaque cancer du poumon comporte des altérations moléculaires différentes, qui peuvent les différencier d'un point de vue de la sensibilité ou de la résistance à certains traitements. L'identification de variants somatiques - acquis au sein de la tumeur par les cellules cancéreuses - permet de comprendre les mécanismes biologiques à l'œuvre dans sa progression et ainsi mieux la traiter.

Le séquençage systématique des patients fait entrevoir l'avènement de la médecine « de précision » basée sur une caractérisation fine du patient dans le but d'être plus prédictive et d'optimiser l'efficacité des traitements et la gestion de la maladie (prédispositions, effets secondaires d'un traitement, etc). C'est dans l'optique de contribuer à l'émergence de cette nouvelle médecine basée sur le séquençage du génome que Prenostics a été fondée en 2015, devenant l'un des premiers acteurs privés français du domaine.

En cherchant à développer des tests destinés au diagnostic de maladies génétiques et à la recherche de prédisposition aux cancers héréditaires de profilage moléculaire de tumeurs, l'entreprise a été confrontée à choisir le type de test NGS le plus pertinent dans chaque cas, parmi plusieurs approches disponibles. Il est en effet possible actuellement de séquencer l'ADN des individus à différentes échelles. Alors que les tests de panels de gènes ciblés ne séquencent qu'une cohorte spécifique de gènes, le séquençage du génome entier identifie des anomalies dans tout le génome. Le séquençage de l'exome quant à lui séquence de manière ciblée les régions codant pour les protéines.

Actuellement, en France, les analyses génétiques sont fondées en premier lieu sur le séquençage de panels ne ciblant que les gènes d'intérêt et sur la recherche d'anomalies cytogénétiques. Malheureusement, l'identification des gènes en cause s'avère très difficile en raison de la diversité des gènes potentiellement impliqués

pour un phénotype donné et de la diversité des variants parmi les 3 milliards de bases qui constituent le génome humain. Ainsi, un grand nombre de patients demeure sans diagnostic moléculaire pourtant essentiel à l'élaboration de stratégies thérapeutiques individuelles et au conseil génétique des familles.

Ainsi, les laboratoires réfléchissent actuellement à changer leurs stratégies, passant d'une analyse ciblée basée sur quelques loci à une analyse systématique du génome entier. La complexité de l'interprétation et de la validation des variants reste le principal défi de la mise en œuvre généralisée de ces technologies. Il existe en effet des limitations freinant l'inclusion de cette technologie dans les procédures de routine. Premièrement, l'importante quantité de données produites à traiter et analyser induit des difficultés techniques. Deuxièmement, l'interprétation de la grande majorité des variants dont l'effet est inconnu est complexe. Troisièmement, le coût financier inhérent au séquençage du génome entier est important.

Une solution intermédiaire au séquençage du génome entier consiste à séquencer uniquement l'exome, qui représente toutes les parties codantes du génome, soit environ 2% de celui-ci. Il permet donc de s'affranchir des régions non codantes souvent peu contributives car leurs impacts fonctionnels sont encore difficilement prévisibles en l'état actuel des connaissances. Le coût d'un séquençage d'exome est également nettement réduit par rapport au séquençage du génome entier, passant d'environ 2000\$ à 500\$ (en 2016). Malgré les atouts du WES, il n'est en général pas utilisé en routine dans les laboratoires de génétique que ce soit pour le diagnostic de maladies rares ou en oncologie, au détriment du séquençage ciblé de panels de gènes, car ses bénéfices par rapport à ce dernier sont jugés trop faibles en considérant son coût et sa complexité.

## Objectifs de la thèse

Dans le cadre de cette thèse en co-tutelle entre les milieux académiques et industriels, nous nous sommes fixés plusieurs objectifs, scientifiques et techniques, afin de répondre aux besoins de l'entreprise.

L'objectif scientifique majeur a été de définir l'approche de NGS la plus pertinente en milieu hospitalier, afin de proposer aux patients celle qui pourrait contribuer au mieux

à leur prise en charge. En particulier, ce travail vise à évaluer les bénéfices et les limites de l'exome par rapport à une approche classique de séquençage ciblé, dans différents contextes cliniques. Pour ce faire, nous avons tenté de répondre aux questions suivantes:

1. L'exome systématique permet-il de poser un diagnostic plus souvent qu'une approche clinique de routine se basant sur des tests ciblés ou des panels dans les anomalies du développement?
2. L'analyse de l'exome systématique peut-il avoir un rôle dans les familles BRCAx ?
3. L'exome tumoral permet-il une modification des traitements en oncologie par rapport à une approche classique ?

Dans le cadre d'une thèse CIFRE cette approche permettait également de mettre en place les pipelines de produits pour l'entreprise. En effet, la réponse apportée par nos cohortes pilotes pourrait valider ou invalider la création de produits innovants.

Ce projet implique le séquençage de plusieurs cohortes tests d'individus atteints de maladies d'origine génétique non diagnostiquées, et de profilage moléculaire de tumeurs. Ainsi, le second objectif intrinsèque a été de découvrir de nouveaux variants voire de nouveaux gènes impliqués causant les pathologies des individus étudiés.

L'objectif global visé par l'entreprise a été de développer des analyses de séquençage d'exome fiables, praticables et pertinentes pour le diagnostic de maladies génétiques et le profilage moléculaire de tumeurs.

En résumé, ce travail de thèse a pour but de démontrer la pertinence et la faisabilité du séquençage d'exome dans le contexte de maladies rares et en cancérologie, pour la prévention des risques (conseil génétique), le diagnostic et le choix thérapeutique. La réalisation de ce travail correspond par ailleurs à un réel besoin de l'entreprise de développer des analyses optimisées pour améliorer la prise en charge des patients atteints de maladies génétiques et de cancers.

## Déroulement de la thèse, contribution et aboutissements

Le manuscrit est structuré en quatre grands chapitres. Le premier chapitre est une revue de la littérature détaillée reprenant la bibliographie de tous les thèmes abordés. Les trois suivants traitent chacun de l'application du NGS dans des contextes médicaux distincts et répondent aux trois grands objectifs cités plus hauts. Le manuscrit suivra donc un plan réalisé de manière thématique. Cette mise en place permettra de bien distinguer chaque partie, et de comprendre le fil conducteur de la thèse.

Le manuscrit débute donc par une synthèse bibliographique des applications et de l'état de l'art actuel de l'apport des technologies de NGS en médecine.

Le second chapitre est consacré à l'apport du WES pour le diagnostic de maladies rares. Ces travaux ont permis la publication de deux articles dans des revues à comité de lecture, pour la découverte de nouvelles étiologies moléculaires. L'un concerne la mise en évidence d'un nouveau syndrome décrit pour la première fois, en collaboration avec une équipe américaine, causé par une altération du gène *ASCC3* et l'autre traite de l'étude fonctionnelle d'un variant synonyme dans le gène *EFTUD2* chez une jeune fille atteinte de dysostose mandibulofaciale avec microcéphalie (MFDM).

Le troisième chapitre traitera de l'intérêt de l'utilisation du WES dans le contexte de famille BRCAx. Nous tenterons d'évaluer son implication et les enjeux notamment pour le conseil génétique de ces familles. Les résultats ont fait l'objet d'une communication affichée et d'un article manuscrit non soumis.

Dans un quatrième chapitre nous évaluerons les bénéfices du séquençage d'exome tumoral dans le cadre d'un profilage moléculaire des tumeurs réfractaires, après les échecs du séquençage de panels pour la personnalisation des traitements.

Le travail de traitement des échantillons et de séquençage a été réalisé sur la plateforme de séquençage LIGAN au sein de l'UMR1283 et le développement des analyses et interprétations des données a été effectué au sein des unités de bioinformatiques de l'UMR1283 et de la société Prenostics.

Durant ces trois années, j'ai effectué la majeure partie des étapes des analyses génétiques et de leur conception, dans le but de répondre aux besoins de

l'entreprise. Mon travail a consisté en l'organisation de la collecte des échantillons à l'hôpital ou directement en contactant les patients, la réception des échantillons et l'extraction d'ADN des tissus. J'ai par la suite mis l'accent sur la mise en place de pipeline d'analyse optimal permettant le traitement, l'analyse et l'interprétation des données brutes de séquençage. J'ai également créé et rédigé les rapports rendus à destination des patients.

# Chapitre 1 : Synthèse bibliographique : le séquençage du génome humain, l'essor du NGS et ses applications en médecine

Dans cette revue de la littérature, nous développerons l'état actuel des connaissances en matière de séquençage, d'analyse des données, d'interprétation fonctionnelle des variants génétiques et de leurs applications médicales actuelles et futures.

## I - Le séquençage du génome humain : du séquençage à l'interprétation des données

Nous nous concentrerons sur le NGS, dont le processus se décompose en plusieurs étapes, en partant de l'échantillon biologique initial jusqu'à l'interprétation des variants (Figure 2), que nous décrivons dans ce premier chapitre.



**Figure 2** : Les étapes du processus de séquençage ADN par NGS

### 1. Les technologies de séquençage

Dans toute étude NGS, le premier objectif est de créer une carte de haute qualité des variations du génome ou de l'exome par rapport au génome de référence. Cette étape cruciale jette les bases de toutes les analyses en aval destinées à l'interprétation du génome et à la découverte génétique. Les variants qui ne sont pas identifiés avec précision ne seront pas pris en compte dans les analyses post-séquençage. Les méthodes d'analyse utilisées pour annoter les variations du génome dépendent fortement de la technologie de séquençage utilisée et de la profondeur de couverture obtenue.

#### 1.1 Les différents types de séquençage



En 1975, Sanger est le premier à développer, par des approches fondamentales, une méthode de détection de la séquence de bases de l'ADN [1]. En raison de sa grande précision, de sa simplicité et de sa grande disponibilité, la méthode Sanger a été la technologie dominante du séquençage de l'ADN pendant trois décennies. Le domaine du séquençage de l'ADN a ensuite rapidement évolué en termes de capacités et d'applications. Des avancées dans plusieurs domaines ont été combinées pour progressivement passer au séquençage à l'échelle du génome. Ainsi en 1990, fort du développement de la réaction en chaîne par polymérase (PCR) et du séquençage d'ADN automatisé par fluorescence, le projet du génome humain a été lancé dans le but d'identifier la séquence de toutes les paires de bases de nucléotides constituant l'ADN humain [2]. Au final, le projet, d'un coût de 3 milliards de dollars, a duré 13 ans et mobilisé un vaste consortium international de scientifiques [3]. En avril 2003, le génome complet était annoncé et le projet aboutissait avec un génome de référence humain de haute qualité [4].

Depuis lors, les technologies de séquençage ont évolué à un rythme soutenu. En 2005, l'introduction de la technologie de pyroséquençage 454 de Roche a marqué le début de la révolution NGS [5]. Ces technologies sont caractérisées par leur capacité à « lire » plusieurs millions de séquences de fragments d'ADN, en parallèle avec un débit élevé et à un coût bien moindre que les premières technologies de séquençage [6].

Il existe actuellement trois stratégies générales de NGS : (1) un NGS à lectures courtes utilisant la technologie Illumina, qui produit actuellement des lectures appariées de 150 pb avec des taux d'erreur faibles compris entre 0,1% et 0,5%; (2) WGS à lectures longues utilisant les technologies de monomolécules uniques de Pacific Biosciences ou Oxford Nanopore Technologies, qui permettent d'obtenir des lectures de 10 à 100 kb - et parfois beaucoup plus longues - avec des taux d'erreur élevés de l'ordre de 10% à 15 %; et (3) WGS à lecture liée utilisant la technologie de 10X Genomics, qui génère des lectures courtes d'Illumina codées à partir de molécules plus longues (par exemple, 50 kb). En raison de considérations de coût, de facilité d'utilisation et de précision, la très grande majorité des études de génétique humaine utilisent un NGS à lectures courtes avec le système Illumina HiSeq ou NovaSeq. Nous nous concentrerons donc principalement sur l'analyse de ce type de données.

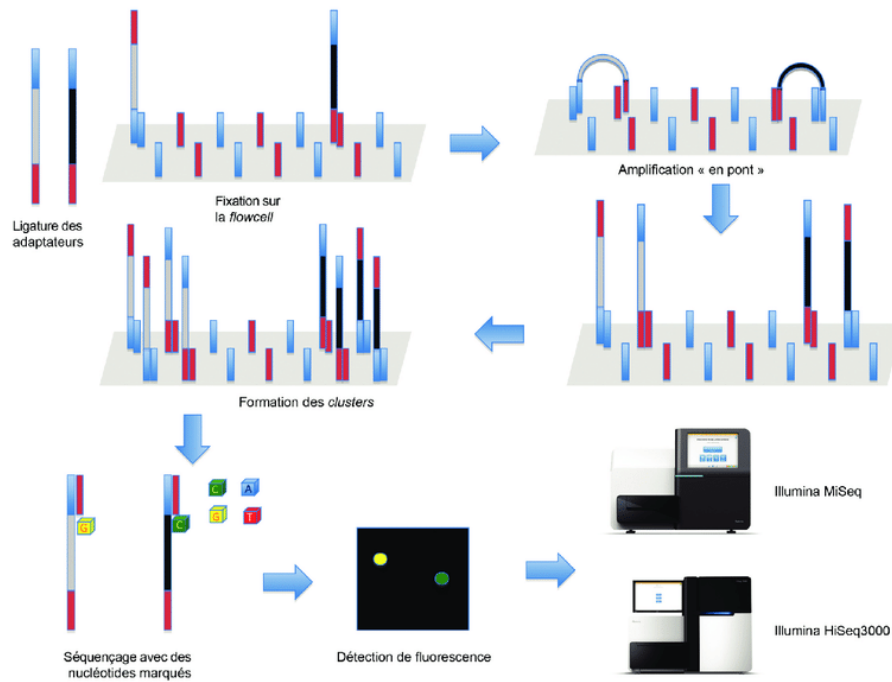
Néanmoins, il est clairement important de produire des lectures plus longues pour améliorer les résultats de diverses analyses, telles que l'assemblage *de novo* de grands génomes très répétitifs ou la détection d'évènements structuraux de grande taille (délétion, insertions, translocations). Comme nous traiterons dans les chapitres suivants uniquement de données de lectures courtes Illumina, le concept de données à lecture longue sera simplement mentionné et ne sera pas discuté plus en détail dans cette thèse. Cependant, produire des lectures plus longues et le séquençage de molécules simples sans étape d'amplification est certainement la direction que prendra le NGS.

## 1.2 Technologie Illumina

Le procédé de séquençage haut débit développé par la société Illumina est apparu comme l'un des procédés NGS les plus largement utilisés pour le séquençage d'ADN [7]. La précision et le débit élevés ainsi que les coûts toujours en baisse ont rendu la plate-forme Illumina particulièrement adaptée à de nombreuses questions biologiques. C'est la plateforme que nous avons utilisée pour tous les séquençage réalisés au cours de cette thèse, c'est pourquoi nous verrons dans cette partie le principe plus détaillé du séquençage Illumina et son protocole.

Dans une première étape de préparation des librairies d'ADN, l'ADN total est fragmenté en millions de fragments d'environ 150bp et les adaptateurs sont ligaturés à leurs extrémités. Les fragments sont alors prêts pour le séquençage. Ils sont fixés sur une plaque. La plaque, également appelée flowcell, consiste en plusieurs canaux avec une « forêt » dense d'amorces complémentaires aux adaptateurs fixées à sa surface et permettant aux fragments de se lier. L'amplification en pont est effectuée pour générer des « clusters » sur la plaque. Dans ce processus, l'adaptateur d'une extrémité libre d'un fragment déjà lié interagit avec l'amorce complémentaire fixée sur la plaque. Ensuite, un pont double brin est généré et après dénaturation, deux « templates » simple brin sont produits, capables de participer à une étape d'amplification suivante (Figure 3). Suite à ces étapes, des groupes de séquences identiques sont produits, nécessaires pour la dernière étape de séquençage : le séquençage en lui-même implique quatre nucléotides fluorescents et bloqués de manière réversible sur leur extrémité terminale 3'OH, si bien que leur incorporation empêche la liaison du nucléotide suivant. Ces derniers sont déposés simultanément

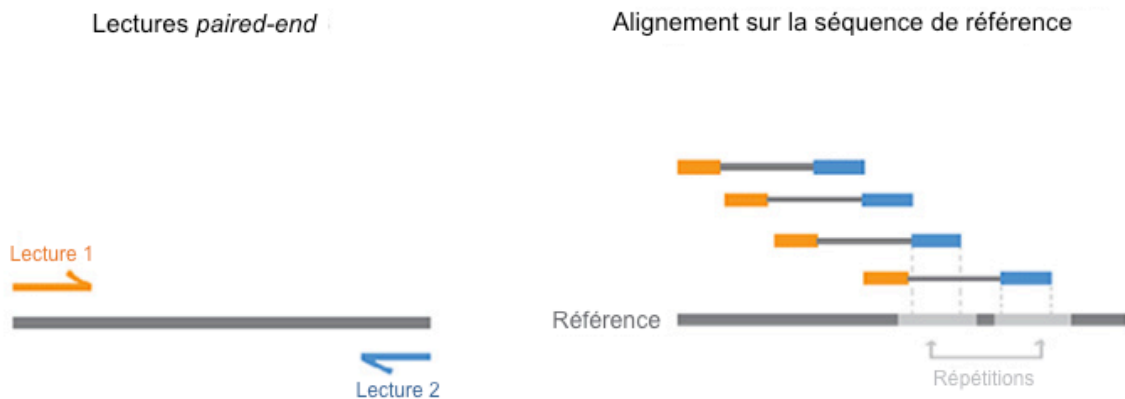
sur la flowcell. Lorsqu'un nucléotide se lie de façon complémentaire au brin matrice, une couleur de fluorescence spécifique est émise et la base nouvellement ajoutée peut être identifiée. Avant le cycle suivant, les fluorophores sont clivés et lavés de la flowcell et le groupe 3'-OH de chaque nucléotide est régénéré. À chaque tour de séquence, les lectures sont allongées d'au moins un nucléotide et finalement sauvegardées sous forme de simples chaînes de texte dans un fichier FASTQ [8] pour un traitement ultérieur (Figure 3).



**Figure 3:** Principe du séquençage Illumina: Après la ligature des adaptateurs sur les fragments d'ADN, ils sont fixés sur la flowcell en s'appariant à la forêt d'amorces. S'ensuit une amplification en pont permettant la formation de clusters de fragments identiques, qui sont alors séquençés à l'aide de nucléotides marqués. La détection de la fluorescence dans les appareils Illumina identifie la séquence des fragments. Source: Dortet L et al. Impact du séquençage d'ADN à haut débit sur la surveillance des épidémies de bactéries multi-résistantes aux antibiotiques. 2017, Feuilles de biologie, Vol 354.

Outre les typiques lectures *single-end* produites par une exécution standard d'Illumina, la technologie permet également la production de lectures dites *paired-ends*. Dans une conception à *paired-ends*, chaque fragment est séquençé à partir des extrémités 5' et 3', de sorte que deux lectures correspondantes soient produites à partir de chaque fragment. Comme la taille moyenne des fragments dans la librairie est connue, le calcul de l'espace séparant les deux lectures associées est possible

(Figure 4). Ces informations supplémentaires améliorent de manière significative le processus d'assemblage et la mise en correspondance des lectures dans les régions répétitives du génome.



**Figure 4 :** Alignement des lectures *paired-end* : le séquençage *paired-end* permet de séquencer les deux extrémités du fragment d'ADN. Comme la distance entre chaque lecture appariée est connue, les algorithmes d'alignement peuvent utiliser cette information pour mapper plus précisément les lectures sur des régions répétitives. Il en résulte un meilleur alignement des lectures, en particulier sur les régions répétitives et difficiles à séquencer du génome. Sources : [www.illumina.com](http://www.illumina.com)

### 1.3 Notion de profondeur de séquençage

Il est important de considérer lors de la conception des études de NGS, le niveau de profondeur souhaité. Elle correspond au nombre de fois qu'est lue une séquence d'ADN. Plus la profondeur est grande, plus la qualité du séquençage est élevée et permettra de distinguer les vrais variants des erreurs. Une couverture plus profonde améliore la sensibilité de détection des variants et améliore également la précision en permettant des schémas de filtrage plus sophistiqués. En général, les cas de maladies rares familiales ou isolées ciblent une couverture plus profonde (> 30x) pour assurer une détection robuste des variants hétérozygotes rares ou *de novo*. Les études de traits complexes à grande échelle peuvent cibler une couverture un peu plus basse (par exemple > 20x) pour augmenter la taille de l'échantillon pour un budget donné, tout en permettant une détection sensible des variants rares. En cancérologie, la profondeur utilisée est généralement plus grande, aux alentours de 100x, en raison de l'importante hétérogénéité tumorale et la présence de populations

de cellules peu représentées dans l'échantillon. De plus, certaines biopsies peuvent contenir qu'une minorité de cellules cancéreuses qui coexistent parmi des cellules normales (ou l'inverse) et il est alors nécessaire d'augmenter la profondeur de séquençage car on est alors dans une situation de mosaïsme cellulaire et donc génomique.

Au final la couverture optimale dépend des objectifs de l'étude et des échantillons étudiés.

## 2. Alignement sur le génome de référence

Comme nous l'avons vu l'ADN est fragmenté pendant la préparation des bibliothèques avant le séquençage. Toutes les lectures doivent ensuite être réassemblées informatiquement pour reconstruire la séquence du génome. Cela peut être effectué soit selon une séquence de référence déjà connue (assemblage basé sur une référence), soit sans utiliser les connaissances préalables (assemblage *de novo*).

De nombreux programmes ont été développés pour mapper les lectures sur une séquence de référence, avec des algorithmes et une vitesse variable. Certains outils couramment utilisés [9] pour les données de séquençage d'ADN sont BurrowWheeler Aligner (BWA) [10] et Bowtie [11]. La suite logicielle SAMtools [12] est largement utilisée pour convertir, trier et indexer des fichiers d'alignement (SAM / BAM) produits par les outils d'alignement actuels.

Fréquemment, il arrive que certaines lectures soient apparemment alignées avec la même qualité sur plusieurs emplacements du génome de référence. Cela peut être dû à des régions répétitives présentes dans un génome. Elles mettent en lumière la nécessité de faire évoluer le NGS vers la production de lectures plus longues qui permettraient d'aligner sans ambiguïté les séquences complexes du génome sur le génome de référence.

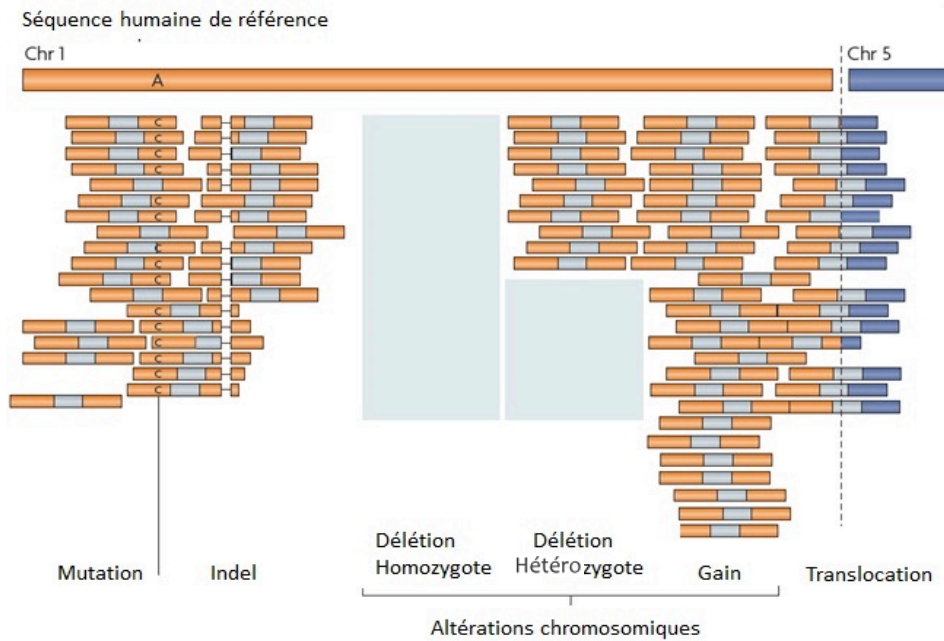
## 3. Les différents types de variants

L'alignement des séquences d'un individu sur le génome de référence met en évidence les variations génétiques de son génome communes et rares. Il existe plusieurs types de variants selon leur nature et leur taille.

### 3.1 Les variants mononucléotidiques (SNVs) et variants à petites insertions / délétions (indels)

Les SNVs et les indels (<50 pb) constituent la grande majorité des variants dans la population humaine. Dans le génome de chaque humain comparé à un génome de référence typique on retrouve en moyenne entre 3 et 4 millions de SNVs et environ 0,5 millions d'indels. Le catalogue dbSNP (build 151) compte plus de 660 millions de SNV et d'indels issus de diverses études de séquençage. Bien que l'immense majorité de ces variants n'a pas d'impact fonctionnel au niveau moléculaire ou phénotypique, chaque génome a plus d'une centaine de variants tronquant les protéines (*protein-truncating variant*, PTV) qui introduisent un codon d'arrêt prématuré, dont plus de 20 sont rares dans la population humaine et potentiellement délétère [13]. Des SNV faux-sens ou des insertions/délétions entraînent des modifications des acides aminés, qui peuvent être entièrement bénignes ou provoquer une maladie grave. Enfin, ces variants peuvent affecter la régulation des gènes en affectant les éléments régulateurs transcriptionnels et post-transcriptionnels. Ceux-ci incluent des motifs d'ADN spécifiques auxquels se lient des facteurs de transcription et d'épissage, souvent dans des régions régulatrices non codantes comme les régions 5'UTR, 3'UTR, et les séquences introniques donneur et accepteur d'épissage.

Ces petites variations de séquence constituent la classe de variants la plus facile à détecter à partir de données à lecture courte. Elles sont identifiées par les algorithmes de détection SNV/indel qui analysent le génome de référence à la recherche de lectures alignées présentant des incohérences, des insertions ou des suppressions de manière à suggérer une variation de la lignée germinale plutôt que des erreurs de séquençage ou d'alignement (Figure 5).



**Figure 5** : Types d'altérations génomiques détectées par séquençage haut débit. Sources : adapté de biorigami.com

Les outils existants largement utilisés (par exemple, [14]) sont très efficaces dans les régions codantes du génome et permettent un alignement précis de la lecture, avec des niveaux de sensibilité et de spécificité supérieurs à 99,5% pour les SNV et 95% pour les indels [15, 16].

Cependant, environ 8,5% du génome est considéré comme extrêmement difficile pour les détections de SNV/indel en raison de la présence de duplications segmentaires et/ou de répétitions de nombre élevé de copies qui entraînent un défaut d'alignement en lectures courtes [15]. Ces régions dupliquées comprennent des gènes cliniquement pertinents, et la mauvaise détection des variants de ces gènes est une faiblesse majeure du NGS basé sur les lectures courtes.

### 3.2 Les variants structuraux

Les variants structuraux (SV) sont une forme diversifiée de variation du génome d'une taille supérieure à 50pb qui inclut les variations du nombre de copies (CNV), les réarrangements, translocations et les insertions d'éléments mobiles. Les SV sont peu nombreux par rapport aux SNVs et aux indels mais ont des conséquences généralement plus importantes en raison de leur taille. Les SV peuvent exercer des

effets fonctionnels en fusionnant des gènes, en modifiant le dosage des gènes, en perturbant leur fonction ou en réarrangeant des éléments régulateurs et/ou des gènes pour modifier le contexte génomique. Sans surprise, les variants structuraux (délétions ou duplications chromosomiques) extrêmement grands qui suppriment ou dupliquent de nombreux gènes ou même des chromosomes entiers ont généralement des effets phénotypiques importants et ne sont pas observés chez la plupart des individus. Les formes de SV plus petites et plus répandues n'affectent en général qu'un ou quelques gènes ou se situent dans des régions non codantes. Bien que les SV ne représentent que 0,2% environ du total des variants répertoriés, des études récentes basées sur le WGS ont estimé qu'ils représentaient 3% à 7% des variants ayant des effets sur l'expression des gènes, et 4% à 12% des variants à effet délétère dans les régions codantes [17] suggérant leur impact plus élevé que les petits variants.

Les SV sont reconnus comme la forme de variation la plus difficile à détecter de manière fiable à partir de données NGS à lectures courtes. En effet, les grands éléments ne sont pas visibles entièrement sur une lecture courte et de plus, les SV sont enrichis en éléments répétitifs qui sont confondus durant l'alignement. Cependant il existe quelques approches générales qui permettent de cartographier les SVs. Dans l'ensemble, l'approche la plus précise et qui offre la plus haute résolution est la "cartographie des points de cassure". Cette méthode repose sur la détection directe de nouvelles jonctions de séquence qui ne sont pas présentes dans le génome de référence [18-21].

Les CNV plus grands (> 1 kb) tels que les grandes délétions ou les gains, ainsi que les translocations peuvent être détectés par analyse de la profondeur de lecture (Figure 5). Cependant cette approche comporte plusieurs limites en raison de sa faible sensibilité, de son taux élevé de faux positif et de sa faible résolution pour déterminer le point de cassure (insertion ou délétion). Seul un nombre modeste (5% à 10%) des CNV est détectable uniquement par une analyse en profondeur de la lecture dont la plupart sont des CNV de grande taille, et sont enrichis en gènes. Récemment une méthode appelée CoDE-seq, a été développée dans notre laboratoire afin d'augmenter la sensibilité, la résolution et l'efficacité pour l'identification de CNVs par analyse de la profondeur grâce à la mise en œuvre d'une nouvelle capture de séquençage d'exome dit « augmenté » [22].



#### 4. Interprétation fonctionnelle des variants

L'analyse fonctionnelle des variants constitue la partie essentielle de l'analyse génétique. Si la grande majorité des variants sont non fonctionnels et neutres, c'est à dire qu'ils n'engendrent pas de phénotype particulier, certains ont un impact sur la fonction et/ou la structure des protéines et peuvent causer des maladies. Il est donc primordial de mettre en place des filtres de sélection des variants permettant de prioriser les variants les plus probablement associés aux traits phénotypiques observés, parmi une liste de plusieurs milliers de candidats. Il faut bien dissocier les variants causant les maladies rares monogéniques et les variants associés aux maladies polygéniques, qui sont identifiés selon des stratégies de filtrages différents. Les premiers sont rares et ont un fort impact sur les régions codant les gènes tandis que les seconds sont plus communs dans la population et se situent souvent dans les régions non codantes. Au cours de cette thèse nous nous sommes intéressés particulièrement aux variants en cause dans les maladies monogéniques.

L'American College of Medical Genetics (ACMG) a publié des directives pour l'interprétation des variants [23]. Ces directives proposent différents critères pour classer les variants, incluant la fréquence dans la population, la prévalence chez les individus affectés, les données de ségrégation, les études fonctionnelles, le type de variant, les similarités avec des variants connus, et les prédictions informatiques. En fonction de la solidité des critères de classification d'un variant, les directives ACMG permettent de le déterminer comme pathogène, probablement pathogène, signification incertaine, probablement bénin ou bénin selon les règles du Tableau 1 [23]. Nous développerons dans cette partie les principaux critères utilisés par cette méthode.

|  | Benign   |   |   | Pathogenic  |   |   |
|--|--|---|---|---|---|---|
|  | Strong   | Supporting  | Supporting  | Moderate  | Strong  | Very strong   |
| <b>Population data</b>                   | MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2 |   |   | Absent in population databases PM2  | Prevalence in affecteds statistically increased over controls PS4 |   |
| <b>Computational and predictive data</b> |  | Multiple lines of computational evidence suggest no impact on gene /gene product BP4<br>Missense in gene where only truncating cause disease BP1<br>Silent variant with non predicted splice impact BP7<br>In-frame indels in repeat w/out known function BP3 | Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3 | Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5<br>Protein length changing variant PM4 | Same amino acid change as an established pathogenic variant PS1   | Predicted null variant in a gene where LOF is a known mechanism of disease PVS1 |
| <b>Functional data</b>                   | Well-established functional studies show no deleterious effect BS3                                       |   | Missense in gene with low rate of benign missense variants and path. missenses common PP2           | Mutational hot spot or well-studied functional domain without benign variation PM1  | Well-established functional studies show a deleterious effect PS3 |   |
| <b>Segregation data</b>                  | Nonsegregation with disease BS4  |   | Cosegregation with disease in multiple affected family members PP1                                  | Increased segregation data →  |   |   |
| <b>De novo data</b>                      |  |   |   | De novo (without paternity & maternity confirmed) PM6   | De novo (paternity and maternity confirmed) PS2                   |   |
| <b>Allelic data</b>                      |  | Observed in <i>trans</i> with a dominant variant BP2<br>Observed in <i>cis</i> with a pathogenic variant BP2  |   | For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3   |   |   |
| <b>Other database</b>                    |  | Reputable source w/out shared data = benign BP6   | Reputable source = pathogenic PP5   |   |   |   |
| <b>Other data</b>                        |  | Found in case with an alternate cause BP5   | Patient's phenotype or FH highly specific for gene PP4  |   |   |   |

**Tableau 1** : Directives ACMG pour la classification des variants comme bénins ou pathogènes, basée sur les types de preuve et la force des critères. D'après Richards et al., 2015 [23].

#### 4.1 Fréquence allélique des variants (MAF)

La fréquence allélique est un facteur clé à considérer dans l'analyse. Plus de 95% des variants d'un individu sont communs dans la population, c'est à dire d'une fréquence >5%.

Comme nous l'avons dit plus haut, la majorité des variants retrouvés chez un individu sont bénins, mais quelques-uns peuvent être délétères. La plupart des variants délétères sont rares à cause de la sélection négative subie selon les lois de la sélection naturelle. La rareté d'un variant dans la population reflète les pressions de sélection agissant sur lui ce qui en fait probablement l'indicateur le plus puissant de sa pathogénicité. Ainsi, les études de génétique sur les maladies Mendéliennes ou

congénitales ayant un fort impact sur la capacité de reproduction du porteur utilisent typiquement la fréquence allélique comme premier critère de sélection de variants.

Plusieurs ressources existent afin d'estimer la fréquence d'un variant dans la population.

De nombreuses bases de données de référence à grande échelle sont devenues accessibles pour la communauté biomédicale. Le projet 1000 Genomes (1000G) inclut des données de génotypes individuels à partir de données de séquences de 2 504 génomes entiers et d'exomes [24]. Le Exome Variant Server, créé dans le cadre du projet de séquençage de l'exome (ESP) contient des informations de fréquence couvrant 6 503 exomes [25]. Exome Aggregation Consortium (ExAC) a rassemblé des données de séquençage d'exome complet provenant de 60 706 individus non apparentés et fournit des informations sur différentes populations humaines [13]. La base de données Genome Aggregation Database (gnomAD) récemment publiée est une vaste source de variants, comprenant 123 136 séquences d'exome et 15 496 séquences de génome entier d'individus non apparentés, séquencées dans le cadre de diverses études génétiques spécifiques à une maladie et à l'ensemble de la population (<http://gnomad.broadinstitute.org/>).

Plusieurs études ont tenté d'évaluer le seuil de la fréquence allélique au-dessous duquel un variant est considéré pathogénique. Une étude, ayant passé au crible un ensemble de 1 508 variants pathogènes impliqués dans 79 maladies (39 dominantes et 40 récessives), a révélé que 97,3% avaient un MAF <0,01% selon la base de données ExAC [26]. Selon les directives ACMG-AMP 2015, une «fréquence d'allèle supérieure à celle attendue pour un trouble» devrait être considérée comme une preuve solide d'une classification bénigne [23]. D'après les conclusions de l'étude ci-dessus, les fréquences globales d'allèles ExAC > 0,01% devraient être considérées comme «plus grandes que prévu» pour les maladies Mendéliennes, et ce seuil pourrait être encore abaissé pour certains gènes [26]. D'autre part, certaines études ont suggéré une méthode de calcul des seuils de MAF utilisant un cadre statistique prenant en compte la prévalence de la maladie, l'hétérogénéité génétique et allélique, le mode d'héritage, la pénétrance de la pathologie [27].

Cependant, des seuils MAF plus stricts peuvent augmenter le risque de classification incorrecte des variants pathogènes avec une fréquence élevée comme bénignes [28, 29]. C'est le cas pour certains variants observés à une fréquence de 1% dans la population européenne et ne devant pas être classés comme bénins sans une revue

de littérature approfondie; ainsi, une évaluation minutieuse est nécessaire pour évaluer le pouvoir pathogène d'un variant [29].

#### 4.2 Prédictions informatiques

La validation expérimentale de tous les variants faux-sens est souvent irréalisable techniquement. Lorsqu'elle l'est, elle entraîne un coût considérable en main-d'œuvre et en ressources. Pour faire face à ces limitations, un nombre croissant de méthodes de calcul *in silico* ont été développées, basées sur l'homologie de séquence, la structure de la protéine et la conservation évolutive [30-32]. En général, ces méthodes peuvent être classées en trois types: (i) méthodes de prédiction de fonction qui prédisent la probabilité qu'un variant faux-sens donné provoque des modifications pathogènes de la fonction des protéines, (ii) méthodes de conservation qui utilisent des alignements multiples pour mesurer le degré de conservation sur un site nucléotidique donné et (iii) des méthodes d'ensemble intégrant ces deux composants.

Les cinq algorithmes les plus couramment utilisés sont Polymorphism Phenotyping (PolyPhen) [33] and Sorting Intolerant From Tolerant (SIFT) [34], Protein Variation Effect Analyzer (PROVEAN) [35], Combined Annotation-Dependent Depletion (CADD) [36] et Mutation Taster. Les directives ACMG préconisent de tenir compte des prédictions de ces algorithmes qui suggèrent soit aucun impact, soit un effet néfaste sur un gène ou produit de gène. En revanche les directives alertent sur le risque de surestimer les preuves informatiques et leurs capacités de prédiction. Par conséquent, il est important d'interroger plusieurs algorithmes différents et de leur faire confiance uniquement lorsque tous les algorithmes de prédiction s'accordent [23]. C'est en effet une source majeure de discordance de résultats entre différents laboratoires qui utilisent des processus subjectifs pour choisir certains critères d'utilisation différents des algorithmes *in silico* [37].

#### 4.3 Type de variant et données fonctionnelles

La pathogénicité d'un variant repose essentiellement sur le type du variant et sa localisation dans le gène. Comme indiqué dans les directives ACMG, chaque type de variant a un effet différent sur le produit du gène. De plus, chaque gène a ses spécificités et des domaines fonctionnels particuliers, impliqués dans les

mécanismes par lesquels le gène induit la maladie [38]. Il faut donc prendre en compte lors de l'analyse d'un variant s'il est situé dans un domaine fonctionnel du produit du gène. À partir de ce constat ACMG a établi un classement des variants selon les cas de figure [23] (Tableau 1). Ainsi un variant tronquant dans un gène pour lequel un mécanisme de perte de fonction cause la maladie est considéré comme une preuve très forte de pathogénicité. Les variants qui présentent un même changement d'acides aminés qu'un variant pathogène déjà établi sont des preuves fortes. Les preuves modérées de pathogénicité incluent les variants sur un point névralgique de variants (« hotspot ») ou un domaine fonctionnel, ainsi que les nouveaux changements d'acide aminé dans un variant faux-sens pathogène précédemment signalé.

En revanche les variants faux-sens dans un gène où seule la troncature cause la maladie, ainsi que les variants synonymes sans prédictions d'impact sur l'épissage sont considérés comme des preuves modérées de bénignité (Tableau 1).

#### 4.4 Ségrégation des variants

Les directives ACMG incluent également la ségrégation ou non du variant avec les individus malades au sein de familles présentant plusieurs cas. Ce paramètre permet de lier un variant avec un phénotype, d'autant plus fortement que le variant co-ségrége avec une grande proportion de membres de la famille atteints par la maladie (et est absent des sujets sains). Ces données de ségrégation doivent cependant être prises avec prudence car elles peuvent au mieux indiquer une association entre un variant et une maladie mais pas sa pathogénicité intrinsèque [23, 39]. Des seuils pour exprimer les degrés de ségrégation ont été mis en place, basés sur le nombre de membres atteints partageant le variant. Cette méthode de mesure a permis de séparer trois niveaux de co-ségrégation d'un variant avec une maladie : faible pour  $\geq 3$  individus partageant une maladie dominante, moyenne pour  $\geq 6$  individus et forte pour  $\geq 10$  individus. Ceci n'est vrai que pour les variants à transmission dominante et correspond à la notion ancienne de lodscore, c'est-à-dire de probabilité que cette co-transmission ne soit pas due au hasard.

#### 4.5 Prédictions fonctionnelles

Enfin, le dernier élément à considérer lors de la prédiction de l'effet d'un variant sur le produit du gène est l'existence d'études fonctionnelles, déterminant si le variant a un impact délétère ou non sur le produit du gène [23]. Il convient de définir la valeur de l'étude fonctionnelle, qui dépend de la qualité de l'expérience qui prouve un effet fonctionnel sur le variant. Il sera jugé la reproductibilité des résultats, l'importance des effets observés, et les modèles expérimentaux utilisés [38]. En effet, une étude fonctionnelle a d'autant plus de valeur qu'elle reproduit le plus fidèlement les effets du variant sur le phénotype, ainsi une souris knock-out pour un gène donné sera un meilleur modèle pour imiter un variant perte de fonction récessive qu'un variant dominant. Nous verrons dans le prochain chapitre des approches expérimentales pour la caractérisation fonctionnelle des VUS (*variants of unknown significance*).

## 1.5. Validation fonctionnelle de l'impact des variants

Étant donné que notre capacité à prédire les effets moléculaires des variants génétiques est imparfaite, l'approche complémentaire consiste à mesurer empiriquement les effets moléculaires des variants génétiques chez les individus qui les portent. Les tests fonctionnels sont donc la seule option permettant d'obtenir des preuves concluantes de la pathogénicité des variants identifiés. Dans certains cas, des données fonctionnelles de variants génétiques peuvent déjà être disponibles, par exemple dans des bases de données contenant des données fonctionnelles. Dans de nombreux cas, ces informations ne seront pas disponibles et de nouvelles investigations sur le variant génétique en question sont nécessaires. Il existe de nombreux moyens d'obtenir des preuves de la pathogénicité des variants génétiques au moyen d'essais fonctionnels spécifiques au niveau de l'ARN et au niveau protéique

### 5.1 Au niveau de l'ARN

Au niveau de l'ARN, les variants génétiques peuvent affecter les niveaux de transcription, l'épissage de l'ARNm ou la stabilité du transcrit. Les prédictions *in silico* sont généralement utilisées pour identifier les variants susceptibles d'altérer l'épissage de l'ARNm [40]. La validation technique de ces prévisions comprend des expériences telles que l'analyse de l'ADN complémentaire (ADNc) dans des tissus de porteurs du variants [41, 42]. Par exemple, l'analyse de l'ARN de trois nouveaux

VUS au niveau de sites d'épissages de PMS2, c.538-3C>G, c.164-2A>G et c.989-2A>G, a confirmé leur impact fonctionnel. En effet, les analyses de l'ADNc via le séquençage des produits de RT-PCR ont confirmé la présence de transcrits aberrants de PMS2 responsables du syndrome de Lynch [43].

## 5.2 Au niveau protéique

D'autre part, les analyses fonctionnelles au niveau des protéines sont importantes dans l'évaluation de la pathogénicité d'un variant, à condition que la propriété évaluée soit pertinente pour le mécanisme de pathogénicité du gène de cette maladie. Un variant peut affecter la protéine à différentes étapes moléculaires, par exemple au niveau de son expression et de sa stabilité, de sa localisation subcellulaire, sa capacité à formation de complexe, ou d'une fonction spécifique. Dans le cas où un le gène muté est impliqué dans une voie de signalisation (kinase, ubiquitinyase, méthyltransférase etc), une stratégie classique est d'observer dans les cellules du patient si ses cibles en aval sont modifiées par la protéine mutée de la même manière que chez le contrôle.

## 5.3 Stratégies de corrélation génotype-phénotype

Il y a en général deux manières d'établir une corrélation génotype-phénotype comme preuve de la pathogénicité des variants génétiques. La première stratégie consiste à démontrer que les cellules de l'individu atteint présentent un phénotype biochimique ou biologique mesurable dont on sait qu'il est dépendant du gène d'intérêt, puis à « sauver » le phénotype en introduisant une version sauvage (WT) du gène. En général, des cellules dérivées du patient sont utilisées, par exemple des lymphoblastes ou des fibroblastes de la peau [44].

La deuxième stratégie consiste à tester les variants génétiques dans des systèmes modèles. Ceux-ci ne doivent pas nécessairement être des modèles dérivés de l'individu, mais peuvent par exemple être des modèles de culture cellulaire ou des modèles animaux. Ces dernières années, la technologie d'édition de gènes CRISPR / Cas9 (répétitions palindromiques courtes régulièrement groupées) a transformé la génomique fonctionnelle, permettant ainsi aux chercheurs de modifier n'importe quelle région du génome, et ainsi de recréer un variant donné dans le modèle souhaité. Ainsi, une étude a récemment reproduit une dizaine de VUS dans le gène

MSH2 associé au syndrome de Lynch par CRISPR / Cas9 dans des cellules souches embryonnaires humaines. La caractérisation de l'impact des variants sur les fonctions cellulaires de réparations de l'ADN ont donné d'importantes informations sur leur potentielle implication dans le syndrome de Lynch [45].

La modélisation des variants candidats par la technologie CRISPR / Cas9 a permis la caractérisation de nouveaux gènes impliqués dans le développement de maladies, ainsi que les variations génétiques somatiques guidant la progression des tumeurs.

## II - Les applications du NGS en clinique

Il ne fait aucun doute que l'émergence des technologies NGS a constitué un tournant décisif dans l'avancement de notre compréhension de la biologie des maladies. La capacité du NGS à identifier les variations génétiques du patient à l'échelle du génome entier, dans un délai raisonnable et à moindre coût, a marqué une révolution dans la prise en charge des patients. La démocratisation du NGS a permis de séquencer des centaines de milliers de génomes et d'exomes durant ces dix dernières années et de constituer d'immenses bases de données telles que 1000Genome ou encore gnomAD. Elles ont été essentielles à l'identification de nouveaux gènes et altérations génétiques impliqués dans une multitude de pathologies.

Cela a permis d'ouvrir la voie à la médecine personnalisée, qui vise à optimiser et personnaliser les stratégies de prévention, de détection et de traitement en se basant sur les caractéristiques individuelles de chaque individu. Ces nouvelles connaissances moléculaires ont fortement contribué à l'avancement de la médecine de précision en fournissant un meilleur diagnostic aux patients atteints de cancer, et de maladies Mendéliennes et dans certains cas des stratégies spécifiques de thérapies ciblées.

Dans ce chapitre, nous présenterons les applications présentes et futures du NGS sur la prise en charge clinique des patients souffrant de maladies rares mendéliennes, et de maladies complexes telles que le cancer. Nous évoquerons dans un deuxième temps les limites actuelles de cette technologie en clinique, ainsi que les enjeux actuels de la recherche en génomique.



## 1. Stratégie de NGS en clinique : WGS versus WES versus séquençage ciblé

Selon l'objectif et/ou le budget de l'étude, on ne souhaite ou ne peut pas toujours séquencer le génome en entier. On peut vouloir séquencer uniquement les parties du génome codant pour des protéines ou bien un panel de gènes particulier associés à une maladie. La sélection des régions d'intérêt s'effectue en amont du séquençage, à la suite de l'étape de fragmentation, et a pour objectif d'enrichir la librairie en fragments d'ADN désirés. Cet enrichissement peut se faire soit par capture à l'aide de sondes s'hybridant aux fragments désirés, soit par PCR où les régions d'intérêt sont amplifiées.

Le séquençage ciblé d'un panel de gènes est le plus accessible des types de séquençage. Il est très utilisé en clinique car il permet d'obtenir rapidement les résultats et de diminuer les coûts. Dès lors qu'un diagnostic est suspecté, le panel de gènes adéquat associé aux pathologies observées est séquencé dans le but d'identifier le variant causal. Il existe de nombreux panels de gènes proposés pour presque toutes les pathologies. En revanche, le séquençage ciblé se limite au panel de gènes choisi et ne permet pas une analyse exhaustive à grande échelle comme le WGS et le WES. Il est donc trop limité dans les cas où le phénotype est nouveau ou incertain et dans les maladies très hétérogènes génétiquement, et ne permet l'identification de variants uniquement dans des gènes connus. Lorsque le séquençage ciblé d'un panel de gènes est négatif, un WES ou un WGS doit être envisagé pour identifier le variant causal.

Le WGS et le WES présentent chacun des avantages et des inconvénients qu'il convient de prendre en compte avant de commencer un projet de séquençage NGS. Sans surprise le WGS est plus complet car il permet l'examen de tous types de variants dans les régions codantes et non codantes du génome alors que le WES omet toutes les régions intergéniques et les régions régulatrices telles que les promoteurs et les amplificateurs. De plus, le WGS a une couverture de séquence plus fiable que le WES en raison des différences d'efficacité de l'hybridation des sondes de capture WES.

Quant au WES, il ne cible que les régions codantes pour les protéines ce qui permet de réduire significativement les coûts de séquençage, de stockage et d'analyse bio-informatique grâce à la nette diminution du nombre de variants à prioriser et à valider. De ce fait, cela permet de séquencer avec une plus grande profondeur

moyenne qu'en WGS, et d'économiser du temps lors du traitement des données. Les coûts réduits permettent d'augmenter le nombre d'échantillons à séquencer, ce qui permet des études sur des populations plus importantes.

Dans le cadre des travaux de cette thèse, nous avons toujours opté pour le WES en défaveur du WGS. En plus des inconvénients du WGS notamment son coût élevé, il est très difficile à l'heure actuelle d'interpréter des régions non codantes qui constituent 98% du génome. En effet, dans la recherche de variants diagnostiques, la grande majorité des variants interprétables actuellement se situent dans les régions codantes des gènes. La complexité du génome fait que la fonction des régions non codantes reste en grande partie inconnue, de ce fait il est encore difficile de prédire l'impact des variants dans les régions non codantes du génome et l'immense majorité des variants de ces régions sont de signification inconnue (VUS).

## 2. Stratégies de NGS d'identification de nouveaux gènes/loci associés aux maladies

En ce qui concerne les stratégies de recherche de nouveaux gènes ou loci associés à des maladies ou traits phénotypiques, il faut dissocier les maladies mendéliennes dites monogéniques reposant sur l'altération de la copie d'un seul gène, et les maladies polygéniques qui reposent sur l'accumulation de plusieurs variants d'impact faible à modéré chez les individus mais qui provoquent ensemble une prédisposition à une pathologie. Certaines pathologies complexes comme le cancer et le diabète se sont avérées sous des formes monogéniques ou polygéniques.

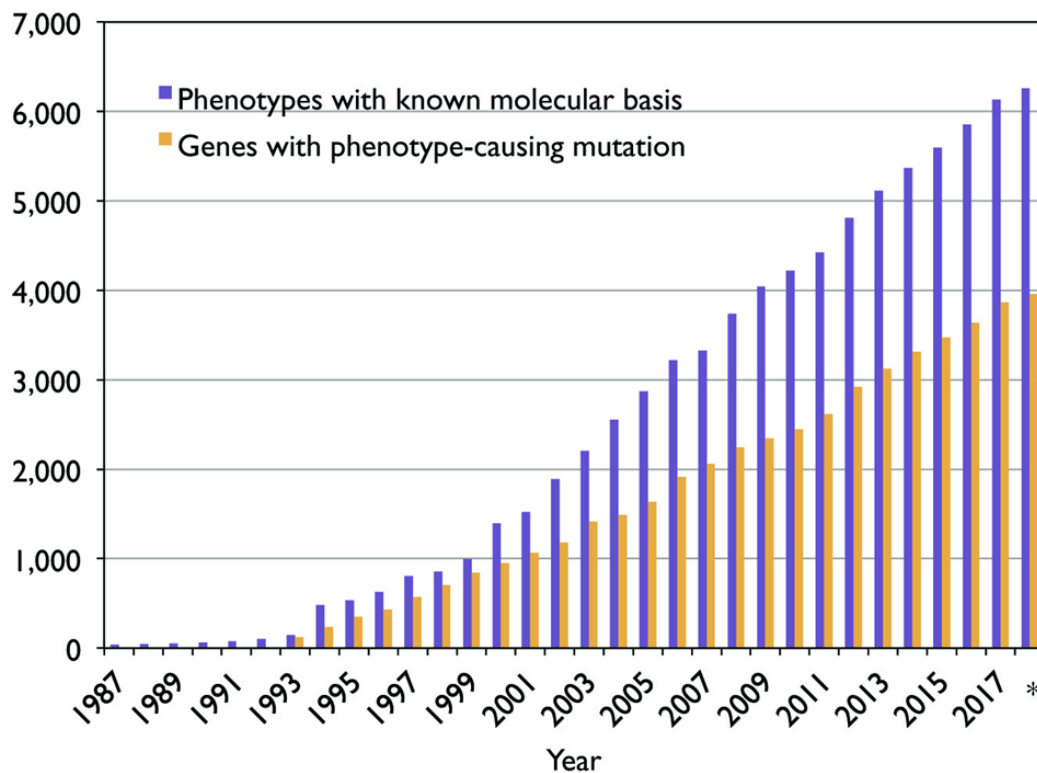
### 2.1 Maladies mendéliennes monogéniques

Le nombre de gènes nouvellement identifiés associés à une maladie a augmenté de façon très importante dans tous les domaines de la médecine depuis l'émergence de la technologie NGS, passant de 1300 associations gène-phénotypes connues en l'an 2000 contre plus de 6000 aujourd'hui (figure 6). Cette explosion de connaissances a eu lieu grâce au NGS, qui a pu être utilisé pour séquencer l'ensemble du génome des patients de manière rapide et sensible, au lieu de ne cibler que quelques gènes de susceptibilité.

Trois principaux tests basés sur le NGS sont utilisés dans la recherche sur les variants rares. Ces tests peuvent être classés en fonction du coût, de la facilité

d'analyse et de la portée, et comprennent: 1) le séquençage ciblé de séquences codantes (exons) de panels de gènes liés par des phénotypes similaires ou se chevauchant (panel de gènes); 2) le WES, dans lequel toutes les régions codantes connues du génome humain sont séquencées; et 3) le WGS, qui analyse l'ensemble du génome humain. Le WES a dominé la recherche sur les variants rares ces dernières années car il offre le meilleur compromis scientifique (couverture de tous les gènes contrairement au séquençage ciblé) et économique (plus économique et rentable que le WGS).

L'identification de nouveaux gènes associés à une maladie mendélienne humaine contribue à notre compréhension des bases biologiques des maladies et revêt une grande importance pour la gestion des patients et, dans certains cas, même pour une intervention thérapeutique [46]. Deux stratégies principales ont été utilisées pour isoler les gènes responsables de la maladie: 1) le séquençage d'un groupe de patients présentant les mêmes caractéristiques cliniques puis le filtrage de variants localisés dans un gène commun à tout ou partie du groupe ; et 2) l'analyse de patients isolés conjointement avec les parents et/ou les membres informatifs de leur famille, et le filtrage des variants selon différents modes de transmission (hypothèse autosomique dominante, récessive, liée à l'X ou *de novo*) afin de réduire le nombre de variants à un nombre suffisamment petit pour permettre l'identification du gène causal.



Source: *Online Mendelian Inheritance in Man*

**Figure 6** : Évolution du nombre d'associations gène-phénotype recensées dans la base de données OMIM au cours des 30 dernières années. D'après Amberger et al., 2019 [47].

## 2.2 Maladies complexes et stratégies « GWAS »

De nombreuses maladies parmi les plus courantes et les plus sévères, telles que les maladies cardiovasculaires, le cancer, la maladie d'Alzheimer, la maladie de Parkinson et le diabète de type 2 ne sont le plus souvent pas causés par des variants uniques [48, 49]. Ces maladies sont plutôt influencées par une combinaison de multiples facteurs de risque génétiques et environnementaux, et ne suivent donc pas le schéma classique d'un héritage mendélien [50]. Le modèle de départ «un gène, un variant, une conséquence» posé n'est pas applicable pour élucider la biologie de ces maladies. Les caractères complexes, par définition, sont influencés par de nombreux gènes - la taille humaine, par exemple, semble être affectée par la variation génétique de plusieurs centaines ou milliers de loci sur le génome [51]-, qui peuvent interagir de manière additive ou non.

Partant d'une série d'études de 2005 liant un variant proche du gène du facteur H du complément au risque de dégénérescence maculaire liée à l'âge [52-54], les

chercheurs du monde entier ont mené des études d'association pangénomiques (GWAS) pour identifier les loci qui portent des variants génétiques (généralement des SNP) associées à toutes sortes de maladies et de traits complexes. Les GWAS plus récents combinent les données des participants de plusieurs bases de données sous la forme d'une méta-analyse [55] pour analyser l'association de millions de variants avec un phénotype donné.

L'ère GWAS a été une réussite, en ce sens que des milliers de loci ont été associés statistiquement au risque de maladies et de traits phénotypiques, et qu'un nombre considérable de ces loci l'ont été de manière répétée, suggérant qu'elles sont de vraies associations [56]. Cependant, les mécanismes par lesquels ces variants agissent et prédisposent à certaines maladies sont en grande partie inconnus. En effet, nombre de ces loci sont situés dans des régions non codantes, et ont probablement des rôles de régulation souvent peu connus. Après plus d'une décennie d'études GWAS et la découverte de centaines de milliers de loci associés à des milliers de traits phénotypiques, nous entrons dans une nouvelle ère post-GWAS dont les enjeux sont la compréhension fonctionnelle des mécanismes biologiques par lesquels les loci identifiés par GWAS influent sur le phénotype [57].

Au cours de cette thèse nous avons considéré toutes les pathologies étudiées comme étant mendéliennes et monogéniques, et nous nous sommes limités à la recherche de variants rares, c'est pourquoi nous ne détaillerons pas plus dans cette introduction les stratégies de recherche sur les maladies polygéniques.

### 3. Application du WES pour le diagnostic de maladies rares mendéliennes

Les maladies génétiques rares non syndromiques sont causées par le dysfonctionnement d'un seul gène. Selon les gènes affectés, elles peuvent n'être causées par l'altération que d'un seul allèle du gène et se transmettre de manière autosomique dominante, ou bien nécessiter l'affection des deux allèles du gène et se transmettre de manière autosomique récessive. Les situations d'« haplo-insuffisance » par exemple, ont lieu lorsque le produit d'un seul allèle, bien qu'actif, est synthétisé en quantité insuffisante pour permettre le fonctionnement normal de la cellule provoquant des maladies transmises de manière dominante.

Dans les cas où la maladie génétique atteint plusieurs membres de la famille sur plusieurs générations, la stratégie pour identifier le variant en cause est

principalement d'identifier les variants qui co-ségrègent avec les individus atteints. Dans les cas où les parents de l'enfant atteint sont sains, le mode de transmission de la maladie se transmet généralement par des variants *de novo* ou hétérozygotes composites.

### 3.1 Maladies autosomiques dominantes rares : le rôle de premier plan des variants *de novo*

Les variants *de novo* sont des variants apparaissant par hasard chez un individu sans être héritée de l'un des parents et sont en principe hétérozygotes à effet dominant. De ce fait ils sont de bons candidats dans les cas où un enfant atteint d'une maladie rare est issu de parents non atteints. Le fait qu'ils soient hétérozygotes les rend susceptibles d'être causaux, c'est à dire que la présence d'un seul allèle muté suffit pour que la maladie se déclare.

La disponibilité de puces à ADN génomiques à haute résolution au cours de la dernière décennie a permis une analyse non biaisée des CNV sur l'ensemble du génome bien avant que cela ne soit possible pour les SNV et les indels *de novo*. De telles analyses des CNV ont révélé l'importance des variations génomiques *de novo* dans les troubles neuro-développementaux tels que la déficience intellectuelle, les troubles du spectre autistique (TSA) et la schizophrénie [58, 59]. Les CNV *de novo* de plus de 100 kb sont peu fréquents dans la population générale (présents chez environ un individu sur cinquante). Une fréquence d'environ 10% est toutefois estimée chez les patients présentant une déficience intellectuelle sporadique [59, 60], un TSA [61] ou une schizophrénie [62].

Le NGS a mis en évidence le rôle prépondérant des SNV *de novo* dans le développement de maladies rares, et en particulier de troubles du développement neurologique tels que l'épilepsie, des défauts de morphogenèse cérébrale, une déficience intellectuelle, des TSA et la schizophrénie. Les variants *de novo* représentent la forme la plus à risque de variations génétiques rares: elles sont en moyenne plus délétères que les variations héritées, car elles n'ont pas été sélectionnées à travers les générations par le processus de l'évolution naturelle. Les SNV *de novo* fournissent un mécanisme par lequel les maladies précoces létales pour la reproduction restent fréquentes dans la population. Ils sont donc des candidats idéaux pour provoquer des maladies génétiques graves qui se produisent sporadiquement. En outre, ils peuvent se produire dans la lignée germinale, au cours

de la genèse de l'embryon ou de manière somatique. Comme les SNV *de novo* ne sont pas des événements rares dans leur ensemble, elles pourraient être responsables d'une fraction importante des maladies les plus courantes.

### 3.2 Maladies récessives rares: le rôle de premier plan des variants hétérozygotes composites et homozygotes

Dans les troubles génétiques récessifs, les deux copies d'un certain gène sont défectueuses. Cela signifie que la copie d'un gène transmise par la mère aussi bien que par le père est porteuse d'un variant pathogène. Il existe deux cas de figures possibles pour la mutation de deux copies d'un gène :

- Les deux copies du gène peuvent être porteurs du même variant pathogène, auquel cas le variant est homozygote,
- Soit elles sont porteuses de variants différents, auquel cas le variant est dit hétérozygote composite.

L'apparition de variants pathogènes est un processus aléatoire et leur survenue dans les gènes récessifs est relativement rare dans la population humaine issue de parents non consanguins sauf si la prévalence de ces variants est importante (comme dans la mucoviscidose 1/50 en Europe). Il est donc plus probable de retrouver le même variant chez le père et la mère dans des familles où les parents sont consanguins. Dans le cas d'un couple non consanguin, la probabilité que deux allèles du gène mutés différemment soient présents chez l'enfant est plus élevée.

Cela se traduit par une simple règle empirique: si les parents sont consanguins, l'hypothèse la plus probable est que la maladie récessive soit causée par un variant homozygote chez l'enfant. En revanche si les parents sont non consanguins, l'explication la plus probable d'une maladie récessive est l'hétérozygotie composée par deux variants pathogènes différents.

Les notions de variants *de novo* et hétérozygotes composites sont très utiles pour trouver la cause génétique d'une maladie chez un enfant issu de parents non apparentés. Cela justifie l'intérêt d'effectuer les tests WES en trio et permet pendant l'analyse de faire un filtrage supplémentaire ce qui réduit significativement le nombre de gènes candidats.

### 3.3 Quelques exemples concrets

L'année 2010 a été marquée par les deux premières applications du WES pour le diagnostic de maladies rares. Dans le premier cas, une équipe américaine a découvert la cause génétique du syndrome de Miller, caractérisé par une micrognathie grave, une fente labiale et / ou palatine, une hypoplasie ou une aplasie des éléments postaxiaux des membres, un colobome du paupières et mamelons surnuméraires [63]. Ils ont effectué le WES sur quatre personnes non apparentées atteintes de ce syndrome dans trois familles indépendantes. Après avoir filtré leurs SNPs avec les bases de données publiques, et recherché des gènes comportant des variants rares chez chacun des quatre individus, ils ont pu identifier un seul gène candidat, *DHODH*, qui code pour une enzyme clé de la biosynthèse de novo de la pyrimidine. Le séquençage de Sanger a confirmé la présence de variants dans *DHODH* dans trois autres familles atteintes du syndrome de Miller. La même année, la même approche a permis d'identifier les gènes responsables des syndromes de Kabuki (*MLL2*) [64], de Schinzel-Giedion (*SETBP1*) [65] et de Sensenbrenner (*WDR35*) [66].

Peu de temps après, des stratégies différentes ont commencé à paraître. Goldstein et ses collaborateurs en 2012 ont expérimenté le trio-WES, en séquençant douze individus atteints de maladies génétiques différentes non diagnostiquées avec leurs parents biologiques. Ils ont réussi à confirmer un diagnostic génétique chez six des patients et ont même été en mesure de proposer un traitement potentiel basé sur le diagnostic de certains d'entre eux [67]. La même année, un bilan du National Institutes of Health (NIH) des deux premières années de l'application de la technologie NGS pour l'établissement de diagnostics de maladies rares a été publié. Sur un total de 160 patients (47% d'enfants et 53% de troubles neurologiques) un diagnostic a pu être établi chez 39 patients soit 24% d'entre eux [68]. En 2013, une autre étude américaine a signalé un taux de diagnostic presque identique à 25% dans une étude dans laquelle WES avait été appliqué au diagnostic clinique d'une cohorte de 250 patients [69]. Le taux de diagnostic est resté inchangé lorsque la population de l'étude a été portée à 2000 individus [70]. Dans toutes les études, les taux de diagnostic étaient plus élevés lorsque la stratégie du trio-WES, plutôt que celle du patient uniquement, était appliquée. Les taux respectifs rapportés par Farwell et ses collègues [71] étaient de 37% contre 21%. De plus le rendement diagnostique du WES s'est avéré être 50% plus élevé que celui du séquençage de Sanger (par criblage séquentiel de nombreux gènes candidats) [72], en particulier



dans les cas de troubles hautement hétérogènes tels que la surdit , la c civit , les maladies mitochondriales et les troubles moteurs [69]. Dans la section suivante, nous passerons en revue, mais de mani re non exhaustive, les g nes nouveaux les plus r cents et / ou les plus pertinents qui ont  t  identifi s dans des maladies neurod veloppementales et neuromusculaires rares gr ce au NGS.

### 3.3.1 Troubles neurod veloppementaux

Les perturbations du d veloppement qui surviennent pendant la vie f tale sont g n ralement caus es par des modifications de la migration et / ou de la prolif ration des cellules neuronales conduisant   des troubles de la morphog nese c r brale.

Les  tudes WES ont jou  un r le important dans le diagnostic de la d ficience intellectuelle. Elles ont  tabli que des alt rations *de novo* sont la cause majeure des cas de d ficience intellectuelle dans la population, et identifi  des centaines de g nes responsables [73-75].

De m me, le WES a permis de mieux comprendre les bases mol culaires de l' pilepsie infantile, qui  taient jusqu'alors tr s m connues. Des  tudes de patients ont identifi  de nouveaux g nes associ s   cette pathologie et identifi  que les variants *de novo* dominants en sont la cause principale [76, 77].

Enfin, concernant les TSA, un groupe de maladies neurod veloppementales qui pr sentent une forte h ritabilit , le NGS a permis de mieux en comprendre les bases mol culaires, qui semblent  tre caract ris es par de rares  v nements *de novo* dans de nombreux g nes diff rents [78-80]. Les exomes de l'autisme sporadique r v lent des variants *de novo* dans des g nes dont les produits forment un r seau de prot ines hautement interconnect es [81].

Au total, les troubles psychiatriques n'avaient jamais  t  consid r s comme des maladies monog n tiques jusqu'  r cemment, mais l'utilisation du NGS chez les patients a permis de mettre en  vidence un paradigme mutationnel *de novo*, comprenant des SNV et des CNV [82].

### 3.3.2 Troubles neuromusculaires

Les troubles neuromusculaires sont des affections de l'unit  motrice ou de la composante sensorielle r sultant de troubles ou de l sions des motoneurones

(amyothopies spinales), des ganglions rachidiens, des racines dorsales, des nerfs crâniens et périphériques (neuropathie), des jonctions neuromusculaires (myasthénie) et des muscles (myopathie). Depuis l'utilisation du NGS, de nombreux gènes ont pu être associés à ces pathologies, et de nouvelles formes de myopathies en fonction du gène muté ont pu être déterminées.

Plusieurs nouveaux gènes se sont avérés sous-jacents à de nouvelles dystrophies musculaires en gaine, incluant *LIMS2* [83], le gène *BVES* [84], *TOR1AIP1* [85], et du gène *POGLUT1* [86].

De plus de nouvelles formes de myopathies myofibrillaires ont été découvertes, telles que la myopathie myofibrillaire 7, associée au gène *KY* [87], et la myopathie myofibrillaire 8, associée au gène *PYROXD1* [88]. Deux nouvelles myopathies centronucléaires ont également été caractérisées au niveau moléculaire: la myopathie centronucléaire 4, associée au gène *CCDC78* [89], et la myopathie centronucléaire 5, associée au gène *SPEG* [90]. Très récemment, un autre nouveau syndrome myasthénique congénital était associé à des variants du gène *MYO9A* [91], qui code une myosine non conventionnelle qui fonctionne comme un moteur moléculaire à base d'actine. Des variants récessifs entraînant la perte de fonction de gènes du complexe ASC-1, telles que *ASCC1* et *TRIP4* ont également été identifiées comme responsable d'une atrophie musculaire spinale (SMA) chez le fœtus, associées à des fractures osseuses congénitales [92-95].

Enfin, trois nouveaux troubles de neuropathies sensorielles héréditaires (1D, 1E, 1F) ont été caractérisés par WES impliquant les gènes *ATL1*, *DNMT1* et *ATL3*, respectivement [96-98].

Au total le NGS a mis en évidence de multiples formes moléculaires de maladies neuromusculaires présentant des phénotypes différents, ce qui a permis une plus grande précision dans le diagnostic de ces pathologies.

### 3.3.3 Troubles moteurs

La NGS a également conduit à des avancées significatives dans des troubles neurodégénératifs du mouvement tels que l'ataxie spinocérébelleuse (ACS), caractérisée par des problèmes de coordination causés principalement par une atrophie du cervelet, et la dystonie, un trouble du tonus musculaire lié au

dysfonctionnement des ganglions de la base. De récentes études de WES ont permis la découverte de nouveaux gènes associés aux troubles du mouvement.

La première découverte a été l'implication du gène *DOCK3*, qui joue un rôle important dans la croissance axonale et la réorganisation du cytosquelette, dans un trouble lié à une hypotonie, à une ataxie et à une déficience intellectuelle [99]. Un SNV *de novo* du gène *GRIK2*, membre de la famille des gènes du récepteur du glutamate ionotrope, s'est avérée être la cause de la maladie chez un individu souffrant d'ataxie, de retard de langage et de déficience intellectuelle [100]. Un variant homozygote de *SLC30A9*, impliquée dans l'homéostasie du zinc intracellulaire, a été découvert chez une famille consanguine présentant une détérioration neurologique évoluant vers une déficience intellectuelle grave, une ataxie profonde, une camptocormie et une apraxie oculomotrice [101].

La profusion actuelle de découvertes de nouveaux gènes liés à ces troubles met en évidence leur grande hétérogénéité moléculaire et démontre l'importance des études WES en trio pour le diagnostic et la compréhension fine des mécanismes moléculaires mis en jeu, essentielle pour l'application d'une médecine personnalisée dans la prise en charge des maladies rares.

#### 4. Application du WES pour le diagnostic et le traitement des cancers

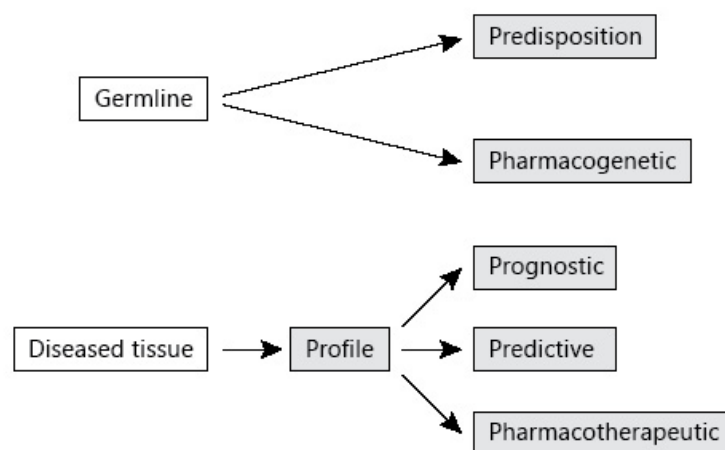
Le cancer est essentiellement une maladie du génome qui évolue et progresse avec l'accumulation de variants somatiques, acquises par l'individu, ou bien des variants de la lignée germinale, c'est à dire transmis de manière héréditaire [102, 103]. En effet la transformation cancéreuse des cellules est étroitement liée aux altérations génétiques de gènes spécifiques, par exemple de proto-oncogènes, de gènes suppresseurs de tumeur et de gènes de réparation de l'ADN. L'explosion des avancées dans le secteur du NGS et des analyses informatiques traitant des données volumineuses ont permis d'analyser de manière exhaustive les profils génomiques du cancer en recherche et en clinique, en utilisant le séquençage ciblé de panels de gènes, le WES, le WGS et le séquençage de l'ARN (ARN-Seq) [104, 105]. À ce jour, plus de 50 000 génomes de tumeurs ont été séquencés et accumulés afin d'explorer les altérations génomiques du cancer, notamment par le Cancer Genome Atlas (TCGA) [103, 106] et le Consortium international sur le

génomique du cancer (ICGC) [107]. D'ici 2030, des centaines de millions de patients atteints de cancer verront leur génome tumoral séquencé [108].

De nombreux variants – principalement dans les régions codantes - ont été découverts dans tous les types de tumeurs humaines courantes et rares. Ces études systématiques de ces données sur le génome du cancer ont révélé de nombreux nouveaux gènes et voies de signalisation oncogéniques [103]. Elles ont notamment identifié des variants des lignées germinales et somatiques récurrentes de plusieurs gènes suppresseurs de tumeurs classiques, tels que *RB1*, *TP53* et *APC*, et une analyse des CNV a révélé une amplification de certains oncogènes et activateurs d'oncogènes sous-jacents, tels que *HER2 / ERBB2* et *MYC* [102, 109].

Certains variants de la lignée germinale et somatique peuvent être utilisés comme biomarqueurs génétiques, en prédisant par exemple l'initiation de la maladie, son évolution ou bien la sensibilité aux traitements. D'autre part certaines altérations génétiques oncogènes peuvent être traitées directement par des thérapies ciblées.

On peut classer les biomarqueurs génétiques exploitables en 6 catégories distinctes, en fonction du type d'informations qu'ils fournissent. On y retrouve les variants de prédisposition, pharmacogénétiques, pronostiques, prédictifs et « pharmacothérapeutiques » désignés comme les «6 P des biomarqueurs génétiques» par Ilyas [110] (Figure 7).



**Figure 7** : Les différents types de biomarqueurs génétiques. D'après Ilyas, 2017 [110].

Ces connaissances peuvent être exploitées pour développer des tests génétiques du cancer basée sur l'identification de biomarqueurs génétiques ayant une valeur pronostique, prédictive, ou thérapeutique et que nous présenterons dans cette partie.

#### 4.1 Variants de la lignée germinale

##### 4.1.1. Les variants de prédisposition

Comme précédemment évoqué, il existe des variants de prédisposition à certains types de cancers dans le génome germinale. Ils confèrent à ceux qui les portent une augmentation du risque de développer un certain type de cancer. Ces gènes sont généralement soit des oncogènes soit des gènes suppresseurs de tumeurs et interviennent dans des voies de signalisation clés pour la transformation des cellules en cellules cancéreuses. Ils varient selon les types de cancers, mais sont souvent impliqués dans des processus clés de la progression tumorale comme les voies de réparation de l'ADN, ou de division cellulaire. Ces variants ont une pénétrance plus ou moins élevée. Chaque cancer a ses spécificités et ses gènes de prédisposition particuliers. Nous prendrons l'exemple du cancer du sein (CS) pour décrire plus en détail les variants constitutionnels de prédisposition au cancer, que l'on peut en général classer en trois catégories (Figure 8):

- Les variants rares à très rares, à haut risque

Des études de liaison et des études de famille atteintes de syndromes de CS héréditaires spécifiques ont déterminé qu'ils étaient causés par des variants constitutionnels monogéniques dans les gènes *BRCA1* et *BRCA2* [111]. Ces syndromes présentent un risque relatif très élevé de cancer (de 5 à 10 fois), mais ils sont rares dans la population en général, représentant moins de 25% des cancers du sein héréditaires [112] et seulement 5% du nombre total de cas de cancer du sein [113].

Des variants monogéniques présentant un risque similaire ont été recherchés au cours des deux décennies qui ont suivi la découverte de *BRCA1 / 2*, sans succès à ce jour. Ces dernières années de nombreuses études ont tenté d'identifier d'autres facteurs génétiques à haut risque et des gènes candidats ont été identifiés par WES. Thompson et al. ont démontré dans l'étude que les variants rares des gènes de réparation de l'ADN, *FANCC* et *BLM*, sont des allèles de susceptibilité potentiels du cancer du sein [114]. Kiiski et al. ont également

découvert que le gène *FANCM* sensibilise au cancer du sein triple négatif en utilisant le WES dans une recherche familiale sur le cancer du sein [115]. D'autre part, Noh et al. ont également découvert que 7 variants dans les gènes *XCR1*, *DLL1*, *TH*, *ACCS*, *SPPL3*, *CCNF* et *SRL* pourraient être associées au cancer du sein par WES dans une étude de familles sans altération des gènes *BRCA* [116].

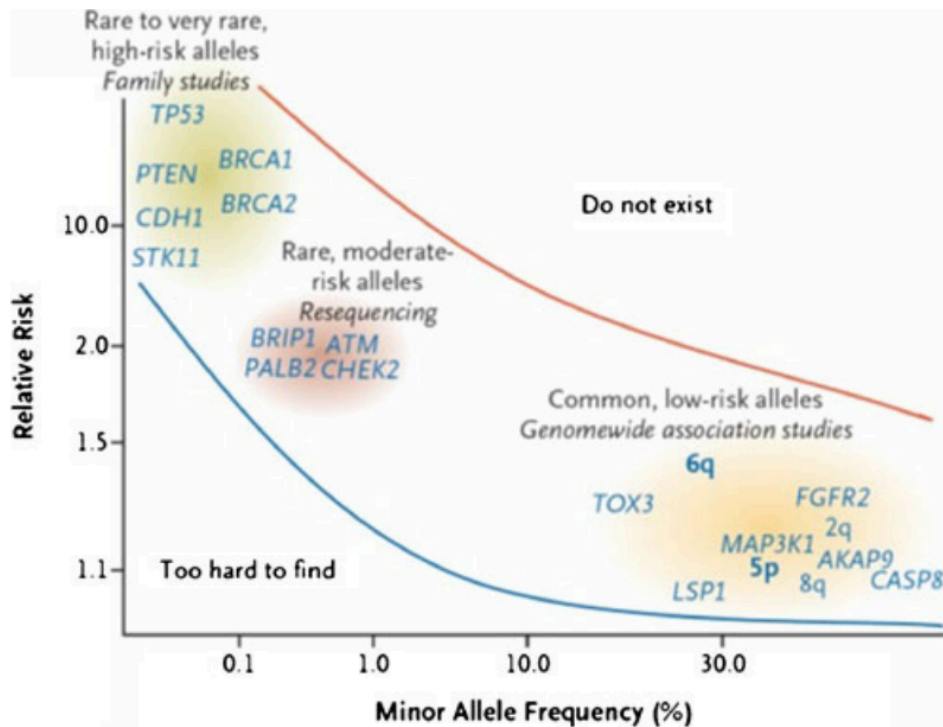
- Les variants rares, de risque intermédiaire

Plus récemment, des allèles à risque modéré ont été identifiés par de grandes études d'association cas-témoins et des études de familles, par séquençage de gènes candidats. Des variants dans des gènes de réparation de l'ADN ont été identifiés, lesquelles confèrent un risque de cancer du sein multiplié par 2. Le variant 1100delC du gène *CHEK2* résulte en une protéine tronquée et confère un risque relatif de 2,3 et un risque absolu de 13% à l'âge de 70 ans [117, 118]. Parmi les autres allèles à risque modéré, citons les variants des gènes *ATM*, *BRIP1* et *PALB2*, qui présentent également des risques relatifs compris entre 2 et 2,5 [117, 119-121] (Figure 8).

Cependant, ces variants à risque modéré sont également relativement rares, laissant 75% des patients présentant un type de cancer familial sans réponse génétique.

- Les variants communs, à faible risque

Pour expliquer ces cas sans réponse génétique, il a été proposé que la susceptibilité au cancer du sein soit en grande partie «polygénique», c'est à dire qu'elle soit conférée par plusieurs loci, chacun ayant un léger effet sur le risque de cancer du sein. La recherche de ces 10 dernières années s'est concentrée sur l'identification par GWAS de variants génétiques relativement communs ayant des effets faibles sur le risque génétique individuellement (risque relatif de 1,1–2), mais pouvant agir de manière multiplicative. De très nombreuses études GWAS et méta-analyses portées sur le génotype et phénotype du CS de centaines de milliers d'individus ont identifié de nombreux variants fréquents conférant un risque faible, comme par exemple dans les gènes *AKAP9*, *ANKLE1*, *MAP3K1*, ou *CASP8* (Figure 8).



**Figure 8** : Les principaux gènes de prédisposition au cancer du sein, regroupés dans un diagramme risque / fréquence. D'après de Buin et al., 2012

D'autres types de cancers ont d'importants facteurs génétiques de prédisposition. Par exemple, le syndrome de Lynch est une affection purement génétique qui favorise la survenue de plusieurs types de cancers à un âge jeune (moins de 50 ans) tels que le cancer du foie, de l'estomac, des ovaires et plus particulièrement du côlon. Les gènes responsables de ce syndrome sont MSH2, MSH6, MLH1 et PMS2, et sont impliqués dans la réparation de l'ADN. D'autres gènes à risque dans le cancer du côlon comprennent par exemple APC, MUTYH ou SMAD4.

#### 4.1.2 Les variants pharmacogénétiques

Il s'agit des variants de la lignée germinale qui permettent de prédire comment un individu répondra à un type de médicament particulier. Cela aura des implications sur le fait de donner ou non le médicament au patient, ainsi que sur la posologie du médicament administré [122-124]. Par exemple, dans le cadre d'un cancer diagnostiqué, des variants de la lignée germinale dans les gènes de réparation par recombinaison homologue (RH) comme *BRCA1*, *BRCA2*, *ATM*, *CHEK2*, ou *PALB2*

suggèrent une susceptibilité à des essais de chimiothérapie aux sels de platine et aux inhibiteurs de PARP [125]. De la même manière, une instabilité microsatellite ou un défaut dans la réparation de l'ADN causés par des variants de *MSH2*, *MSH6*, *MLH1* ou *PMS2* peuvent suggérer une vulnérabilité aux inhibiteurs de PD-1 [126].

#### 4.2 Les variants de la lignée somatique

Ce type de variant fait référence aux variants identifiés uniquement dans les tissus malades. Dans les tumeurs, différents variants peuvent être détectés et peuvent être utilisés pour établir un profil génétique des tumeurs et de les classer - à la fois en complément de la classification morphologique et en tant qu'alternative à la classification morphologique. Ils représentent leur signature moléculaire et sont révélateurs des mécanismes biologiques qui sous-tendent la progression tumorale et les potentielles résistances acquises aux traitements.

Certains variants sont caractéristiques de certaines tumeurs et peuvent donc être utilisés pour confirmer un diagnostic voire le préciser [127]. Ils peuvent receler des informations pronostiques, prédictifs ou thérapeutiques.

##### 4.2.1 Les variants pronostiques

Les variants pronostiques visent à prédire l'évolution de la maladie. Outre la classification habituelle des tumeurs en anatomopathologie selon leur grade et leur potentiel invasif, qu'on associe automatiquement à un pronostic plus ou moins favorable, l'étude du profil génétique des tumeurs peut constituer également un facteur pronostique important. En effet la présence de certains variants somatiques est capable de pronostiquer l'évolution du cancer quant au risque de rechutes, et de survie globale à moyen ou long terme [128].

Par exemple dans les adénocarcinomes canauxaires pancréatiques, le variant délétère G12V du gène *KRAS* est associé à un mauvais pronostic et une survie globale abaissée [129]. Par ailleurs, de nombreuses études ont confirmé l'association de variants de *TP53* avec des tumeurs plus agressives et un pronostic globalement plus mauvais dans divers types de cancer [130-136]. À l'inverse, des variants somatiques peuvent être prédictif d'un meilleur pronostic, comme ceux dans les gènes régulateurs du cycle cellulaire dans les cancers du sein triple-négatifs [137].



De plus, le séquençage de l'ADN circulant peut également être considéré comme un outil pronostique précieux. La caractérisation moléculaire et de variants oncogènes dans l'ADN circulant représente un moyen non invasif prometteur de prédire la réponse au traitement et les rechutes dans différents types de cancers [138, 139].

#### 4.2.2 Les variants prédictifs

Les variants prédictifs visent à prédire la réponse du patient à un traitement donné [140]. Avec les données provenant d'un nombre croissant d'études corrélatives disponibles établissant un lien entre les profils mutationnels et les résultats des traitements, la valeur prédictive des données d'un test basé sur le NGS continue à augmenter.

Tout d'abord les technologies de NGS ont permis d'identifier des variants permettant de prédire la réponse aux chimiothérapies communément utilisées. Étant donné les effets secondaires de ce traitement à court et à long terme, il est important de déterminer quels sont les patients à traiter et les patients pour lesquels un traitement est peu susceptible de générer un bénéfice significatif. Ces traitements sont composés d'agents cytotoxiques comme la cisplatine qui induisent des dommages à l'ADN. D'une façon générale, les variants affectant les gènes de réparation de l'ADN prédisent une sensibilité accrue des cellules cancéreuses aux chimiothérapies, en les empêchant de résister au traitement. Ainsi il a été montré que des altérations des gènes *ATM*, *RB1*, *FANCC*, et *ERCC2* impliqués dans la réparation de l'ADN prédisent une réponse et un bénéfice clinique de la chimiothérapie par cisplatine dans les cancers de la vessie [141, 142]. De la même manière, dans les cancers de l'ovaire, un défaut de réparation de l'ADN par l'inactivation de gènes impliqués dans la recombinaison homologe (*BRCA1/2*, gènes *RAD*, *CDK12*) [143-145], et dans la réparation par excision de bases comme *ERCC1* [146], sensibilise les cellules cancéreuses à la chimiothérapie aux sels de platine. Des variants somatiques dans les gènes *RPA1* et *DNAH* sont également associés à une meilleure réponse aux chimiothérapies dans différents types de cancer [147, 148]. À l'inverse, d'autres variants génétiques sont des marqueurs d'une résistance à la chimiothérapie, telles que des mécanismes de stabilisations de la fourche de réplication induits par des pertes de *PTIP*, *PARP1* et *CHD4*, ou la restauration de recombinaison homologue dans des cellules déficientes via la réversion de variants dans les gènes *BRCA1/2*

[149, 150]. De plus, il a été montré que des altérations des gènes de régulation du cycle cellulaire et des points de contrôle immunitaires sont des marqueurs de résistance à la chimiothérapie retrouvés dans des tumeurs post-traitement, suggérant une évolution génomique dynamique d'acquisition de la résistance au cours du traitement [151]. Certains variants peuvent également prédire une toxicité de la chimiothérapie, marquée par de lourds effets secondaires, par exemple chez les patients porteurs d'un variant bi-allélique dans le gène *FANCM* [152].

D'autre part, l'identification des variants somatiques d'une tumeur par WES permet d'indiquer la pertinence de l'utilisation d'un traitement par immunothérapie, basée sur l'inhibition du point de contrôle immunitaire par des anticorps anti-programmed cell death protein 1 (PD-1), anti-PD-L1, et anti-CTLA-4. Rizvi et al. ont démontré qu'une augmentation de la charge mutationnelle, définie par le nombre total de variants ponctuels non-synonymes dans la tumeur, est associée à une bonne réponse aux immunothérapies, ainsi qu'un bénéfice clinique durable et une meilleure survie globale [153, 154]. Une charge mutationnelle élevée peut également indiquer la présence de défauts des voies de réparation d'ADN, notamment des gènes *BRCA1/2*) hérités ou acquis somatiquement, qui peuvent être particulièrement sensibles aux inhibiteurs de PARP comme l'*olaparib*. Ces données mettent en lumière l'importance des interactions mutation-traitement.

Plusieurs tests génomiques permettant d'évaluer le pronostic des cancers (risque de rechutes) et / ou de prédire la réponse à un traitement adjuvant ont été mis au point en utilisant les dernières données des études corrélatives. Les tests disponibles comprennent le *test du score de récurrence de 21 gènes* par Oncotype DX et les *70 gènes signatures* par MammaPrintR, qui sont avérés comme utiles cliniquement et recommandés par l'association américaine d'oncologie clinique pour la prise de décision thérapeutique [155, 156].

#### *4.2.3 Les variants pharmacothérapeutiques: identifier les cibles thérapeutiques*

Des études de séquençage de tumeurs à grande échelle ont révélé l'identité de nouvelles aberrations génomiques responsables de la croissance et de la progression tumorales, appelées variants *driver*. De plus en plus, ces variants *drivers* peuvent être inhibées cliniquement à l'aide d'un répertoire croissant de thérapies à

ciblage moléculaire [157]. La médecine de précision en oncologie est entrée dans une nouvelle ère il y a près de 20 ans avec le développement et l'approbation de deux traitements ciblés: le trastuzumab pour le traitement du cancer du sein métastatique avec amplification de HER2 et l'imatinib pour le traitement de patients atteints de leucémie myéloïde chronique avec fusion BCR-ABL [158, 159]. Actuellement, les thérapies basées sur le génotype et approuvées par la *Food and Drugs Administration* (FDA) sont disponibles pour de nombreux types de tumeurs (Tableau 2). Nous verrons plus en détail les marqueurs génomiques les plus étudiés, tels que EGFR, BRAF et les CDK.

| <b>Cible moléculaire</b> | <b>Thérapie ciblée</b> | <b>Type de cancer</b>                                     | <b>Variant spécifique approuvé pour la sélection des patients</b> | <b>Références</b> |
|--------------------------|------------------------|---|---|-------------------|
| ALK                      | Alectinib              | CPNPC   | <i>Variant ALK</i>  | [160]             |
|                          | Brigatinib             | CPNPC   | <i>Variant ALK</i>  | [161]             |
|                          | Ceritinib              | CPNPC   | <i>Variant ALK</i>  | [162]             |
| ALK, MET, ROS1           | Crizotinib             | CPNPC   | <i>Fusion de ALK</i> ou <i>fusion de ROS1</i> fusion              | [163]             |
| <i>BCR-ABL1</i>          | Bosutinib              | Leucémie myéloïde chronique (LMC)                         | <i>Fusion de BCR-ABL1</i>   | [164]             |
|                          | Dasatinib              | LMC   | <i>Fusion de BCR-ABL1</i>   | [165]             |
|                          | Imatinib               | LMC   | <i>Fusion de BCR-ABL1</i>   | [166, 167]        |
|                          | Nilotinib              | LMC   | <i>Fusion de BCR-ABL1</i>   | [168]             |
|                          | Ponatinib              | LMC   | <i>Fusion de BCR-ABL1</i>   | [169]             |
| BRAF                     | Dabrafenib             | CPNPC, mélanome, cancer de la thyroïde anaplastique       | <i>BRAF V600/K</i>  | [170-172]         |
| CDK4/6                   | Abemaciclib            | Cancer du sein HR+ HER2-                                  |   | [173]             |
|                          | Palbociclib            | Cancer du sein HR+ HER2-                                  |   | [174]             |
|                          | Ribociclib             | Cancer du sein HR+ HER2-                                  |   | [175, 176]        |
| EGFR                     | Cetuximab              | Cancer colorectal, cancer de la tête et du cou (squameux) | Exprime EGFR et non muté pour KRAS (cancer colorectal)            | [177]             |
|                          | Erlotinib              | CPNPC, cancer du pancréas                                 | <i>EGFR</i> (délétion de l'exon 19/L858R) pour CPNPC              | [178-180]         |
|                          | Gefitinib              | CPNPC   | <i>EGFR</i> (délétion de exon 19/L858R)                           | [181]             |
|                          | Necitumumab            | CPNPC (squameux)  |   | [182]             |
|                          | Osimertinib            | CPNPC   | <i>EGFR</i> (T790M/exon 19 deletions/L858R)                       | [183-185]         |
|                          | Panitumumab            | Cancer colorectal   | <i>Non muté pour KRAS et NRAS</i>                                 | [186, 187]        |

|            |                           |  |   |            |
|------------|---------------------------|--|---|------------|
| EGFR/ERBB2 | Afatinib                  | CPNPC  | <i>EGFR</i> (exon 19 deletions/L858R/S768I/L861Q/G719X) | [188]      |
| ERBB2      | Ado-Trastuzumab Emtansine | Cancer du sein HER2+   |   | [189, 190] |
|            | Trastuzumab               | Cancer du sein HER2+<br>Cancer de l'estomac HER2+  |   | [191-193]  |
|            | Pertuzumab                | Cancer du sein HER2+   |   | [194]      |
|            | Lapatinib                 | Cancer du sein HER2+   |   | [195]      |
|            | Neratinib                 | Cancer du sein HER2+   |   | [196, 197] |
| KIT        | Imatinib                  | Mastocytose agressive systémique   | Absence du variant D816V dans <i>c-Kit</i>              | [198]      |
|            |                           | Tumeurs gastrointestinales stromales   | <i>Kit</i> (CD117) positif                              | [199]      |
| MEK        | Trametinib                | CPNPC, mélanome, cancer de la thyroïde anaplastique  | <i>BRAF</i> V600E/K                                     | [170, 171] |
| mTOR       | Everolimus                | Cancer du sein HR+ HER2-; carcinome rénal; Tumeur neuroendocrine du pancréas, gastro-intestinale, ou du poumon |   | [200-204]  |
| PDGFR      | Imatinib                  | Trouble myelodysplasique/myelo prolifératifs   | Réarrangement de <i>PDGFR</i>                           | [205]      |
|            |                           | Dermatofibrosarcome  | <i>COL1A1-PDGFB</i> fusion                              | [206]      |

**Tableau 2** : Thérapies ciblées approuvées par la FDA, d'après Chan et al., 2019.

#### a. Récepteur au facteur de croissance épidermal (EGFR)

C'est dans l'utilisation des inhibiteurs de l'EGFR dans les traitements du cancer du poumon non à petites cellules (CPNPC) que les bénéfices des thérapies ciblées personnalisées basées sur la génomique est le plus évident par rapport aux traitements de chimiothérapie classiques. Les variants somatiques de l'*EGFR* sont parmi les cibles les plus étudiées en raison de leur prévalence élevée (15 à 50%) chez les patients atteints de CPNPC. Les deux variants les plus courants de l'*EGFR* sont les délétions de l'exon 19 et la substitution -L858R de l'exon 20, représentant ensemble jusqu'à 90% des variants de l'*EGFR*. Ces altérations résultent en une tyrosine kinase mutée hyper-phosphorylée, qui active de manière constitutive les

signaux en aval (RAS / MAPK et P13K / Akt), conduisant à la tumorigenèse [207]. Il a été démontré que ces deux variants prédisent une réponse aux inhibiteurs de tyrosine kinase (ITK) chez les patients atteints de CPNPC [208, 209]. Les données de plusieurs essais cliniques ont montré que jusqu'à 67% des individus atteints de CPNPC porteurs d'un variant sensibilisant ont répondu à l'erlotinib ou au gefitinib (ITK de première génération) [210]. Les ITK se sont montrés plus bénéfiques que la chimiothérapie traditionnelle en terme de taux de réponse et de survie globale, uniquement chez les porteurs d'un variant sensibilisant dans le gène *EGFR* [211, 212].

Par ailleurs, chez les patients atteints d'un cancer colorectal métastatique, des thérapies dirigées contre l'*EGFR*, telles que le cetuximab et le panitumumab, sont spécifiquement administrées à des patients ne présentant aucun variant oncogénique dans *KRAS* et *NRAS* [213].

#### *b. Protéine kinase sérine-thréonine et mitogen-activated protein kinase*

Environ la moitié des mélanomes avancés sont porteurs d'une altération de *BRAF* de la protéine kinase sérine – thréonine, V600E étant le variant délétère le plus fréquent (90% des mutations totales de *BRAF*). La protéine *BRAF* mutée phosphoryle et active les protéines MEK, ce qui active à son tour la voie MAPK, conduisant à une croissance cellulaire et à une prolifération incontrôlées [214]. La forte prévalence de l'altération de *BRAF* dans le mélanome en fait un candidat idéal pour des thérapies ciblées. Le dabrafenib et le vémurafénib, deux inhibiteurs de *BRAF*, ont tous deux montré un taux de survie amélioré par rapport au traitement par chimiothérapie classiquement utilisé chez les patients atteints d'un mélanome métastatique *BRAF*-V600 [215, 216]. Cependant, la majorité des patients ayant reçu une monothérapie avec l'inhibiteur de *BRAF* ont développé une résistance dans les mois suivant le traitement due à une réactivation de la voie MAPK par une voie alternative induite par un second variant [217]. Il a été démontré récemment que l'emploi d'une thérapie combinée de deux agents distincts diminue les résistances au traitement et augmente son efficacité [170-172].

#### *c. Cyclin-dependant kinase 4/6*

L'importance des kinases 4/6 dépendantes des cyclines (CDK4 et CDK6, respectivement) dans la division cellulaire et leur hyperactivation dans plusieurs types de cancer a été bien établie. De nombreuses études ont été menées pour développer des médicaments ciblant l'activité des CDK, et trois d'entre eux sont maintenant approuvés pour le cancer du sein ER + avancé ou métastatique. L'ajout d'un inhibiteur de CDK4 / 6 (ribociclib, palbociclib et abemaciclib) à l'hormonothérapie a permis d'augmenter la survie sans progression de la maladie par rapport aux traitements classiques [173, 176, 218]. Malgré leur efficacité supérieure dans les traitements contre le cancer, les déterminants de la sensibilité des inhibiteurs de CDK4 / 6 sont encore inconnus.

Le nombre de thérapies ciblées progresse, et de nombreux essais cliniques sont en cours pour tester l'efficacité de nouveaux médicaments. Le nombre de thérapies ciblées devrait donc normalement augmenter, et permettre à un plus grand nombre de personnes de profiter de traitements personnalisés en fonction du profil génétique de chaque tumeur. Pour ce faire, le développement du NGS est essentiel, pour identifier le plus rapidement et efficacement les modifications génomiques pouvant être ciblées dans tous les types de tumeurs.

#### 4.3 Essais cliniques prospectifs et randomisés sur différents types de tumeurs

La première étude randomisée en oncologie de précision dans le cancer avancé a été l'essai SHIVA (Institut Curie), qui a évalué les bénéfices des thérapie moléculaire ciblée basée sur le profil moléculaire de la tumeur par rapport aux traitements conventionnels [219]. Le critère d'évaluation de l'étude était la survie sans aggravation. Cette étude n'a pas identifié de différence de survie globale entre les deux groupes [219]. Toutefois celle-ci n'incluait que des thérapies ciblées disponibles en France en 2011 et comportait plusieurs autres limitations.

D'autre part, l'étude IMPACT (Initiative for Molecular Profiling and Advanced Cancer Therapy) développée au MD Anderson [220, 221] a étudié de manière rétrospective le profil moléculaire de patients avec un cancer avancé. Elle a montré que le profil moléculaire de la tumeur et la sélection d'un traitement ciblé correspondant sont associés à des taux de réponse, de survie sans aggravation et de survie globale plus élevés par rapport aux patients non traités avec le traitement ciblé correspondant.

Suite à ces résultats encourageants, l'essai MATCH (Molecular Analysis for Therapy Choice) a été lancé afin de tenter de faire correspondre des anomalies moléculaires à des agents ciblant des anomalies spécifiques. Les résultats de cette étude, basée sur 6000 individus, ont fait l'objet de plusieurs articles décrivant des associations entre des altérations moléculaires et les réponses aux thérapies ciblées. L'utilisation de l'inhibiteur de la PI3-kinase, le tselisib, a été associée à une survie sans aggravation  $\geq 6$  mois chez 27% des patients présentant des variants activateurs du gène *PIK3CA* [222]. L'utilisation d'adotrastuzumab emtansine chez les patients atteints de tumeurs sur-exprimant HER2, à l'exclusion des CS, a entraîné un taux de réponse partielle de 8,1% (3 sur 37) et un taux de maladie stable de 43% (17 sur 37; durée médiane, 4,6 mois) [223]. Le taux de survie à 6 mois était de 24,8%. Le taux de réponse partielle chez les patients présentant des altérations de la voie du récepteur du facteur de croissance des fibroblastes (FGFR) traités avec l'inhibiteur sélectif du FGFR AZD4547 (N = 41) était de 5% et le taux de maladie stable de 51% [224]. Le taux de survie sans aggravation à 6 mois était de 17%. Enfin, l'utilisation de l'inhibiteur de PD1 (nivolumab) a abouti à une réponse objective de 24% (8 sur 33) chez les patients atteints de cancers non colorectaux réfractaires et comportant un défaut de réparation des mésappariements de l'ADN.

L'objectif ultime est de pouvoir utiliser chez tous les patients, au moment du diagnostic, leur profil génomique tumoral, leurs marqueurs immunitaires et d'autres caractéristiques biologiques pour déterminer le traitement optimal. Il est nécessaire d'envisager la conception d'autres essais cliniques innovants intégrant les changements biologiques dynamiques et la complexité du cancer.

## 5. Limitations du NGS

Bien que l'on souhaite utiliser le NGS comme méthode unique pour détecter toutes les variations génétiques cliniquement pertinentes, des limitations importantes existent actuellement. Ces limitations concernent la sensibilité analytique de la détection des variants, les zones du génome difficiles à séquencer ou à analyser, la détection des variants structuraux et du nombre de copies, et surtout la validation de causalité et l'intégration des données génomiques aux soins médicaux. Ces limitations seront discutées plus en détail dans la partie suivante.

### *5.1 Sensibilité analytique*

Le NGS peut détecter des SNV dont la fréquence allélique est au minimum présente dans 5% à 10% des cellules testées [225-227]. Bien que cette sensibilité soit suffisante pour la plupart des tests de maladies héréditaires, elle limite les tests en oncologie pour une maladie résiduelle, lorsqu'un faible pourcentage de tissu tumoral est présent, ou pour détecter un variant faiblement présent du à l'hétérogénéité intrinsèque de la tumeur. Les raisons possibles de cette sensibilité limitée sont notamment l'utilisation d'échantillons provenant de tissus fixés au formol et inclus dans la paraffine (FFPE) et les erreurs de séquençage systématiques [228].

En effet, les échantillons FFPE produisent des artefacts plus importants que les échantillons de tissu frais. De plus, des études ont montré que les erreurs systématiques conduisent à un taux d'erreur de 4% à 6%; elles peuvent être des erreurs spécifiques à une séquence, des erreurs à un emplacement particulier de la lecture (par exemple, les extrémités pour les séquenceurs Illumina), ou liées au contenu en nucléotides de la séquence (zones riches en GC pour Illumina) [229, 230]. Etant donné que ni la PCR ni le séquençage ne provoquent d'insertions / délétions (indels), en dehors des régions répétées, la détection des petits indels est moins sujette aux erreurs systématiques que celle des SNV [231].

### *5.2 Les régions difficiles à séquencer*

Les régions homologues, les régions répétitives et les régions riches en GC ne peuvent pas être interprétées de manière fiable par les plates-formes NGS actuelles et les algorithmes bio-informatiques standard. Les régions homologues, y compris les pseudogènes, sont des zones du génome présentant une similarité de séquence élevée qui peuvent différer du gène d'intérêt par seulement quelques paires de bases. Les fragments d'ADN séquencés à partir du gène cible et des régions homologues peuvent avoir une séquence presque similaire qu'il est impossible de les distinguer.

Dans l'analyse des données de NGS, les fragments d'ADN d'un gène cible et de ses régions homologues ont souvent une mauvaise qualité d'assemblage [232]. Un mauvais alignement peut conduire à la fois à la détection de faux positifs et de faux négatifs. Cela peut avoir des conséquences sur les tests médicaux, car de nombreux



gènes cliniquement pertinents, tels que *PMS2* pour le syndrome de Lynch, possèdent des pseudogènes, et sont donc difficiles à interpréter par le NGS [232, 233]. Ce problème pourrait être résolu dans le futur par des techniques de NGS dits de troisième génération plus récentes, permettant le séquençage de lectures beaucoup plus longues.

D'autre part, les régions riches en GC semblent présenter un bruit de fond plus important et une qualité de séquençage inférieure. En particulier, les séquenceurs Illumina provoquent des erreurs de substitution dans les zones à haute teneur en GC et les homopolymères longs G / C [230].

### *5.3 Limitations des bases de données et des connaissances*

Bien qu'il soit techniquement possible d'effectuer une analyse du génome entier, en particulier pour les maladies héréditaires, les connaissances actuelles ne permettent pas d'interpréter tous les variants et la majorité d'entre eux sont considérés comme des VUS, notamment dans les régions non codantes. Les outils pour aider à l'interprétation des variants comprennent les bases de données (publiques, privées ou spécifiques à un laboratoire), les outils de prédiction bioinformatiques, les connaissances génétiques et médicales, la littérature médicale, le contexte clinique et l'expérience clinique.

Bien que les bases de données soient indispensables pour interpréter les variants, elles restent pour l'instant incomplètes et peuvent présenter des erreurs dues à une mauvaise qualité de séquençage. De plus, elles peuvent ne pas être à jour ou présenter des données contradictoires [234]. La littérature médicale et les bases de données doivent donc être utilisées avec précaution.

L'importance des variants dans les introns ou les régions non traduites est souvent inconnue, et les variants exoniques rares ou nouveaux peuvent également être difficiles à interpréter. Des variants rares qui entraînent un décalage du cadre ou qui transforment un acide aminé en un codon stop seront généralement pathogènes. En revanche, les variants faux-sens sont plus difficiles à interpréter. De nombreux facteurs et prédictions *in silico* sont pris en compte lors de l'interprétation de ces variants [23].

### *5.4 Variations structurales et variations du nombre de copies*

Le NGS identifie relativement bien les SNVs et les petits indels, mais détermine moins bien la présence de grands réarrangements structuraux et les CNVs, surtout dans le cas où une capture est effectuée pour réaliser un séquençage ciblé [235]. À l'heure actuelle, les laboratoires cliniques utilisent une deuxième technique complémentaire comme les puce d'hybridation génomique comparative ou *Comparative Genomic Hybridization Array* (CGH array) pour détecter ce type de variants.

Il est toutefois possible de détecter les CNVs à partir des données de NGS en utilisant les variations de profondeur et d'en déduire la délétion ou duplication d'une région génomique. La détection de ces grandes modifications requiert des algorithmes bioinformatiques différents de ceux utilisés pour la détection des SNVs. Ces outils analysent les différents niveaux de profondeur de séquençage pour détecter les duplications / amplifications ou les suppressions [22, 236].

# Chapitre 2 : Apports du séquençage d'exome pour le diagnostic de maladies rares

## 1. Introduction

Les maladies génétiques sont une des principales causes de décès chez les enfants de moins de dix ans [237-240]. Établir une étiologie diagnostique chez les enfants suspectés de maladies génétiques est important pour la mise en œuvre rapide de la médecine de précision, en particulier pour guider les décisions cliniques importantes telles que les choix thérapeutiques, les chirurgies et les soins palliatifs [241]. Le diagnostic étiologique nécessite l'identification de la base moléculaire causale. Dans la pratique, cela est difficile pour plusieurs raisons. Tout d'abord, comme nous l'avons vu précédemment, il existe environ 6000 troubles génétiques pour lesquels la base moléculaire a été établie (OMIM). Il existe donc une importante hétérogénéité génétique avec de très nombreux gènes potentiellement impliqués. D'autre part, les présentations de maladies génétiques chez les jeunes enfants sont souvent approximatives et imprécises induisant une forte hétérogénéité clinique. De plus, la progression de la maladie est plus rapide chez les enfants, transformant la recherche diagnostique en une course contre la montre [242].

Actuellement, l'établissement des diagnostics moléculaires consiste généralement en une combinaison d'un CGH array et d'un séquençage ciblé de panels de gènes. Ces tests sont employés en première ligne pour les enfants atteints de maladies génétiques afin de guider le diagnostic différentiel [243]. De nombreux autres tests de routine sont également fréquemment utilisés, notamment, les tests métaboliques, de cytogénétique et d'hybridation in situ par fluorescence chromosomique [244]. Ces stratégies conventionnelles par étapes peuvent être coûteuses et longues et néanmoins ne fournissent un diagnostic moléculaire que pour une petite fraction des patients.

Le séquençage ciblé de panel est une approche fondée sur des hypothèses cliniques qui peuvent amener les médecins à se limiter sur un organe en particulier ou à un composant phénotype spécifique et se restreindre à des panels de gènes qui peuvent ou non refléter le diagnostic différentiel complet. Ainsi, la prise en charge

hospitalière des enfants chez lesquels des maladies génétiques sont suspectées reste largement empirique, sur la base des diagnostics cliniques [241].

Au total, il est estimé que le taux de diagnostic par séquençage ciblé de panel toutes pathologies confondues est de 24% en moyenne [245] et varie en fonction des types de pathologies. Par exemple, pour les troubles du mouvement, il n'est que de 11 à 22% [246, 247], et avoisine les 29% pour les troubles neuro-développementaux [248]. Ces faibles taux de diagnostic entraînent une longue errance diagnostique pour les patients et leurs familles. C'est pourquoi nous nous sommes donné l'objectif, au travers d'une étude clinique, de déterminer si le séquençage de l'exome peut être considéré comme une alternative plus efficace.

Au cours des cinq dernières années, le WGS et le WES ont commencé à être largement utilisés pour le diagnostic génétique des nourrissons et des enfants présentant des suspicions de maladies génétiques [249, 250]. En permettant l'examen concomitant de tous ou de la plupart des gènes, le WGS et le WES ont le potentiel d'augmenter le rendement diagnostique chez les personnes soupçonnées de troubles génétiques. Bien que les méthodes WGS et WES se soient considérablement améliorées au cours de cette période, il n'y a pas encore de directives pour leur utilisation par les cliniciens et très peu de laboratoires médicaux l'utilisent en première ligne.

Il existe donc un réel besoin de déterminer si l'utilisation précoce du séquençage à l'échelle du génome est digne d'intérêt pour le diagnostic de maladies génétiques. Dans le cadre de l'activité R&D de l'entreprise Prenostics, nous avons étudié de façon prospective une cohorte d'enfants âgés de 0 à 18 ans présentant une suspicion de maladies génétiques et avons comparé les performances du séquençage ciblé de panel (techniques de routine actuellement utilisée à l'hôpital), au séquençage d'exome.

## 2. Matériel et méthodes

### 2.1. Échantillon et cohortes utilisées

Nous avons eu accès de manière prospective à des échantillons sanguins d'une cohorte de 25 enfants atteints de maladies génétiques rares sans diagnostic préalable en provenance du Centre de Génétique Chromosomique de l'Unité de

Diagnostic Prénatal de l'hôpital américain de Paris, dont la particularité est que la recherche de CNV par CGH array a été non contributive.

Les enfants malades et leurs parents ont été séquencés dans le cadre de leur prise en charge clinique par l'entreprise Prenostics. Les jeunes individus inclus dans cette étude présentent des symptômes cliniques divers que nous avons pu classer en 4 groupes : maladies neurodéveloppementales (7 individus), trouble du spectre autistique avec déficit intellectuel (4 individus), malformations congénitales (7 individus) et troubles du mouvement (8 individus). Les données cliniques sur les patients incluant le sexe, l'âge et les signes cliniques sont présentées dans les tableaux 5, 6, 7 et 8.

Les parents ont donné leur consentement signé pour l'investigation d'un séquençage clinique d'exome pour leur enfant et eux-mêmes.

## 2.2 Extraction d'ADN

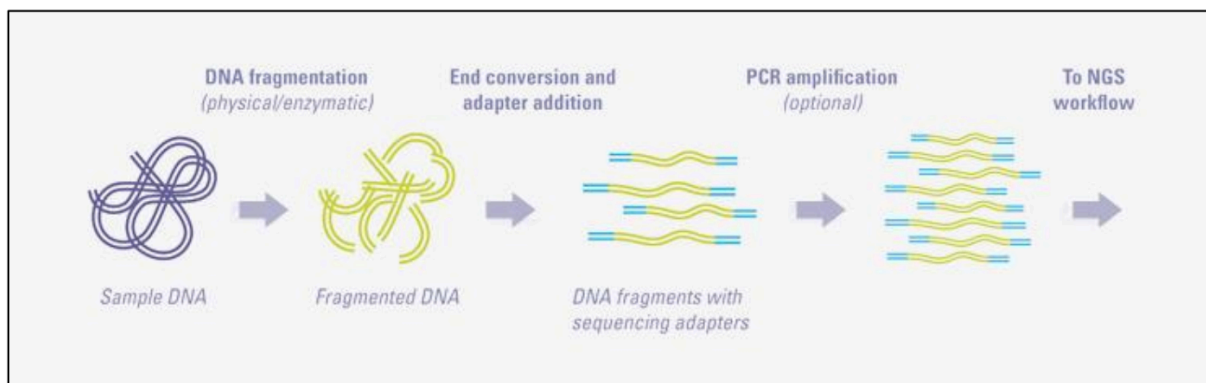
L'ADN génomique a été isolé du sang périphérique en utilisant le kit Dneasy Blood and tissue (Qiagen, Hilden, Germany). L'ADN génomique des tissus FFPE a été extrait en utilisant le kit QIAamp DNA tissue kit (Qiagen, Hilden, Germany). Les échantillons ont été quantifiés à l'aide du kit de test haute sensibilité Qubit dsDNA (ThermoFisher Scientific, USA) afin de déterminer les concentrations d'ADN extraits qui seront utilisés pour la réalisation de la librairie.

## 2.3. Séquençage de l'exome

### 2.3.1 Librairie

Nous avons utilisé 1µg d'ADN de chacun des échantillons d'ADN extrait. Ces ADN ont été fragmentés à l'aide d'un appareil de sonication Covaris E220 de la société Covaris (Woburn, USA), pour obtenir des fragments de taille compatible avec le séquençage d'environ 200 bases. Les extrémités soniquées des fragments obtenus ont ensuite été réparées grâce au kit HTP library preparation kit de la société KAPA (Bâle, Suisse), pour permettre ensuite leur ligation aux adaptateurs de séquençage Illumina. Cette étape, effectuée sur un automate Hamilton STARL et de la société Hamilton (Reno, USA), est nécessaire à la fixation des fragments à séquencer sur la *flowcell* ou puce de séquençage. Les échantillons ont ensuite été amplifiés par

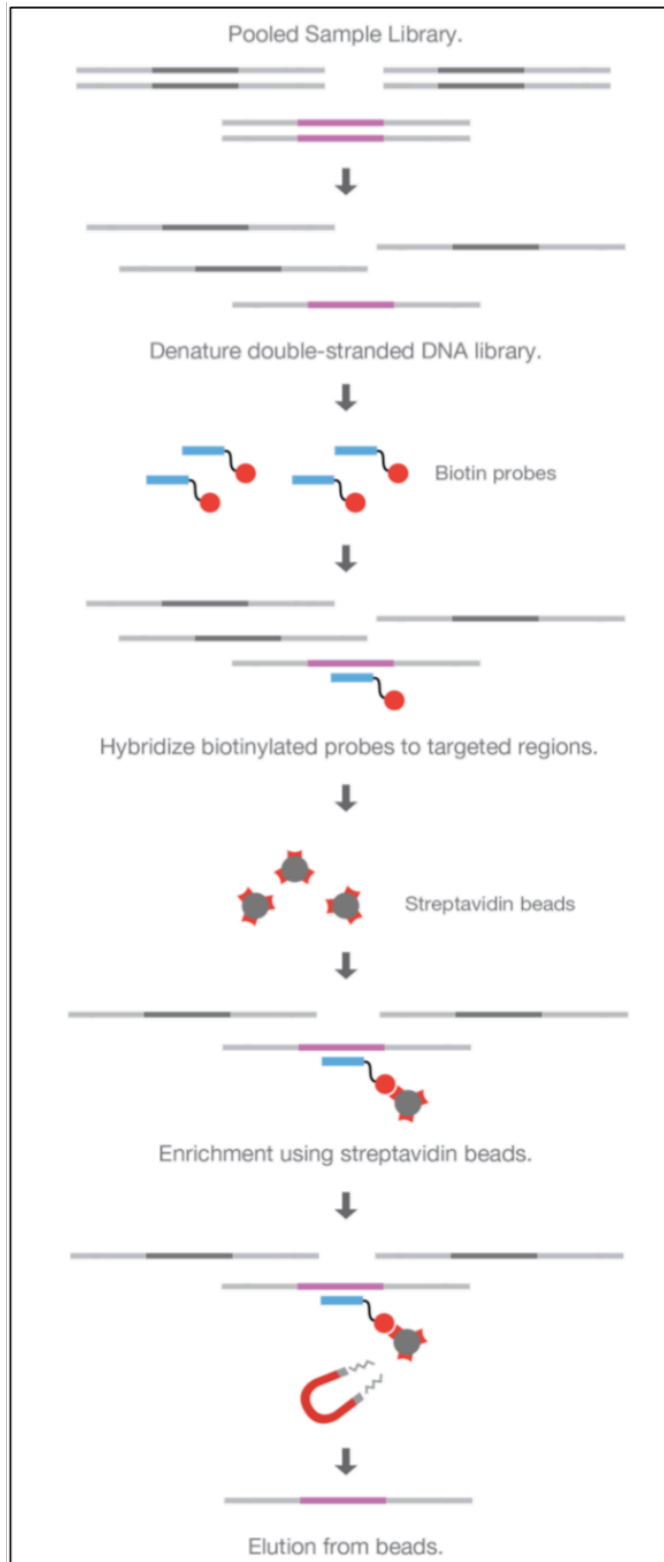
réaction en chaîne par polymérase, ou PCR, en utilisant les 91 amorces complémentaires des adaptateurs de séquençage. Les échantillons amplifiés ont ensuite été purifiés lors d'une étape de sélection de taille pour ne garder que ceux compatibles avec le séquençage d'environ 200 bases, puis quantifiés par l'appareil LabChip GX de Perkin Elmer (Wellesley, USA) (Figure 9).



**Figure 9:** Schéma des étapes de la préparation d'une librairie d'ADN pour séquençage de nouvelle génération

### 2.3.2. Enrichissement

À la suite de ces étapes, les librairies ont été réparties en pools de 5 échantillons constitués de 1 µg d'ADN chacun pour procéder à l'étape d'enrichissement. Lors de cette étape, les pools d'échantillons ont été mis en présence des sondes biotinylées correspondant à la capture de l'exome pendant 72h à 47°C (capture NimbleGen SeqCap EZ MedExome Enrichment, Roche). Les pools d'échantillons capturés ont ensuite été purifiés à l'aide des kits de lavage *SeqCap Hybridization* et *Wash Kit* de Roche (Bâle, Suisse) sur l'automate *Bravo Automated Liquid Handling Platform* de la compagnie Agilent (Santa Clara, USA), afin d'éliminer tous les fragments non capturés. Ces captures ont ensuite été amplifiées grâce au kit *Hifi HotStart ReadyMix* de la société Nimblegen et quantifiées à l'aide des appareils *Qubit fluorometric quantitation assays* (Thermo Fisher Scientific, Waltham, USA) l'appareil *Caliper LifeScience Labchip GX* (Perkin Elmer, Waltham, USA) (Figure 10).



**Figure 10** : Schéma des différentes étapes du procédé de capture (D'après Illumina.com)

### 2.3.3. Séquençage

Les échantillons ont été séquençés sur les séquenceurs HiSeq 4000 ou NovaSeq d'Illumina (San Diego, Californie, États-Unis), en poolant 5 échantillons par ligne sur les *flowcells*, et en utilisant le protocole de 2x150 bp reads suivant les indications du fournisseur.

### 2.4. Analyse bioinformatique

Les reads séquençés ont été démultiplexés, c'est-à-dire réattribués à un échantillon, à l'aide du logiciel bcl2fastq (Illumina, version 2.19.1), générant des fichiers fastq pour chaque échantillon. Ces fastq ont ensuite été alignés sur le génome en version hg38, en utilisant le logiciel BWA (BurrowWheeler Aligner, version 0.7.15) [251]. En moyenne, 84 millions de reads ont été alignés pour chacun des échantillons, dont 94,2%  $\pm$ 0,8% ont été alignés correctement sur le génome de référence. L'étape suivante, celle de la recherche des variants, a été réalisée avec le logiciel GATK (Genome Analysis ToolKit, version 3.7) [252]. Tous les SNPs ainsi que les petites insertions et délétions référencés ont été annotés par l'outil Ensembl Perl Application Program Interfaces (version 89) ainsi que par des scripts Perl développés au laboratoire pour inclure les bases de données dbSNP (version 135) et dbNSFP (annotation des SNV non-synonymes) (version 3.4). Ces bases de données fournissent également une prédiction fonctionnelle des variants non synonymes et des sites d'épissage identifiés [253].

Le fichier VCF combiné a ensuite été téléchargé et analysé avec le logiciel Ingenuity Variant Analysis (IGV) de Qiagen.

### 2.5. Analyse et interprétation des variants

Pour l'interprétation des variants ponctuels, nous nous sommes basés sur les critères ACMG [23], qui fournissent un tableau de critères validés ou non par le variant (Tableau 1), ce qui permet ensuite, en les combinant, de les classer en variants « pathogénique », « probablement pathogénique », « signification incertaine » ou « bénin » (Tableau 3). L'ensemble des variants considérés lors de cette classification devaient répondre à des critères stricts de qualité : ils devaient avoir une couverture minimum de 20 reads en profondeur, et devaient posséder un phred



score de qualité supérieure à 30. Le phred score est communément utilisé pour caractériser la qualité du séquençage. Un score de 30 est jugé acceptable car il indique que la probabilité qu'une base ait été désignée incorrectement est de 1 pour 1000. Ainsi, nous avons pu reconnaître et éliminer des portions de séquences de faible qualité.

|                        |  |
|------------------------|--|
| Pathogenic             | <ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 1</math> Strong (PS1–PS4) OR</li> <li>(b) <math>\geq 2</math> Moderate (PM1–PM6) OR</li> <li>(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) OR</li> <li>(d) <math>\geq 2</math> Supporting (PP1–PP5)</li> </ul> </li> <li>(ii) <math>\geq 2</math> Strong (PS1–PS4) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(b) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> Supporting (PP1–PP5) OR</li> <li>(c) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul> </li> </ul> |
| Likely pathogenic      | <ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND 1 moderate (PM1–PM6) OR</li> <li>(ii) 1 Strong (PS1–PS4) AND 1–2 moderate (PM1–PM6) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(iv) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(v) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(vi) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul>  |
| Benign                 | <ul style="list-style-type: none"> <li>(i) 1 Stand-alone (BA1) OR</li> <li>(ii) <math>\geq 2</math> Strong (BS1–BS4)</li> </ul>  |
| Likely benign          | <ul style="list-style-type: none"> <li>(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) OR</li> <li>(ii) <math>\geq 2</math> Supporting (BP1–BP7)</li> </ul>  |
| Uncertain significance | <ul style="list-style-type: none"> <li>(i) Other criteria shown above are not met OR</li> <li>(ii) the criteria for benign and pathogenic are contradictory</li> </ul>   |

**Tableau 3** : Règle de combinaison des critères pour la classification des variants, d'après Richards *et al* [23].

## 2.6 Isolement d'ARN et RT-PCR

Les cellules mononucléées du sang périphérique ont été isolées par centrifugation sur gradient de densité Ficoll-Paque™. L'ARN total a été extrait avec du Trizol. Après élimination de l'ADN génomique par digestion par la DNase (kit Turbo DNA

free, Applied Biosystems), l'ARN total (1 µg) a été soumis à une transcription inverse avec oligodT (Promega) à l'aide du kit de synthèse d'ADNc SuperScript™ VILO™ (Invitrogen). L'analyse PCR a été réalisée avec un appareil MasterCycler (Eppendorf) à partir de 2 µL d'ADNc en utilisant des amorces d'IDT. Les séquences d'amorce directe et inverse étaient respectivement: 5' GTGGAATACATGCTTATTAATCCATTGACC 3' et 5' GAGCAAGAGAGAGGTGTAGGCATC 3'.

Après une étape initiale de dénaturation à 94 ° C pendant 2 minutes, 30 cycles ont été réalisés. Un cycle était composé d'une étape de dénaturation à 94°C pendant 30 secondes, appariement des amorces à 55°C pendant 30 secondes, puis une extension à 72°C pendant 45 secondes. Une dernière étape d'extension à 72°C a été poursuivie pendant 5 min. Les produits de PCR ont été analysés sur un gel d'agarose à 2%. La longueur de l'amplicon a été évaluée en utilisant un marqueur de poids moléculaire standard (échelle d'ADN 100bp, Promega). Nous avons effectué l'extraction sur gel en utilisant le kit Clean-up Gel extraction de Macherey-Nagel pour isoler les bandes d'ADN du gel d'agarose pour le séquençage.

## 2.7. Séquençage Sanger

Le variant candidat identifié dans le gène *EFTUD2* identifié chez l'individu A a été validé à l'aide du séquençage capillaire de Sanger. En bref, un segment d'ADN d'environ 250 paires de bases a été amplifié à l'aide du système PCR Expand Long Template (Roche, Meylan, France), conformément aux recommandations du fabricant. La paire d'amorces de PCR utilisée l'amplification du fragment *EFTUD2* est présentée dans le tableau 4. Après purification avec le kit Exostar (GE Healthcare, Little Chalfont, Royaume-Uni), les produits de PCR ont été séquencés de manière bidirectionnelle avec les mêmes amorces en utilisant le kit Big Dye Terminator v3.1 (Life Technologies). Les réactions de séquence ont été effectuées sur un séquenceur ABI PRISM 3730xl (Life Technologies).

| Nom              | Séquence             | %GC | longueur | Tm |
|------------------|----------------------|-----|----------|----|
| <i>EFTUD2-Fw</i> | TTCAAGTTCTCTGGCTCCCA | 50  | 20       | 62 |
| <i>EFTUD2-Rv</i> | CCCTCAGTTCACCCTACCAG | 60  | 20       | 68 |

**Tableau 4** : Liste des primers utilisé pour la validation du variant *EFTUD2* par séquençage Sanger.

### 3. Résultats

#### 3.1. Conception du test et séquençage d'une cohorte de patients

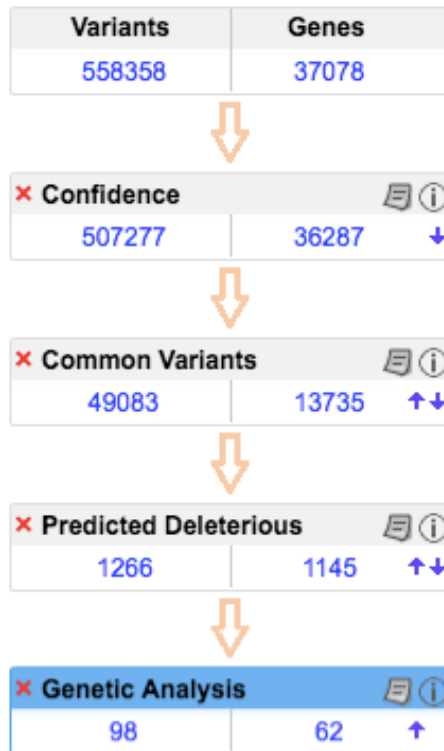
Afin de comparer les performances du séquençage d'exome au séquençage ciblé de panels de façon représentative, nous avons inclus dans l'étude 26 individus atteints de divers troubles du neurodéveloppement supposés d'origine génétique. Ces individus avaient déjà eu un parcours clinique classique consistant en des séries de consultation en génétique et des examens correspondants.

Pour ces individus, le séquençage ciblé de panel de gène et le CGH array, n'a pas permis d'élucider la cause moléculaire de la maladie. Dans cette cohorte le taux de diagnostic était donc de 0%.

L'âge moyen des enfants de la cohorte étudiée est de 4,8 ans. Nous avons pu les classer en quatre groupes selon les troubles qu'ils présentent : troubles neuro-développementaux plus ou moins sévère (6 individus), trouble du spectre autistique avec ou sans déficit intellectuel (4 individus), malformations congénitales (7 individus) et troubles psychomoteurs / hypotonie (7 individus).

Nous avons choisi d'utiliser une approche en trio afin de faciliter l'identification des modes de transmission des variants (*de novo* et hétérozygotes composites en trans, homozygotes).

Une profondeur moyenne de séquençage des régions exoniques de 158X a été atteinte, et 99,1% des variants ont été couverts au moins à une profondeur de 10 X. Le logiciel Ingenuity Variant Analysis (IVA) nous a permis d'appliquer une série de filtres pour réduire la liste des variants identifiés (Figure 11). En particulier, l'analyse conjointe du patient et de ses parents sains sur IVA nous a permis de filtrer les variants selon leur mode de transmission et d'isoler les variants *de novo*, les variants hétérozygotes composites et homozygotes hérités des deux parents, particulièrement intéressant lorsque les parents sont sains. La signification clinique des variants a été interprétée selon les normes et directives ACMG [23].



**Figure 11** : Série de filtres appliqués à l'aide du logiciel IVA (Qiagen). Le premier filtre « Confidence » filtre les variants passant le seuil de qualité fixé à 20, puis le filtre « Common Variants » exclue tous les variants dont la fréquence est supérieure à 0,5% dans la population générale selon 1000 Genome Project, ExaC, GnomAD, ou NHLBI ESP exomes. Puis, le filtre « Predicted Deleterious » conserve uniquement les variants établis comme pathogéniques, les variants frameshift, faux sens, et les variants sur les sites d'épissage. Enfin, « Genetic Analysis » permet de sélectionner le mode de transmission en filtrant le génotype du cas et des parents.

### 3.2 Troubles neuro-développementaux

Ce groupe comporte sept enfants aux signes cliniques hétérogènes, plus ou moins sévères, qui sont décrits dans le tableau 5. Nous avons pu identifier des variants dans des gènes spécifiquement associés aux phénotypes des patients pour six d'entre eux. Nous avons identifié chez les patients A, B, et C des SNPs *de novo* dans les gènes *EFTUD2*, *ADNP* et *CASK*, impliqués dans des syndromes pouvant correspondre aux phénotypes des enfants. Les individus D et E sont des sœurs issues d'un couple consanguin, chez lesquels nous avons découvert deux variants rares dans les gènes *WDR62* et *GRM1*, homozygotes chez elles et hétérozygotes chez les parents. Les pertes de fonctions de ces deux gènes sont impliquées dans des syndromes autosomiques récessifs rares, et qui couvrent l'ensemble des signes

cliniques observés chez les deux sœurs. Enfin, nous avons identifié chez l'individu F un variant dans le gène *ZNF41*, transmis par le chromosome X maternel et impliqué dans une forme de déficience intellectuelle.

Les variants identifiés chez les individus A, B et C ont pu être classés comme pathogéniques selon les critères ACMG, en revanche les variants candidats pour les individus D, E et F sont de signification incertaine. Les variants chez les individus A et C ont pu être classés comme probablement pathogéniques en raison de leur absence chez les parents (variant *de novo* : critère PS2) mais sont intrinsèquement des VUS. Concernant l'individu A, le variant identifié dans le gène *EFTUD2* est synonyme, et donc à priori sans effet sur le produit du gène. Cependant, *EFTUD2* code pour une GTPase fonctionnant dans le complexe du spliceosome impliqué dans la maturation des ARNm et causant le syndrome rare MFDM. Ce dernier semblant correspondre étroitement aux symptômes de l'enfant [254], nous avons effectué une analyse fonctionnelle de ce variant afin de démontrer son impact sur la fonction de *EFTUD2*. La description de ce cas et l'analyse fonctionnelle du variant synonyme dans le gène *EFTUD2* ont fait l'objet de l'article « A *de novo* synonymous variant in *EFTUD2* disrupts normal splicing and causes mandibulofacial dysostosis with microcephaly: case report », publié dans la revue à comité de lecture BMC Medical Genetics, et qui est présenté dans la section suivante.

| Description clinique |       |            |   | Gènes candidats |   |   |   |                    |   |
|----------------------|-------|------------|---|-----------------|---|---|---|--------------------|---|
| Individu             | Sexe  | Age        | Signes cliniques  | Nom du gène     | Variant   | Mode de transmission / Ségrégation              | Référence OMIM  | Fréquence (GnomAD) | Clinvar / ACMG guidelines                 |
| A                    | F     | 7          | <ul style="list-style-type: none"> <li>- Microcéphalie postnatale à -3DS,</li> <li>- Surdit  de perception,</li> <li>- Retard global avec difficult s de compr hension,</li> <li>- Retard de graphisme sans retard moteur.</li> <li>- Dymorphisme facial : micror trognatisme, malocclusion dentaire, grandes oreilles d coll es.</li> </ul>  | <i>EFTUD2</i>   | NM_001142605.1<br>c.597G>T; p.G199G<br>h t rozygote         | Autosomique dominant<br><i>De novo</i>          | 610536<br>syndrome de dysostose mandibulo-faciale-microc phalie                   | 0                  | Probablement pathog nique (PS2, PM2, PM1) |
| B                    | F     | 4 ans      | <ul style="list-style-type: none"> <li>- Retard de langage important, d ficience intellectuelle</li> <li>- retard de coordination, retard pour la marche (23 mois),</li> <li>- Dymorphisme facial</li> </ul>  | <i>ADNP</i>     | NM_001282532.1<br>c.1310dupC;<br>p.G438fs*2<br>H t rozygote | Autosomique dominant<br><i>De novo</i>          | 615873<br>Syndrome de Helsmoortel-Van Der Aa                                      | 0                  | Pathog nique (PVS1, PS2, PM2, PP3)        |
| C                    | M     | 6 ans      | <ul style="list-style-type: none"> <li>- Microc phalie   -2,5DS associ e   un d calage des acquisitions, une  pilepsie, une pachygyrie pari to-occipitale, petite taille</li> <li>- caract ristiques faciales : saillie de la m topique, racine du nez large, strabisme convergent gauche, une macrostomie, un micro- trognathisme</li> <li>- Ataxie statique et cin tique</li> </ul> | <i>CASK</i>     | ENST00000644219.1<br>c.2506C>A ; p.L836M<br>H t rozygote    | Autosomique dominant<br><i>De novo</i>          | 300749<br>Retard mental et microc phalie avec hypoplasie pontique et c r belleuse | 0                  | Probablement pathog nique (PS2, PM2, PP3) |
| D et E               | F ; F | 1 et 5 ans | <ul style="list-style-type: none"> <li>- D ficit intellectuel s v re</li> <li>-  pileptiques,</li> <li>- Cataracte cong nitale,</li> <li>- Microc phalie   -4DS,</li> <li>- Petite stature,</li> <li>- Livedo g n ralis  tr s marqu </li> <li>- R traction de l' tage moyen avec microphthalmie, un front haut, un philtrum court</li> <li>- Retard psychomoteur s v re.</li> </ul>   | <i>WDR62</i>    | c.882+951C>T ; p. ?<br>Homozygote                           | Autosomique r cessif<br>H rit  des deux parents | 604317<br>Microc phalie 2   | 0                  | Signification incertaine (PM2, BP4)       |
|                      |       |            |   | <i>GRM1</i>     | NM_001278064.2<br>c.3206C>T ;<br>p.P1069L<br>Homozygote     | Autosomique r cessif<br>H rit  des deux parents | 614831<br>Ataxie spinoc r bell-euse   | 0,00499%           | Probablement b nin (PM2, BP1, BP4, BP6)   |
| F                    | M     | 32 ans     | <ul style="list-style-type: none"> <li>- Troubles envahissants du d veloppement avec r gression vers l' ge de 24 mois</li> <li>- D ficience intellectuelle</li> </ul>   | <i>ZNF41</i>    | c.73-42A>C ; p.?  | Syndrome li    l'X<br>H rit  de la m re         | -   | 0,065%             | Signification incertaine (PVS1, PP3, BS1) |
| G                    | M     | 13 ans     | <ul style="list-style-type: none"> <li>- Retard de langage important</li> <li>- Troubles de l'attention</li> </ul>  | Pas de candidat | -   | -   | -   | -                  | -   |

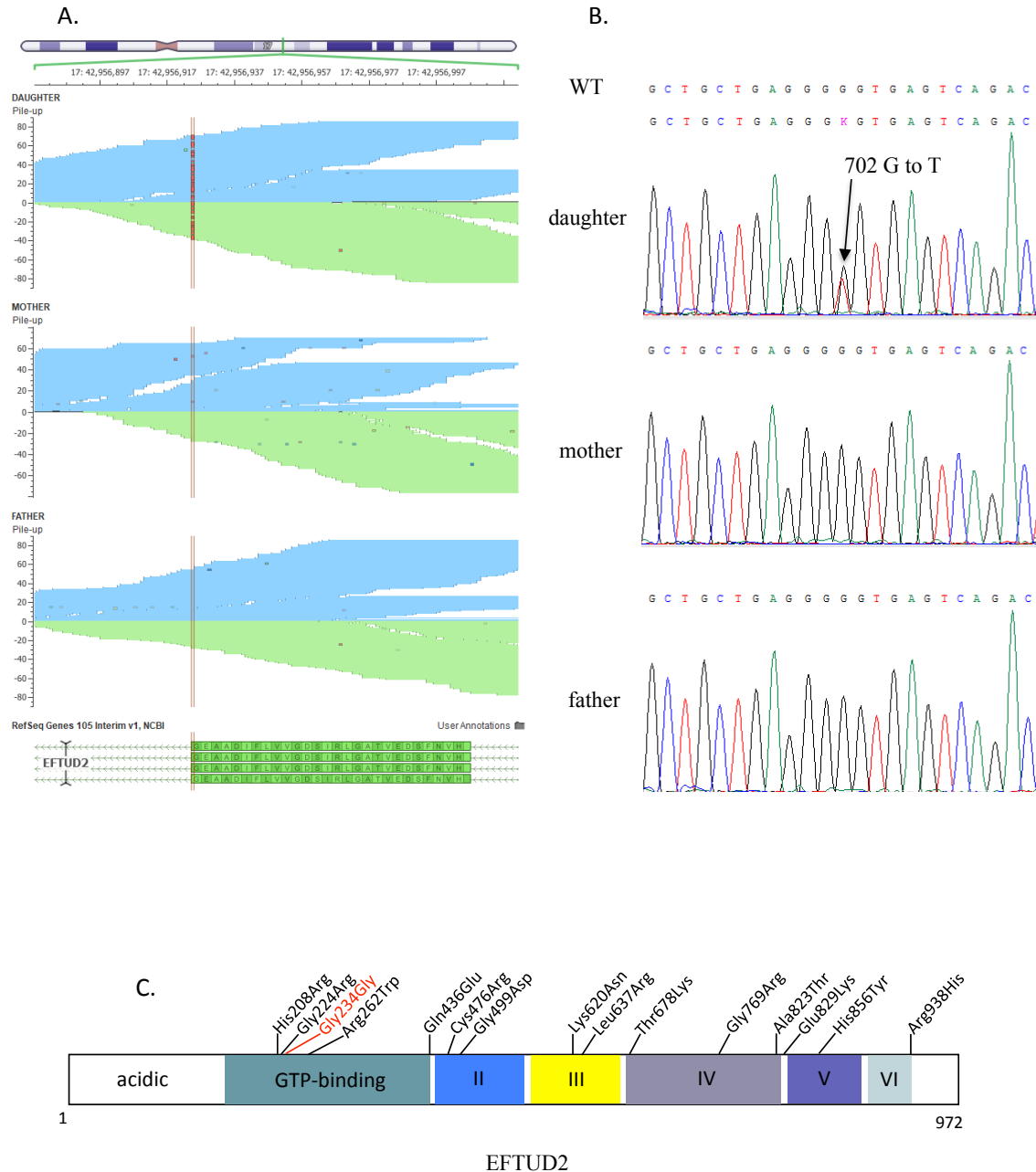
**Tableau 5** : Description des cas atteints de troubles neuro-d veloppementaux.

### 3.2.1 Analyse fonctionnelle de l'impact du variant synonyme de *EFTUD2* identifié chez l'individu A

Cette enfant âgée de 7 ans présente de multiples symptômes incluant une microcéphalie, une surdité de perception, une déficience intellectuelle, un retard graphisme ainsi qu'une dysmorphie faciale.

Nous avons identifié chez elle le variant *de novo* c.702G> T dans l'exon 9 du gène *EFTUD2*, à la position chr17: 42956924 (figure 12A). Ce variant est synonyme, n'entraînant pas de modification du résidu glycine 199 (p.G199G). Il est considéré de signification incertaine, bien que le fait qu'il soit *de novo* permette de le classer comme probablement pathogénique selon les critères ACMG (PS2). L'allèle T à cette position est nouveau dans toutes les bases de données publiques, y compris le projet de séquençage NHLBI Exome, le projet 1000 Genomes et GnomAD, suggérant une très grande conservation de l'allèle G dans la population (PM2).

Le séquençage de Sanger a confirmé la présence du variant *de novo* chez le proband (figure 12B).



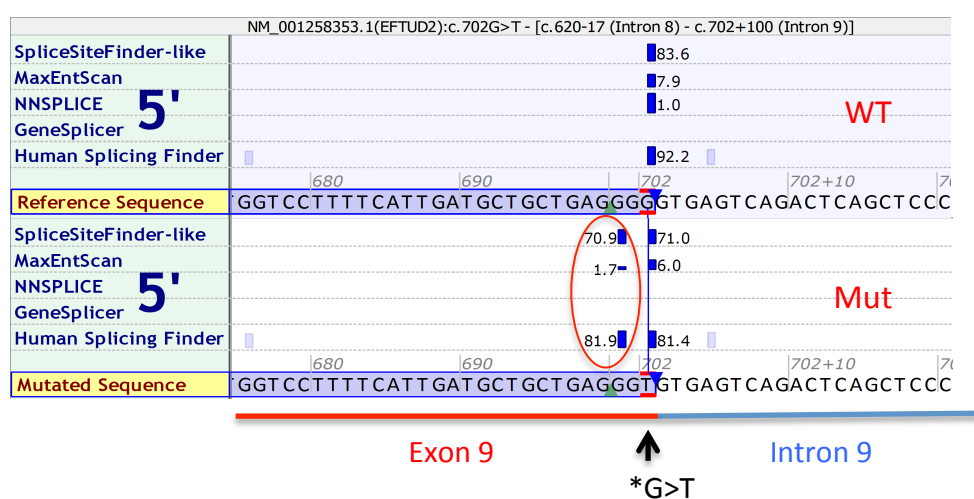
**Figure 12:** Identification du variant *de novo* chez l'individu A. (A) Les graphiques présentent les lectures WES au niveau de la région d'intérêt. Les lignes rouges soulignent le variant *de novo* c.702G> T identifié dans le gène EFTUD2 (graphique du haut) par rapport à ses parents (bas). (B) Électrophérogramme du séquençage de Sanger montrant le changement de nucléotide de G en T chez le probant par rapport au séquençage des échantillons des parents. Ce variant est hétérozygote car les deux allèles possèdent un nucléotide différent. (C) Vue schématique de la structure de la protéine EFTUD2, avec les variantes les plus importantes causant la maladie MFDM. Le variant synonyme que nous avons identifié est en rouge.



### 3.2.1.1 Prédictions in silico

Le résidu muté est le dernier nucléotide de l'exon 9, localisé à la jonction exon / intron adjacent au site donneur d'épissage GT (c.702 + 1 et +2). Selon trois outils de prédiction d'épissage, SpliceSiteFinder-like (SSF), MaxEntScore (MES) et Human Splicing Finder (HSF), le variant c.702G> T affecte le site d'épissage du donneur en créant un site donneur alternatif fort «GT» précédant celui d'origine (figures 13A, 13B).

A.



B.

|                   | SSF<br>[0-100] | MaxEnt<br>[0-12] | NNSPLICE<br>[0-1] | GeneSplicer<br>[0-24] | HSF<br>[0-100] |
|-------------------|----------------|------------------|-------------------|-----------------------|----------------|
| Threshold         | ≥ 70           | ≥ 0              | ≥ 0.4             | ≥ 0                   | ≥ 65           |
| Exon 9 –<br>c.700 | 70.9           | 1.7              | nc                | nc                    | 81.9           |

**Figure 13** : Prévisions in silico de l'impact du variant EFTUD2 c.702G> T sur l'épissage. (A) Capture d'écran des prédictions in silico de l'impact du variant EFTUD2 c.702G> T. Trois outils sur cinq prédisent que le variant G> T crée un nouveau site d'épissage donneur 5' (cercle rouge) dans la séquence mutée (rectangle inférieur) par rapport à la séquence de type sauvage (rectangle supérieur). L'exon 9 (rouge) et l'intron 9 (bleu) sont mis en évidence sur la figure. B) Le tableau représente les scores d'épissage pour le nouveau site d'épissage donneur 5' par cinq outils de prédiction.

### 3.2.1.2 Validation fonctionnelle

Afin de tester la prédiction bioinformatique, nous avons étudié les conséquences de ce variant sur l'épissage *in vivo* du gène *EFTUD2* dans le sang du proband et de ses parents. Après l'isolement de l'ARN à partir des leucocytes, nous avons effectué une RT-PCR et amplifié 360 bases couvrant les exons 8 à 12 de l'ADNc d'*EFTUD2*. Nous avons observé chez l'enfant un deuxième produit de PCR de 280 bases, suggérant la perte d'environ 80 pb dans une partie de l'ADNc *EFTUD2* du proband (Figure 14A).

Le séquençage Sanger de la bande de 280 pb a montré une délétion complète de l'exon 9 (figure 14B, 14C). La longueur de l'exon 9 n'étant pas un multiple de 3 (83 pb), sa délétion doit provoquer un décalage du cadre de lecture menant à un codon stop prématuré terminant la protéine en position p.234 (Figure 15). Ce résultat démontre que le variant synonyme *de novo* identifié dans *EFTUD2* est responsable du défaut d'épissage conduisant au saut de l'exon 9, un exon présent dans tous les isoformes d'épissage d'*EFTUD2*.

WT allele

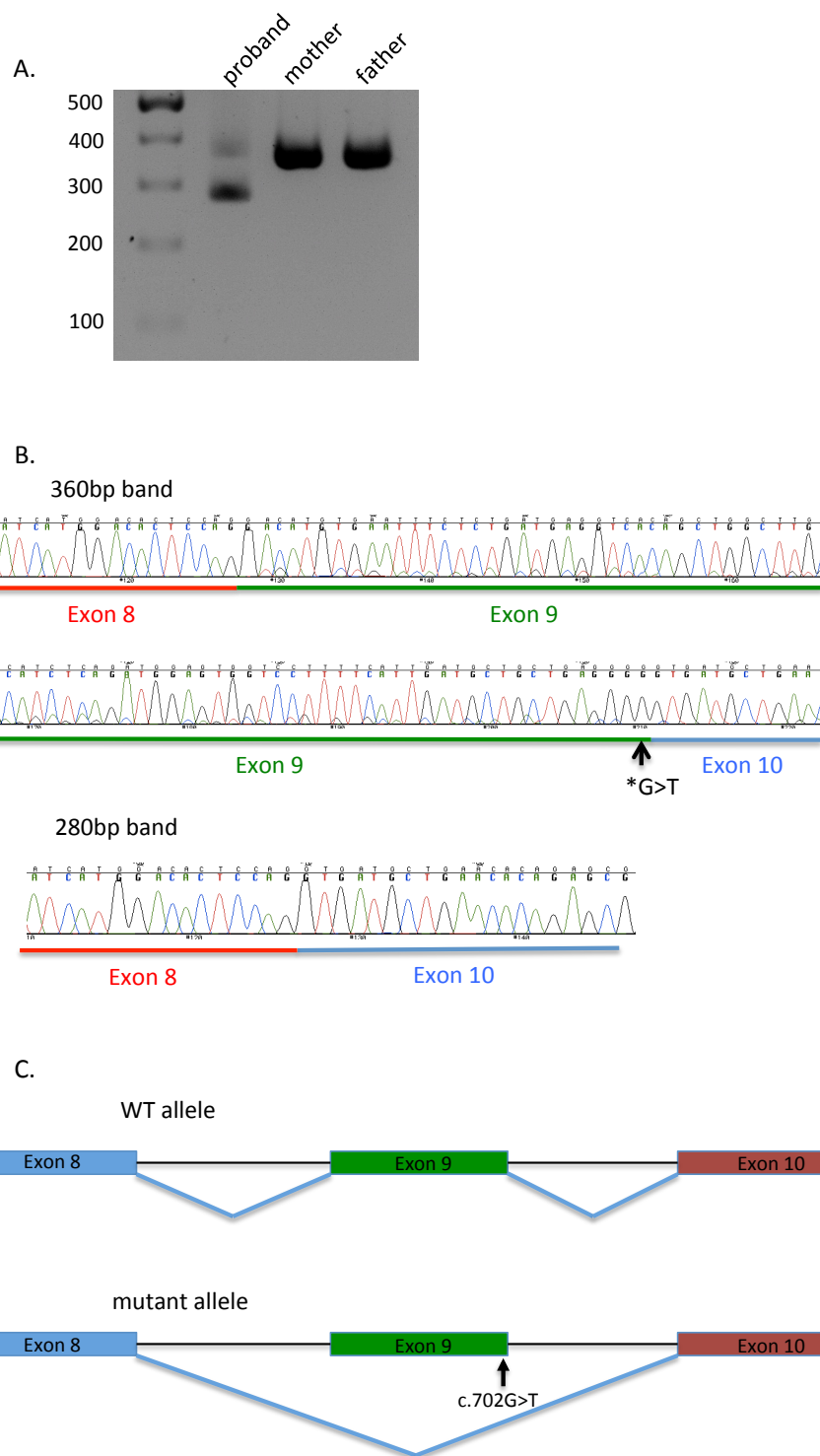
|      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| aaa  | tct  | aaa  | tct  | tat  | ctc  | ttc  | aat  | atc  | atg  | gac  | act  | cca  | gga  |
| K195 | G196 | K197 | S198 | Y199 | L200 | F201 | N202 | I203 | M204 | D205 | T206 | P207 | G208 |
| cat  | gtg  | aat  | ttc  | tct  | gat  | gag  | gtc  | aca  | gct  | ggc  | ttg  | atc  | tca  |
| H209 | V210 | N211 | F212 | S213 | D214 | E215 | V216 | T217 | A218 | G219 | L220 | I221 | S222 |
| gat  | gga  | gtg  | gtc  | ctt  | ttc  | att  | gat  | gct  | gct  | gag  | ggg  | gtg  | atg  |
| D223 | G224 | V225 | V226 | L227 | F228 | I229 | D230 | A231 | A232 | E233 | G234 | V235 | M236 |
| ctg  | aac  | aca  | gag  |      |      |      |      |      |      |      |      |      |      |
| L237 | N238 | T239 | E240 |      |      |      |      |      |      |      |      |      |      |

Mutant allele

|      |      |      |      |      |      |      |      |      |      |      |             |      |      |
|------|------|------|------|------|------|------|------|------|------|------|-------------|------|------|
| aaa  | tct  | aaa  | tct  | tat  | ctc  | ttc  | aat  | atc  | atg  | gac  | act         | cca  | ggt  |
| K195 | G196 | K197 | S198 | Y199 | L200 | F201 | N202 | I203 | M204 | D205 | T206        | P207 | G208 |
| gat  | gct  | gaa  | cac  | aga  | gcg  | gct  | gat  | caa  | gca  | tgc  | ggt         | gca  | gga  |
| D209 | A210 | E211 | H212 | R213 | A214 | A215 | D216 | Q217 | A218 | C219 | G220        | A221 | G222 |
| gag  | gct  | ggc  | agt  | cac  | tgt  | gtg  | cat  | caa  | caa  | gat  | tga         |      |      |
| E223 | A224 | G225 | S226 | H227 | C228 | V229 | H230 | Q231 | Q232 | D233 | <b>STOP</b> |      |      |

Legend :    Exon 8   Exon 9   Exon 10

**Figure 15** : Séquence d'ARNm de l'allèle WT par rapport à l'allèle mutant. Le saut d'exon 9 dans l'allèle mutant provoque un décalage de cadre de lecture, conduisant à un arrêt prématuré du codon. L'exon 8 est en rouge, l'exon 9 en vert et l'exon 10 en bleu.



**Figure 14:** Le variant c.702G> T induit un saut de l'exon 9. (A) Gel d'électrophorèse de l'ADNc *EFTUD2* obtenu après amplification de la région d'intérêt du proband et de ses parents. Le proband affiche 2 bandes, une à 360 pb et une à 280 pb. (B) Électrophérogramme du séquençage de Sanger des bandes de 360 pb et 280 pb du gel d'électrophorèse en A. L'exon 8 est mis en évidence en rouge, l'exon 9 en vert et l'exon 10 en bleu. La jonction exon 8 - exon 10 de l'ADNc d'*EFTUD2* montre que l'exon 9 saute dans l'ADNc de la bande de 280 pb qui n'est présente que chez le proband. (C) Représentation schématique du saut de l'exon 9 dans l'allèle mutant du proband par rapport à l'allèle de type sauvage (WT).

### 3.3. Autisme

Nous avons réalisé l'exome de quatre jeunes patients atteints de troubles du spectre autistique associés à des déficits intellectuels plus ou moins sévères et âgés entre 2 et 8 ans (Tableau 8). Le rendement de l'analyse d'exome est connu pour être particulièrement faible, due à la grande hétérogénéité génétique et la part importante de cas causés par des CNV *de novo* difficilement détectables par l'analyse d'exome [255]. C'est pourquoi avant d'être séquencés en exome, une analyse des CNV a été réalisée par CGH array sur ces patients, sans aboutir à un résultat.

Aucun variant pertinent n'a été identifié dans les gènes associés à l'autisme. En particulier, nous avons recherché des variants dans le gène *PTEN*, responsable de 22% des cas d'autisme accompagné de macrocrânie comme chez les individus I et J [256]. Nous avons également recherché des variants dans les gènes *CDH8* et *CNOT3* pour les individus J et K respectivement car les signes cliniques peuvent correspondre, mais aucun variant n'a été découvert dans ces gènes. L'analyse d'exome en trio n'a pas permis d'identifier de candidats qui puissent être responsable de ces cas de TSA (Tableau 6).

| Individu | Sexe | Age   | Signes cliniques  | Gène candidat   |
|----------|------|-------|---|-----------------|
| H        | M    | 4 ans | <ul style="list-style-type: none"> <li>- Trouble du spectre autistique avec perte de contact visuel à l'âge de 2 ans</li> <li>- Retard de langage</li> </ul>  | Pas de candidat |
| I        | M    | 7 ans | <ul style="list-style-type: none"> <li>- Troubles du comportement du registre autistique</li> <li>- Labilité de l'attention, difficultés de concentration, regard fuyant, colères à la frustration, intégration difficile dans le groupe</li> <li>- Retard de langage. Mémoire hors norme des lieux.</li> <li>- Macrocrânie à +2.5DS</li> <li>- Hyperlaxité</li> </ul>  | Pas de candidat |
| J        | M    | 8 ans | <ul style="list-style-type: none"> <li>- Troubles du registre autistique présents dès la naissance : regard très fuyant d'emblée, sourires réponse tardifs.</li> <li>- Macrosomie à +3DS</li> <li>- Plagiocéphalie occipitale</li> <li>- Épi frontal, philtrum court, une dysplasie des oreilles, des traits épais</li> <li>- Extrémités très douces et hyperlaxes</li> <li>- Dyspraxie, une labilité de l'attention, un retard dans le graphisme.</li> </ul>                               | Pas de candidat |
| K        | M    | 2 ans | <ul style="list-style-type: none"> <li>- Troubles du développement avec risque autistique</li> <li>- Microcéphalie avec un PC de naissance à -2DS et qui est maintenant mesuré à -4,5DS</li> <li>- Décalage des acquisitions,</li> <li>- Troubles des praxies bucco-linguo-faciales,</li> <li>- Électrorétinogramme perturbé</li> <li>- Éléments faciaux (malocclusion dentaire, philtrum long et bombant, hypotélorisme, saillie de la métopique, rétraction de l'étage moyen).</li> </ul> | Pas de candidat |

**Tableau 6** : Description des individus atteints d'autisme séquencés dans notre étude.

### 3.4. Malformations congénitales

Nous avons séquencé sept trios comportant des enfants atteints de malformations dont 4 fœtus, ainsi que leurs parents. Tous les contextes cliniques sont décrits dans le tableau 7.

Nous avons identifié des variants candidats chez les quatre individus L, M, N et O. Les cas L, N et O sont des cas de récurrence de fœtus ayant subi une interruption médicale de grossesse (IMG) en raison de graves malformations cérébrales pour lesquels nous avons identifié des variants hétérozygotes composites en *trans* dans *RTTN* et *NEB* et homozygote lié à l'X dans *OPHN1*. Quant à M, qui présente un tableau complexe de symptômes avec notamment une microcéphalie et une microphthalmie, nous lui avons diagnostiqué une anémie de Fanconi provoquée par un SNV homozygote dans *FANCD2*.

Tous ces variants identifiés ne sont pas décrits dans les bases de données, et correspondent spécifiquement au phénotype observé chez l'enfant et au mode de transmission impliqué. Ils sont classés comme pathogéniques selon les critères ACMG, sauf les variants retrouvés dans *NEB* qui sont de signification incertaine.

Ces quatre familles ont donc été diagnostiquées pour des syndromes à transmission récessive (tableau 10). Ils sont donc hérités des deux parents (à l'exception de l'individu N pour qui la pathologie est liée à l'X et donc uniquement transmise par la mère). Dans ces cas, l'identification du mode de transmission est très intéressante pour le conseil génétique des couples en vue d'éventuelles futures grossesses. Ils ont pu bénéficier d'un dépistage pré-implantatoire afin de sélectionner des gamètes non porteurs des anomalies génétiques en cause.

| Description clinique |      |        |   | Gènes candidats |  |   |  |                    |   |
|----------------------|------|--------|---|-----------------|--|---|--|--------------------|---|
| Individu             | Sexe | Age    | Signes cliniques  | Nom du gène     | Variant  | Mode de transmission / Ségrégation                      | Référence OMIM   | Fréquence (GnomAD) | Clinvar / ACMG guidelines   |
| L                    | M    | -      | <ul style="list-style-type: none"> <li>- IMG pour la deuxième fois consécutive,</li> <li>- Microcrânie avec front étroit et fuyant, racine du nez large et narines antéversées, discret hypertélorisme et oreilles basses implantées,</li> <li>- Microlissencéphalie</li> <li>- Hémi-atrophie cérébelleuse droite.</li> </ul>   | <i>RTTN</i>     | <ul style="list-style-type: none"> <li>- NM_173630.4 c.4650G&gt;T ; p.T1550T hérité du père</li> <li>- NM_173630.4 c.2284C&gt;T ; p.R762* hérité de la mère</li> </ul> | Autosomique récessif<br>Hétérozygote composite en trans | 614833 :<br>Syndrome de microcéphalie, petite stature et polymicrogyrie                                | 0% ; 0,003%        | Signification incertaine (PM2, PM3, PP3) / Pathogénique (PVS1, PP3) |
| M                    | F    | 3 ans  | <ul style="list-style-type: none"> <li>- Microcéphalie,</li> <li>- Malformation corticale,</li> <li>- Microphthalmie,</li> <li>- Livedo,</li> <li>- Syndactylie, synophris,</li> <li>- Otite,</li> <li>- Reflux gastro-oesophagien,</li> <li>- Tâches café au lait et tache dépigmentée lombaire</li> </ul>   | <i>FANCD2</i>   | - NM_001018115.3 c.782A>T ; p.K261M<br>Homozygote  | Autosomique récessif<br>Hérité des deux parents         | 227646 :<br>Anémie de Fanconi et troubles associés   | 0,001%             | Pathogénique (PS4, PM2, PM3, PP3, PP5)                              |
| N                    | M    | Foetus | <ul style="list-style-type: none"> <li>- 2<sup>ème</sup> IMG consécutive pour les mêmes raisons</li> <li>- Dysmorphie faciale,</li> <li>- Au niveau microscopique : Hypoplasie / dysplasie du vermis avec remaniements supratentoriels (pachygyrie, ventriculomégalie, kystes de germinolyse, hypoplasie des voies longues),</li> <li>- Suspicion de syndrome de Zellweger.</li> </ul>  | <i>OPHN1</i>    | - NM_002547.3 c.278dupT ; p.L93fs*8  | Syndrome lié à l'X<br>Hérité de la mère                 | 300486 :<br>Retard mental lié à l'X avec hypoplasie cérébelleuse et dysmorphie faciale caractéristique | 0%                 | Pathogénique (PVS1, PM2, PP3)                                       |
| O                    | F    | Foetus | <ul style="list-style-type: none"> <li>- Mort durant la vie foetale</li> <li>- Hypotrophe, en macération majeure,</li> <li>- Dysmorphie faciale et une hypoxie tissulaire subaigüe,</li> <li>- Retard de croissance et de maturation</li> <li>- Placenta hypotrophe, oedemateux, dysmature et ischémique,</li> <li>- Cerveau macéré, contenu par les méninges, fragmenté à la coupe, hypotrophe.</li> <li>- Hypotrophie des lobes frontaux, avec absence des bulbes olfactifs.</li> </ul> | <i>NEB</i>      | - NM_001164507.1 c.21373C>G ; p.P7125A hérité du père<br>- c.14380G>T ; p.D4794Y hérité de la mère   | Autosomique récessif<br>Hétérozygote composite en trans | 256030 :<br>Nemaline myopathie 2   | 0% ; 0%            | Signification incertaine  |
| P                    | M    | -      | <ul style="list-style-type: none"> <li>- IMG</li> <li>- Hypospade sévère.</li> </ul>  | Pas de candidat |  |   |  |                    |   |

|   |   |        |   |                 |  |  |  |  |  |
|---|---|--------|---|-----------------|--|--|--|--|--|
| Q | M | -      | <ul style="list-style-type: none"> <li>- Séquence de Pierre Robin avec éventration diaphragmatique,</li> <li>- Hypoplasie unguéale,</li> <li>- Absence de cils et de sourcils</li> </ul>  | Pas de candidat |  |  |  |  |  |
| R | F | 8 mois | <ul style="list-style-type: none"> <li>- Retard de croissance intra-utérin,</li> <li>- Thorax court et des malformations étagées des côtes et du rachis : hémivertèbres, fusion vertébrale lombaire, synostoses costales multiples, cyphose modérée.</li> </ul> | Pas de candidat |  |  |  |  |  |

**Tableau 7:** Description des individus atteints de malformations congénitales.

### 3.5. Troubles du mouvement

Nous avons séquencé huit trios présentant des troubles du mouvement, regroupant myopathie, hypotonie, dystonie, tremblements et troubles psychomoteurs. Les cas sont décrits dans le tableau 8. Nous avons pu effectuer le diagnostic génétique des individus S, T, U, V et W pour lesquels nous avons identifié les variants causaux dans les gènes *ATP8A2*, *SCN8A*, *KIF1A*, *ASCC3* et *TNXB* respectivement (Tableau 8).

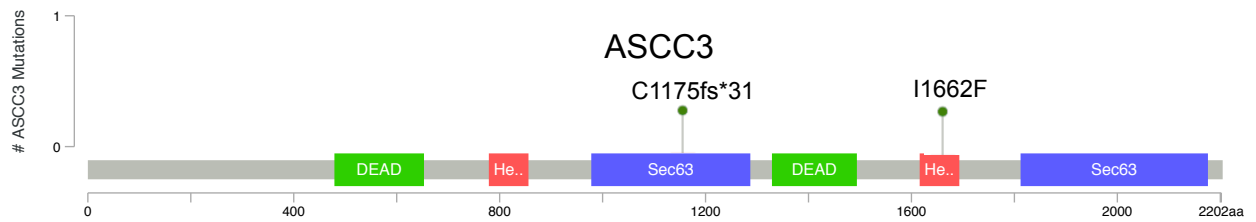
Le variant homozygote *ATP8A2* a déjà été rapporté pour causer un syndrome semblable à celui observé chez l'individu S [257]. Les variants identifiés dans les gènes *SCN8A* et *KIF1A* chez T et U sont *de novo* et classés pathogéniques. Ces variants n'ont jamais été décrits mais sont localisés dans des gènes impliqués dans des troubles du mouvement qui concordent fortement avec les phénotypes de ces individus.

#### 3.5.1 Le cas du patient V : l'identification d'une altération bi-allélique du gène *ASCC3* révèle un nouveau syndrome neuro-musculaire

Le patient V est un garçon âgé de deux ans et né de parents non consanguins. Son examen a montré une hypotonie axiale, scapulaire et pelvienne, des omoplates décollées et une raideur dans les membres inférieurs contribuant à un retard moteur général et à un manque d'expressions faciales (visage hypomimique). Par ailleurs, il présente un EEG normal, un test de potentiel évoqué auditif normal et des niveaux de créatine phosphokinase normaux.

L'analyse du WES a mis en évidence un variant hétérozygote composite dans le gène *ASCC3*. Le variant hérité du père c.3522dupA ; p.C1175fs\*31 est une duplication d'une base A provoquant un changement du cadre de lecture, et le variant hérité du côté maternel c.4984A>T ; p.I1662F est faux sens (figure 16). Ils sont tous les deux de signification incertaine selon les critères ACMG. Toutefois leur absence des bases de données (PM2), et leur pathogénicité prédite par les outils informatiques (PP3) sont des arguments en faveur de leur pathogénicité.

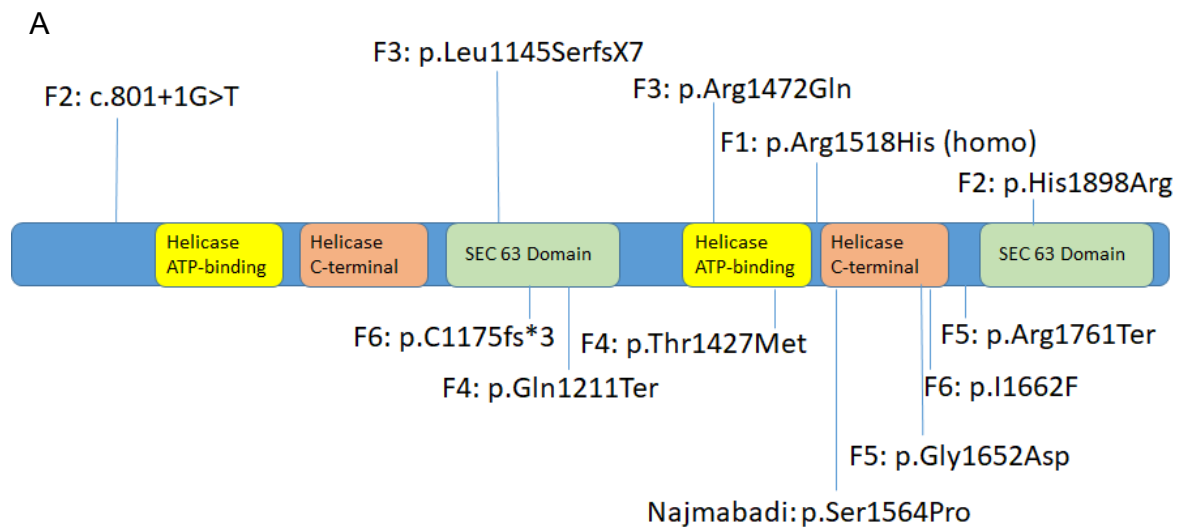




**Figure 16:** Identification de deux nouveaux variants pathogénique en trans dans le gène *ASCC3* chez l'individu V, souffrant d'une forme de myopathie.

D'un point de vue fonctionnel, ce gène n'est lui-même associé à aucune pathologie. *ASCC3* est une hélicase à ADN l'-5 qui génère de l'ADN simple brin pendant la réponse aux dommages à l'ADN. La protéine *Ascc3* forme *in vivo* le complexe ASC-1 en interagissant avec les protéines *Ascc1*, *Ascc2* et *Trip4*. D'après des études de cas récentes, des pertes de fonction bi-alléliques de *ASCC1* et *TRIP4* sont associées aux amyotrophies spinales prénatales avec fractures congénitales des os 1 [MIM 616866] et 2 [MIM 616867] [92]. Ces données fonctionnelles sont fortement en faveur du candidat *ASCC3* pour le diagnostic du patient V. Afin de le démontrer, un deuxième cas similaire d'une famille indépendante est nécessaire pour démontrer l'implication de *ASCC3* dans cette nouvelle maladie. Une recherche sur le site *GeneMatcher.org*, qui permet de connecter des chercheurs du monde entier travaillant sur les mêmes gènes, a permis de prendre contact avec une équipe du Children's Hospital of Philadelphia ayant répertorié cinq autres individus provenant de cinq familles indépendantes présentant des variants bi-alléliques de *ASCC3* associés à des troubles neuromusculaires. Le phénotype neurologique des individus atteints est variable, allant de la fatigue musculaire au retard de développement sévère. Il semble y avoir une corrélation génotype-phénotype car l'individu le plus légèrement atteint est homozygote pour un variant faux sens rare, tandis que les patients les plus sévèrement atteints sont hétérozygotes composites avec un variant faux-sens et une perte de fonction. Les signes cliniques les plus souvent observés sont le retard de développement modéré à sévère (100%) et une hypotonie (83%). Certains individus présentent également une malformation du palais (50%), des difficultés à se nourrir (50%), une extrême fatigue (33%), un visage hypomimique (33%), des oreilles basses (33%), un nez retroussé (33%), une lèvre supérieure fine (33%) et un retard de croissance intra-utérin (29%) (Figure 17). La cohorte ne comportait pas de variants LOF bi-allélique, qui n'est probablement pas viable. Notre cas a pu contribuer à cette étude multi-centrique décrivant le nouveau syndrome neuromusculaire lié à *ASCC3*, donnant lieu à l'article « Biallelic variants in *ASCC3* cause a

novel neuromuscular disorder with variable severity » soumis dans le journal à comité de lecture *Human Genetics and Genomics Advances*.



B

|                                   |                                       |
|-----------------------------------|---------------------------------------|
| Âges                              | Prenatal to 11 yr                     |
| Sexe                              | 3F/4M                                 |
| Retard de développement           | 6/6 (100%), allant de modéré à sévère |
| Hypotonie                         | 5/6 (83%)                             |
| Grandes incisives centrales       | 3/6 (50%)                             |
| Palais anormal                    | 3/6 (50%)                             |
| Difficulté à se nourrir           | 3/6 (50%)                             |
| Constipation sévère               | 2/6 (33%)                             |
| Extrême fatigue                   | 2/6 (33%)                             |
| Visage hypomimique                | 2/6 (33%)                             |
| Oreilles basses                   | 2/6 (33%)                             |
| Nez retroussé                     | 2/6 (33%)                             |
| Lèvre supérieure fine             | 2/6 (33%)                             |
| Retard de croissance intra-utérin | 2/7 (29%)                             |

**Figure 17:** Description des six cas atteints du nouveau syndrome lié à ASCC3. A) Vue d'ensemble des variants bi-alléliques identifiés chez les patients F1 à F6. B) Caractéristiques cliniques des individus présentant des variants bi-alléliques d'ASCC3. Un retard de développement, une hypotonie, de grandes incisives centrales, un palais anormal et des difficultés d'alimentation ont été décrits chez au moins la moitié des enfants.

| Description clinique |      |       |   | Gènes candidats |   |   |  |                    |  |
|----------------------|------|-------|---|-----------------|---|---|--|--------------------|--|
| Individu             | Sexe | Age   | Signes cliniques  | Nom du gène     | Variant   | Mode de transmission / Ségregation              | Référence OMIM                                     | Fréquence (GnomAD) | Clinvar / ACMG guidelines  |
| S                    | F    | 2 ans | - Sévère hypotonie, avec dystonie et mouvements permanents,<br>- Ne peut pas prendre un objet et ne mange que de la nourriture mixée.<br>- Altération de l'électrorétinogramme.                     | <i>ATP8A2</i>   | NM_016529.6<br>c.2104T>C ; p.W702R<br>Homozygote  | Autosomique récessif<br>Hérité des deux parents | 615268   | 0                  | Pathogénique (PS1, PM2, PP3, PP4, PP5)   |
| T                    | F    | 2 ans | - Retard pour la marche et la tenue debout,<br>- Défauts de myélinisation,<br>- Troubles de l'équilibre et tremblement cinétique.   | <i>SCN8A</i>    | NM_014191.4<br>c.3937A>T ; p.M1313L<br>Hétérozygote   | Autosomique dominante<br><i>De novo</i>         | 614306 et 614558 :<br>Troubles cognitifs           | 0                  | Pathogénique (PS2, PM1, PM2, PP2, PP3, PP4)                                    |
| U                    | M    | 1 an  | - Hypotonie sévère, ataxie statique et aréflexie<br>- Absence de contact oculaire.<br>- Périmètre crânien à -1DS.   | <i>KIF1A</i>    | NM_001244008.1<br>c.815A>G ; p.N272S<br>Hétérozygote  | Autosomique dominante<br><i>De novo</i>         | 614255<br>Mental retardation, autosomal dominant 9 | 0                  | Probablement pathogénique (PS2, PM2, PP3, PP4)                                 |
| V                    | M    | 2ans  | - Myopathie des ceintures sévère,<br>- Raideur des membres inférieurs   | <i>ASCC3</i>    | - NM_006828.4 c.4984A>T ;<br>p.I1662F héritée de la mère<br>- NM_006828.4<br>c.3522dupA ; p.C1175fs*31<br>héritée du père   | Autosomique récessif<br>Hétérozygote composite  | -  | 0 ; 0              | Signification incertaine (PM2, PM3) ;<br>Probablement pathogénique (PVS1, PM2) |
| W                    | M    | 8 ans | - Reflux gastro-oesophagien, douleurs intestinales pendant la digestion<br>- Echolalie  | <i>TNXB</i>     | - NM_001365276.2<br>c.6074A>T ; p.D2025V<br>héritée du père.<br>- NM_001365276.2<br>c.6622G>A ; p.D2208N hérité de la mère. | Autosomique récessif<br>Hétérozygote composite  | 606408 :<br>Syndrome d'Ehlers-Danlos               | 0,086% ;<br>0      | Probablement bénin (PM2, BP1, BP4) ;<br>Probablement bénin (PM2, BP1, BP4)     |
| X                    | F    | 2 ans | - Troubles psychomoteurs : allaitement très difficile, retard important pour la posture debout<br>- Nystagmus avec strabisme<br>- Livedo généralisé   | Pas de candidat |   |   |  |                    |  |
| Y                    | F    | 2 ans | - Reflux gastro-oesophagiens<br>- Hypotonie, avec retard moteur<br>- Périmètre crânien de -2DS<br>- Une IRM montre une légère dilatation ventriculaire et des sillons péri-cérébraux un peu larges. | Pas de candidat |   |   |  |                    |  |

|   |   |       |  |                    |  |  |  |  |  |
|---|---|-------|--|--------------------|--|--|--|--|--|
| Z | M | 11ans | - Hypotonie,<br>- Dyspraxie motrice,<br>- Fatiguabilité musculaire | Pas de<br>candidat |  |  |  |  |  |
|---|---|-------|--|--------------------|--|--|--|--|--|

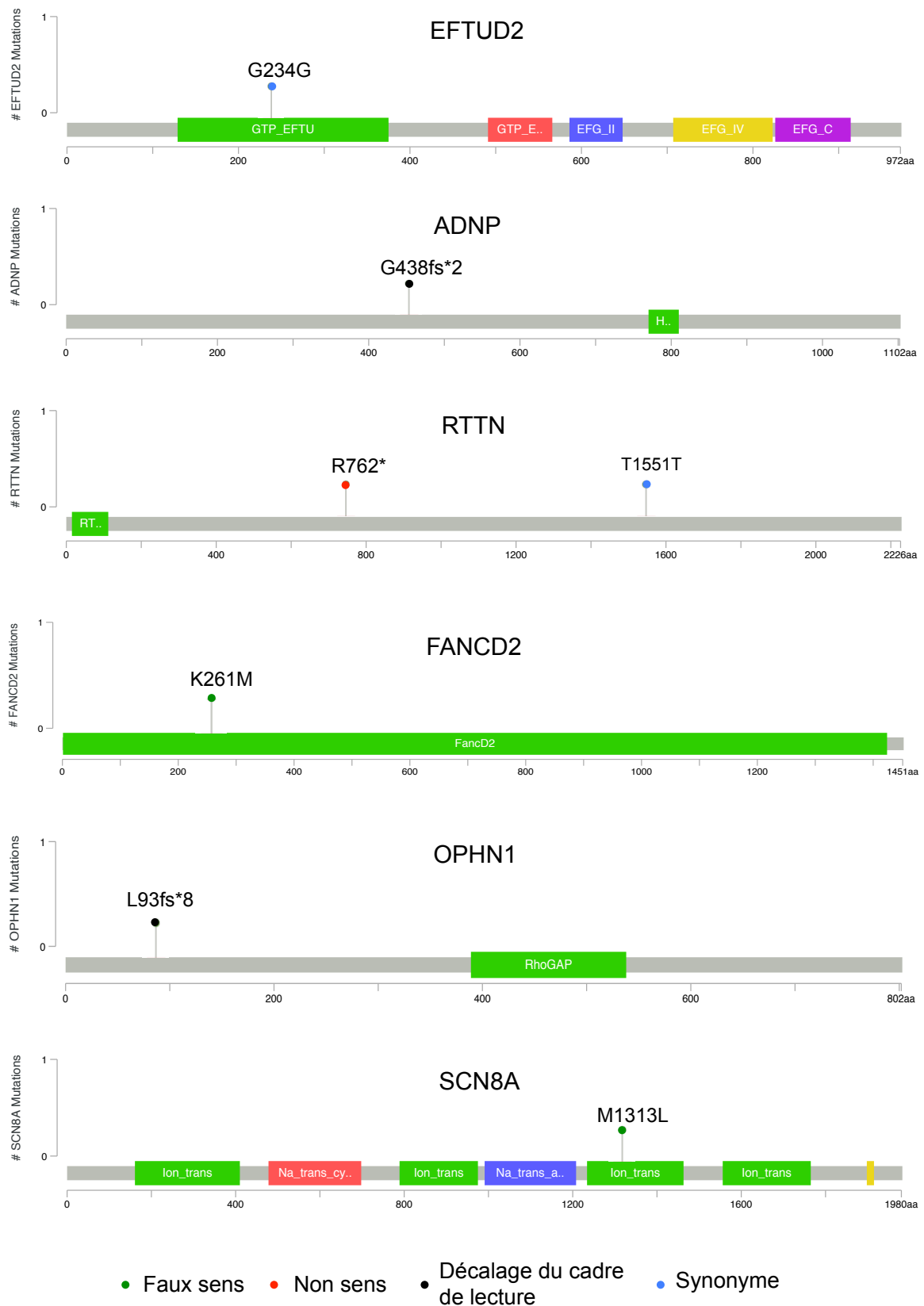
**Tableau 8** : Description des cas de individu atteints de troubles du mouvement

### 3.6. Vue d'ensemble des résultats de la cohorte

Au total, parmi les cas de la cohorte, non élucidés par le séquençage ciblé de panels et les autres examens classiques de biologie moléculaire, le WES nous a permis d'identifier des variants candidats pour quinze cas sur vingt-six, soit 57% des cas.

En réalité, nous avons considéré que le diagnostic WES est positif si les variants identifiés étaient pathogéniques ou probablement pathogéniques selon les directives ACMG et s'ils correspondaient au phénotype et au mode de transmission. Les critères ne sont donc strictement remplis que pour neuf individus sur 26, les six autres ayant été considérés comme des VUS. **Ainsi le rendement strict de diagnostic global du WES dans notre série était de 35% (figure 19A).** La pratique du WES permet donc d'améliorer sensiblement le taux de diagnostic par rapport au séquençage ciblé de panels

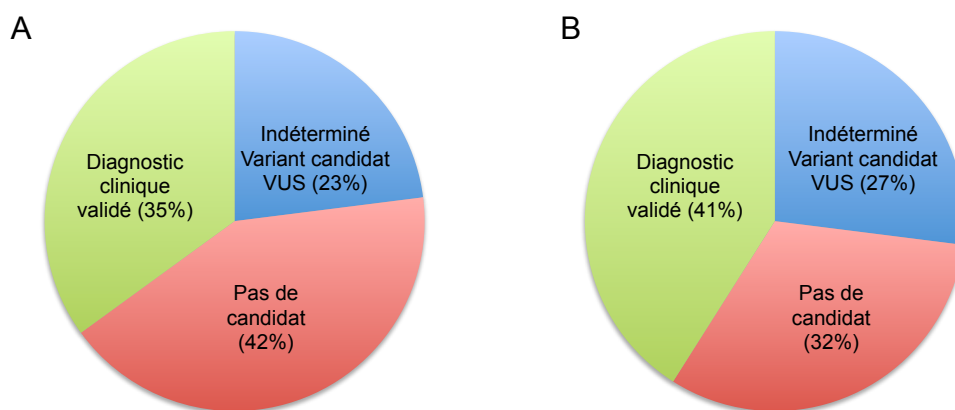
Le groupe positif comprenait neuf individus dont 6 avec une maladie autosomique dominante et trois avec une maladie autosomique récessive. 80% des variants pathogéniques ou probablement pathogéniques identifiés sont nouveaux, selon la base de données GnomAD. Ils sont résumés dans la figure 18. Pour ces patients, la précision du diagnostic génétique par WES a conduit à un meilleur conseil génétique, notamment via la possibilité de diagnostic génétique préimplantatoire ou prénatal.



**Figure 18:** Localisation et impact des nouveaux variants pathogéniques identifiés dans cette étude.

D'autre part, les six VUS affectant des gènes hautement associés au phénotype du patient n'ont pas été considérés officiellement comme des diagnostics positifs et représentaient 23% des individus. Ces variants sont des variants faux-sens, dont il manque des éléments en faveur de leur pathogénicité selon l'application stricte des critères ACMG qui les classe donc comme des VUS malgré leur fréquence très faible ou nulle dans la population. Cependant, la découverte de ces variants pourrait permettre la mise en place d'études fonctionnelles confirmant ou infirmant le rôle du variant dans le phénotype. Par exemple les variants *ATP8A2* (individu S) et *EFTUD2* (individu A) étaient des VUS au moment de l'analyse, juste avant que des études fonctionnelles soient réalisées.

Comme attendu, le rendement de l'exome pour les patients atteints de troubles purement autistiques est faible (0/4). En les mettant à part, le rendement de l'exome pour le diagnostique atteint 41%, et un candidat a pu être déterminé dans 68% des cas (Figure 19B).



**Figure 19:** Diagramme montrant le taux de diagnostic du WES en trio sur notre cohorte d'individus atteints de troubles neuro-développementaux avec (A) ou sans (B) les cas de TSA.

### 3.7 Intérêt du séquençage en trio

Le séquençage en trio, avec les parents de l'enfant malade, s'est avéré très performant pour améliorer l'analyse. Dans les cas où les parents ne sont pas atteints comme c'est le cas dans la présente étude, les modes de transmission possibles sont peu nombreux et pouvoir les identifier est d'une grande aide pour filtrer les

candidats pertinents. Le séquençage des parents nous a permis d'identifier de façon certaine et rapide les variants *de novo*, hétérozygote composites en *trans*, et les variants homozygotes dont chaque copie est héritée d'un parent, ce que n'aurait pas pu déterminer l'exome de l'enfant seul. D'un point de vue des critères ACMG, l'identification formelle d'un variant *de novo*, avec une confirmation de l'absence du variant chez les parents, ajoute le critère PS2, fortement en faveur de la pathogénicité du variant. Pour les maladies autosomiques récessives, la présence d'un variant pathogénique en *trans* ajoute un critère modéré de pathogénicité (PM3). Seul le séquençage des parents permet d'être certain que les variants hétérozygotes composites proviennent des deux parents, et qu'ils affectent donc les deux allèles chez l'enfant.

Pour les maladies dominantes comme récessives, les indications sur le mode de transmission du variant contribuent à classer un variant comme pathogénique. Il est donc évident que le séquençage en trio augmente de manière significative le rendement de l'exome pour le diagnostic en clinique. À l'échelle de notre étude, le séquençage des parents a été nécessaire pour valider le diagnostic de six patients porteurs de variants *de novo* ou hétérozygotes composites en *trans* (A, C, L, M, T et U). Ainsi, malgré son coût plus élevé inhérent, notre étude recommande d'utiliser le séquençage d'exome en trio plutôt que l'exome de l'enfant seul pour le diagnostic de maladies monogéniques rares.

### 3.8. Implication pour l'entreprise et mise en place d'un pipeline

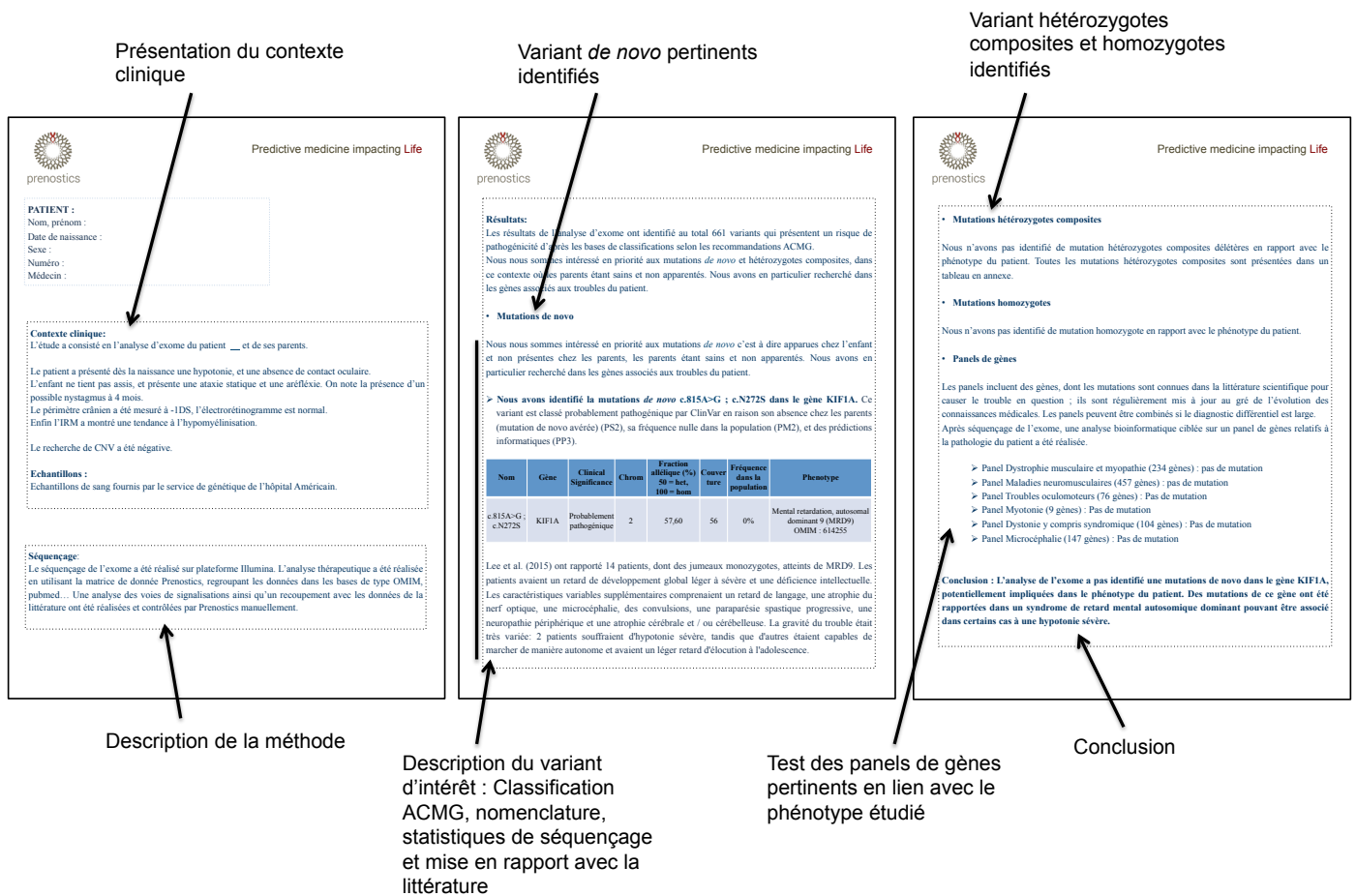
L'amélioration significative des rendements nous ont convaincu de l'utilité de l'analyse d'exome en trio en première approche diagnostique et donc de poursuivre ce type d'analyse dans l'entreprise.

L'analyse de ces cas nous a permis de mettre en place un pipeline performant permettant d'optimiser les délais et la qualité de nos analyses. Nous avons atteint un délai moyen de rendu des résultats de 6 à 8 semaines à compter de la réception des échantillons. La première phase comprend le séquençage d'exome sur la plateforme de génomique à partir des échantillons de sang ou de salive reçus. Les données brutes de séquençage sont aussitôt traitées selon les étapes décrites dans la partie Matériel et Méthodes : les fichiers FASTQ sont démultiplexés et alignés sur le génome de référence Grch37/hg19. Les paramètres de qualité des fichiers FASTQ



sont systématiquement calculés avec FastQC et les statistiques de couverture le sont avec Qualimap-2.2.1. Ensuite a lieu l'étape d'appellation des variants avec GATK. Nous avons automatisé ces étapes de traitement préliminaire des données qui permettent de disposer de données de bonnes qualités à analyser. Elles sont ensuite téléchargées sur les logiciels d'analyse IVA et Enlifer et l'analyse se poursuit manuellement. Pour chaque trio, nous combinons les analyses de l'enfant malade à celles de ses parents sur ces logiciels afin d'isoler les variants *de novo*, les variants homozygotes et les variants hétérozygotes composites. Nous annotons ensuite manuellement les variants délétères et les comparons avec les données de la littérature et la base de donnée en ligne OMIM afin d'identifier des variants candidats potentiellement en lien avec les signes cliniques de l'enfant. Nous avons inclu dans notre pipeline une relecture annuelle des résultats de l'exome pour les individus dont le diagnostic n'a pas pu être trouvé. En effet, de nombreuses nouvelles étiologies moléculaires sont découvertes chaque année, et qui amènent à réexaminer les variants identifiés par l'exome régulièrement.

Le rapport vise à rendre compte des résultats en expliquant la démarche et les éventuels variants identifiés. Celui-ci débute par la description des signes cliniques, des antécédents familiaux et des résultats aux tests déjà réalisés (séquençage ciblé de panel de gènes et CGH array) et se poursuit par la description des éventuels variants *de novo*, hétérozygotes composites, homozygotes identifiées dans l'exome correspondant au phénotype du patient. Les variants pathogéniques identifiés sont mis en contexte de la littérature et des cas précédemment décrits. Dans un deuxième temps, nous présentons la recherche contextuelle de variants dans des panels de gènes en rapport avec le phénotype de l'enfant, puis concluons le rapport par un résumé des candidats identifiés. En annexe, les limitations du test sont rappelées, ainsi que les listes complètes des variants *de novo*, hétérozygotes composites et homozygotes (Figure 20). Un rapport type est présenté en annexe (Annexe 1).



**Figure 20** : Exemple de rapport de séquençage d'exome en trio pour le diagnostic génétique d'une maladie rare non diagnostiquée

## 4. Discussion

Dans une étude prospective, nous avons cherché à évaluer l'apport du WES en trio par rapport au séquençage de panels pour le diagnostic de 26 enfants atteints de diverses affections neurodéveloppementales, de malformations congénitales ainsi que de troubles moteurs non diagnostiqués et dont les parents étaient sains. Avant de réaliser le WES, des séquençages ciblés de panels, ainsi que des CGH array ont été effectués à l'hôpital et se sont révélés négatifs. Trois des familles présentaient des antécédents familiaux.

Notre approche trio-WES nous a permis d'identifier un variant candidat dans 57% des cas (15/26). Parmi ces 15 variants candidats, 9 étaient classés pathogéniques selon les critères ACMG, et ont mené à un diagnostic moléculaire positif pour 35% des patients.

Le rendement du séquençage de panels selon la littérature est autour de 25% (variant entre 22% et 28% selon les études et les types de pathologies) [258-260]. Notre étude a permis de diagnostiquer 35% des cas parmi des cas qui n'avaient pas été élucidés par l'approche classique.

En combinant ces deux résultats et en considérant que les 25% des patients diagnostiqués par séquençage ciblé l'aient été avec la même efficacité par WES, on peut en déduire que le WES appliqué en première approche à tous les individus aurait atteint un rendement de 51,25% soit plus du double que l'approche classique. Notre taux de diagnostic est supérieur aux rendements constatés dans les études précédentes, qui oscillent entre 25 à 40% [67, 69, 70, 261].

Les altérations identifiées étaient de quatre types différents selon le mode de transmission de la pathologie impliquée par le gène. Nous avons identifié 5 variants *de novo* correspondant à des pathologies de transmission autosomique dominante (*EFTUD2*, *ADNP*, *KIF1A*, *CASK* et *SCN8A*), des variants hétérozygotes composites affectant 4 gènes correspondant à des pathologies récessives transmises (*NEB*, *RTTN*, *ASCC3*, *TNXB*), des variants homozygotes identifiées chez des enfants issus d'unions consanguines (*GRM1*, *WDR62*, *FANCD2* et *ATP8A2*), et enfin des variants transmis par le chromosome X maternel (syndrome lié à l'X) (*OPHNA1* et *ZNF41*).

Nous avons constaté un taux de diagnostic particulièrement faible pour les individus atteints de TSA associé à un déficit intellectuel, avec un taux de résolution nul (0/4). Ces derniers sont le plus souvent causés par de larges CNV *de novo* non détectables par WES, c'est pourquoi nous avons pris soin de nous assurer que la recherche de CNV avait été négatif avant de réaliser les analyses d'exome. Une étude récente menée sur 457 patients atteints d'autisme n'avait permis de diagnostiquer que 10,4% des cas par WES, ce qui confirme un faible rendement du WES pour cette pathologie [262].

Pour six autres individus, les variants candidats identifiés, bien que dans des gènes en lien avec le phénotype observé, étaient des VUS selon les critères ACMG et le diagnostic est resté incertain. Ils mettent en évidence une limitation importante que nous avons rencontrée : la classification des VUS. En effet, notre étude a identifié 8 VUS, de type faux sens ou synonymes, nous empêchant de valider le diagnostic de

20% des patients. L'impact de ces variants inconnus sur la fonction de la protéine est difficile à interpréter, et les outils de prédictions bio-informatiques ne sont actuellement pas assez puissants pour les classer selon les critères ACMG.

Pour assigner avec certitude la pathogénicité d'un VUS, la seule preuve irréfutable reste le recours à l'analyse fonctionnelle moléculaire de l'impact du variant. Nous avons pu l'illustrer par notre analyse fonctionnelle du variant synonyme identifié dans *EFTUD2*, qui a mis en évidence son effet délétère sur l'épissage de ce gène chez le patient A. Par extrapolation, ces résultats permettent de confirmer également l'effet délétère du variant synonyme du variant synonyme *RTTN* identifié chez l'individu L, car ces deux variants affectent de façon similaire le dernier nucléotide d'un exon. Nous avons montré que ce variant altère la reconnaissance d'un site exonique consensus donneur d'épissage entraînant un saut d'exon anormal.

Les SNPs synonymes sont souvent interprétés à tort comme étant silencieux compte tenu de leur impact nul sur la séquence protéique. Il est important de vérifier les prédictions de leurs effets *in silico* sur l'épissage des ARN. C'est en interrogeant les outils de prédiction SSF, HSF, MES, que nous avons pu déterminer la pathogénicité des SNP de *EFTUD2*, nous conduisant à effectuer une validation *in vitro* de sa signification clinique.

Certaines régions exoniques sont impliquées dans la régulation de l'épissage dans des sites hautement conservés appelés amplificateurs d'épissage exoniques (CSE) [263]. Ces CSE sont soumis à une forte pression de sélection au niveau du site synonyme, suggérant que des variants synonymes de ces sites pourraient être une cause fréquente de maladies génétiques à locus uniques [264]. Des variants introniques de *EFTUD2* affectant les sites donneurs d'épissage avaient déjà été découverts dans des cas de MFDM, mais, à notre connaissance, notre étude est la première à rapporter un variant synonyme délétère dans cette pathologie. Les variants synonymes ne doivent pas être ignorés, en particulier lorsqu'il est prédit que celles-ci affectent l'épissage selon les outils *in silico*.

D'autre part, le séquençage du patient V a fourni des résultats particulièrement intéressants. Il a permis d'identifier chez cet individu âgé de 2 ans atteint d'une forme sévère de myopathie des ceintures, la présence de variants hétérozygotes composites dans le gène *ASCC3*. Aucune étude publiée à ce jour ne décrit de lien entre *ASCC3* et ce phénotype. Cependant *ASCC3* est membre du complexe ASC-1

avec *ASCC1* et *TRIP4*, qui ont tous les deux été associés à des formes de myopathies similaires [92, 93]. Nous n'avons pas détecté d'autres altérations susceptibles d'expliquer la myopathie chez l'enfant. Cette suspicion de l'implication d'*ASCC3* dans la maladie de cet enfant nous a mené à soumettre ce gène au site GeneMatcher. Nous avons ainsi découvert qu'une équipe américaine travaillait sur des cas similaires présentant des variants de *ASCC3*. Nous avons pu intégrer notre cas dans leur étude présentant cinq autres cas *ASCC3*, confirmant une nouvelle forme de myopathie associée à *ASCC3*.

La détermination du diagnostic moléculaire d'une maladie génétique a un impact important sur le conseil génétique des familles, en particulier lorsqu'un risque de transmission héréditaire est identifié. Ainsi, la présence de variants hétérozygotes composites (familles K, L, S et T), et de variants homozygotes (C, D et M), identifie un risque de 25% de transmission de la maladie lors d'une future grossesse, et contribue donc particulièrement à la planification familiale future. Il en est de même pour les transmissions liées à l'X concernant les individus E et N, qui confèrent un risque de récurrence de 50% si l'enfant est de sexe masculin.

Dans ces cas, le conseil génétique propose à ces familles, dans le cadre de futures grossesses, de procéder à un diagnostic pré-implantatoire afin de sélectionner les embryons non porteurs des altérations causant la maladie. Les délais pour de telles démarches sont généralement longs (plusieurs années). Les parents peuvent se voir proposer le diagnostic prénatal précoce lorsque la future grossesse est en cours. Ce fut le cas des parents du fœtus L, porteurs de variants délétères hétérozygotes composites dans le gène *RTTN*, qui ont pu bénéficier d'un diagnostic prénatal précoce qui s'est avéré positif, ce qui a abouti à une interruption de la grossesse en cours.

L'identification d'un SNP *de novo* sont également très informatifs dans le cadre du conseil génétique des familles, en annonçant à l'inverse un risque de récurrence quasiment nul en vue d'une prochaine grossesse.

Le WES utilisé pour le diagnostic des maladies rares comporte également des limitations techniques. En couvrant uniquement les parties codantes du génome, il est très efficace pour la détection de SNPs pathogéniques, mais sa faible couverture n'est pas optimale pour la détection précise des CNVs et de réarrangements

chromosomiques. Cependant, la détection des CNVs est cruciale car elles sont la cause de la maladie dans une proportion importante des cas. La réalisation de CGH array est pour l'instant un test complémentaire au séquençage indispensable. La détection des CNV par séquençage d'exome s'est cependant nettement améliorée ces dernières années et pourrait prochainement atteindre les mêmes performances que le CGH array, notamment grâce au développement de la capture d'exome augmentée « CoDE-seq » permettant la détection précise et simultanée des CNV et des SNP [22] et à l'amélioration des outils de détection informatiques [265].

## 5. Conclusions

Malgré le coût plus élevé du WES par rapport au séquençage ciblé de panels, cette étude démontre qu'il améliore significativement le rendement diagnostique en permettant le diagnostic de 35% des patients non diagnostiqués par le séquençage de panel. Ces données prospectives soutiennent une stratégie de test qui implique l'utilisation précoce du WES en première ligne.

Les analyses WES négatives ou incertaines restent un problème important, du fait de la complexité de l'interprétation des données génomiques et du manque de connaissances sur les maladies monogéniques. En particulier, les SNP faux sens restent en grande majorité de signification incertaine, ce qui pose problème pour confirmer le diagnostic d'environ 20% des patients.

# Chapitre 3 : Intérêt du WES appliqué aux familles BRCAx et caractérisation de nouveaux facteurs de prédisposition au CS héréditaire

## 1. Introduction

Environ 5 à 10% de tous les CS surviennent dans un contexte héréditaire où un variant génétique induit un risque très élevé de développer la maladie. Les familles à risque de cancer du sein ont une structure héréditaire apparemment dominante et se caractérisent souvent par une apparition précoce de la maladie. Plusieurs gènes de prédisposition augmentent le risque de CS et peuvent être classés en fonction de leur niveau de risque. Les variants des deux gènes à haut risque, *BRCA1* et *BRCA2*, représentent environ 25% du risque familial [266, 267]. Les variants constitutionnels dans les gènes de risque intermédiaire *ATM*, *BRIP1*, *CHEK2*, *NBS1*, *RAD50*, *PALB2*, *XRCC2* sont à l'origine d'environ 5% du risque de CS familial [117-119, 268-270]. Enfin, plus récemment, environ 170 allèles communs de cancer du sein à faible pénétrance ont récemment été identifiés par des études GWAS, et expliquent 18% supplémentaires du risque de CS familial [271].

Au final, plus de 50% des cas de CS familiaux restent inexplicables, avec une prédisposition génétique inconnue. Ces cas sont dits BRCAx en raison de l'absence d'altérations des gènes à risque dans leur génome constitutionnel. Cette absence de diagnostic moléculaire rend difficile le conseil génétique de ces familles et pose problème pour leur prise en charge.

Avec l'introduction des technologies de NGS [272], de nombreuses études ont suggéré qu'un fort taux d'héritabilité du cancer du sein restant peut être attribué à de nouveaux allèles à risque rares qui se transmettent selon un modèle de transmission autosomique dominant. Les multiples études de cas / contrôle impliquant des individus non apparentés n'ont pas permis d'identifier de tels variants autres que *BRCA1* et *BRCA2*, car leur extrême rareté nécessiterait une très grande cohorte pour démontrer leur association avec le cancer du sein.

Les recherches d'association à l'échelle de la population ayant globalement échoué pour identifier de nouveaux facteurs génétiques impliqués, une hypothèse récente stipule que chaque famille BRCAx possède une prédisposition spécifique qui lui est propre. Celle-ci nous a amené à changer de stratégie et de rechercher les facteurs

de prédisposition génétique à l'échelle d'une famille atteinte de CS héréditaires BRCAx. Afin d'identifier l'héritabilité manquante d'une famille, l'idée est d'étudier les éventuels variants rares qui co-ségrégent avec les individus atteints de CS [116, 273-275].

Compte tenu de ce contexte spécifique, l'approche familiale de séquençage d'exome des membres affectés semble être une méthode pertinente pour explorer l'héritabilité manquante des formes familiales de CS. Nous avons fait l'hypothèse que l'étude de ces familles, par le séquençage d'exome, pouvait être pertinente afin d'identifier les facteurs de prédisposition génétique dans ces familles et ainsi améliorer leur suivi médical. Nous avons utilisé une approche de séquençage d'exome de familles BRCAx présentant de nombreux cas de CS et ne comportant pas de variants pathogéniques dans les panels de gènes de prédisposition au cancer du sein négatif (gènes de prédisposition connus *BRCA1/BRCA2*). En analysant la ségrégation des variants à l'échelle de l'exome parmi les individus atteints, nous avons tenté de résoudre le cas de ces familles BRCAx. Ce travail vise à déterminer si l'utilisation systématique du WES peut être pertinent pour la prise en charge des familles BRCAx.

## 2. Matériel et Méthodes

### 2.1. Patients

Dans notre étude de recherche de nouveaux facteurs de prédisposition au cancer du sein héréditaire, nous avons inclus 25 individus issus de 4 familles BRCAx dont 22 atteints et 3 sains. L'accès à cette cohorte a été rendu possible grâce à notre collaboration avec le Pr. Pascal Pujol du service d'oncogénétique du CHU de Montpellier. Ces familles ont été incluses dans l'étude en raison de leur forte prévalence de cancer du sein avec un minimum de 4 individus atteints, et de l'absence de variations génétiques délétères dans les gènes à haut risque *BRCA1* et *BRCA2*. Nous avons inclus dans l'étude des individus sains lorsque cela était possible, afin de les considérer comme « contrôles ». Des échantillons sanguins ont été obtenus pour chacun de ces individus, ainsi que des données cliniques incluant l'âge d'apparition du cancer, le sous type histologique, le grade, le statut des récepteurs hormonaux, et le traitement reçu. Tous les individus participant dans cette



étude de génétique ont donné leur consentement éclairé oral et écrit pour la recherche sur leur génome.

## 2.2. Extraction, préparation des librairies et séquençage d'exome

Pour chaque participant, l'ADN génomique total a été isolé du sang périphérique à l'aide de kit d'extraction DNeasy de Qiagen selon les instructions du fabricant. La pureté et la concentration d'ADN ont été mesurées à l'aide d'un spectrophotomètre NanoDrop™.

La préparation des librairies, ainsi que le protocole de séquençage d'exome a été le même que celui décrit dans la partie Matériel et Méthodes du chapitre 2.

## 2.3. Analyse des données

L'analyse bioinformatique des données de séquençage brut (alignement, appellation des variants) a été effectuée comme décrit dans le chapitre 2.

Les INDEL et les SNP annotés par Ensembl Perl ont été filtrés selon plusieurs critères: (a) nous avons considéré le cancer du sein comme une maladie autosomique dominante et supprimé les variants homozygotes ; (b) nous avons éliminé les variants introniques, intergéniques, afin de se concentrer sur les variants des régions codantes et des sites d'épissage uniquement ; (c) en supposant que les variantes causales sont rares, nous avons supprimé tous les variants avec une fréquence allélique >1% dans GnomAD, Exac, ou 1000 génome, (d) nous avons également supprimé les variants bénins ou tolérés, selon les différents outils de prédiction *in silico* SIFT et Polyphen-2.

## 2.4. Séquençage Sanger

Le séquençage Sanger a été réalisé comme décrit dans le chapitre 2. Le tableau ci-dessous (Tableau 9) décrit les amorces utilisées pour la validation des SNV dans les gènes candidats.

| Nom          | Séquence                 | %GC | longueur | Tm |
|--------------|--------------------------|-----|----------|----|
| CCNF-fw      | GTCTGTGTCTGATGAGGCC      | 60  | 20       | 68 |
| CCNF-rv      | CTCGGGGTCTTCCTGTGTCC     | 65  | 20       | 70 |
| TYRO3-Fw     | AGTGCTGACCCCAAGCAGCG     | 65  | 20       | 74 |
| TYRO3-Rv     | TAGGGCTGATCCCTGCCAGG     | 65  | 20       | 72 |
| TPH1-Fw      | GAACCCATAAGGTAGATGCC     | 50  | 20       | 62 |
| TPH1-Rv      | GTTACTTATCACCAAGAGATT    | 33  | 21       | 56 |
| HIST1H1C-Fw  | TGGGTTTTAGAAGTAGGCGTTCCG | 48  | 23       | 67 |
| HIST1H1C-Rv  | CCCAAAGAAGGCCAAGGTTG     | 55  | 20       | 67 |
| SLC12A3-Fw   | GCTCCTGGCTCAGCCCCAC      | 75  | 20       | 77 |
| SLC12A3-Rv   | TGTCATCTCGACCCCTTTCTGC   | 55  | 22       | 69 |
| EIF4EBP1-Fw1 | ACCTCCCTGTTCCCTTCTAG     | 55  | 20       | 66 |
| EIF4EBP1-Rv1 | CCATGGGGGGCTCATCACTG     | 65  | 20       | 71 |
| EIF4EBP1-Fw2 | TGTGACCAAACACCCCAAG      | 52  | 21       | 68 |
| EIF4EBP1-Rv2 | CCAAGCCCCGACACTCACCG     | 70  | 20       | 74 |

**Tableau 9** : Amorces utilisées pour l'amplification des régions comportant les variants candidats.

## 2.5. Prédiction *in-silico*

Nous avons sélectionné principalement deux outils de prédiction *in silico* pour évaluer les effets fonctionnels des variantes candidates: Sorting Intolerant From Tolerant (SIFT) (<http://sift.jcvi.org/>) pour examiner le degré de conservation des résidus d'acides aminés à travers les espèces et pour trouver des changements dans la structure et la fonction des protéines; PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) pour évaluer l'impact des variants sur la fonction des protéines.

## 3. Résultats

### 3.1. Séquençage d'exome des familles BRCAx

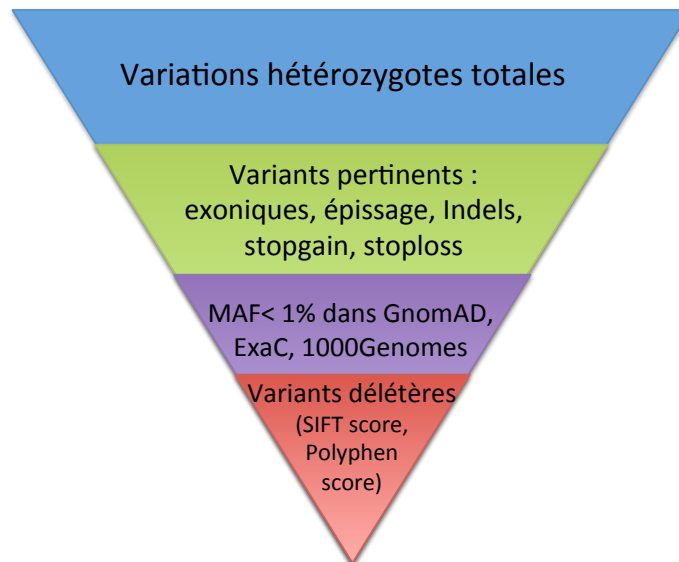
Nous avons sélectionné quatre familles BRCAx présentant un minimum de 4 individus atteints de CS, et qui se sont révélées être négatives aux tests des gènes BRCA1 et BRCA2. Nous avons collecté des échantillons sanguins du plus grand

nombre possible de membres atteints pour chaque famille, et d'au moins un individu sain dans la mesure des disponibilités. L'ADN a été extrait des échantillons afin de procéder à la préparation des librairies d'ADN et au séquençage de l'exome.

Au total, nous avons effectué le séquençage d'exome de vingt-deux femmes atteintes de CS et trois individus sains appartenant à ces familles. La couverture moyenne du séquençage était de 91,59, et 98,54% des régions ciblées ont été couvertes par au moins 8 lectures.

### 3.2. Analyse des données de séquençage et stratégie de filtrage des variants

La cause des cancers du sein familiaux peut être soit monogénique, due à un variant à forte pénétrance dans un gène, ou bien polygénique et causée par l'accumulation de plusieurs variants de faible impact mais dont les effets se multiplient. Ce risque polygénique prend en compte des loci identifiés par GWAS dans des régions intergéniques non couvertes par le WES. Nous avons donc fait l'hypothèse que ces cas de cancer du sein sont monogéniques, hérités de manière autosomique dominante et avons donc appliqué des filtres de sélection restrictifs pour l'identification d'allèles hétérozygotes à haut-risque (Figure 21). De plus, nous avons pris en compte le fait que les variants en cause ne sont pas nécessairement partagés par l'ensemble des membres atteints de la famille, en raison de la possibilité que la cause de la maladie soit génétiquement hétérogène au sein de la famille, ou de la possibilité de cas sporadiques de cancer du sein par un membre de la famille non porteur du variant génétique en cause dans le risque familial. Nous avons recherché des variants présents chez tous ou partie des membres atteints



**Figure 21:** Étapes de filtration des variants à haut-risque

En appliquant les pipelines décrits dans la section méthode nous avons obtenus les variants de type SNV et indels partagés par tout ou partie des membres atteints dans chacune des familles. Le premier filtre appliqué a consisté à ne garder que les variants ayant potentiellement un impact sur la fonction de la protéine à savoir les variants exoniques, les variants situés dans les zones d'épissage, les indels entraînant un décalage du cadre de lecture, les variants générant un codon stop (stopgain), ou encore une perte de codon stop (stoploss).

Dans une deuxième étape de filtrage, nous avons éliminé les variants dont la fréquence est supérieure à 1% de la population dans les bases de données GnomAD, 1000Genomes et ExaC. Ce seuil établi est volontairement légèrement élevé pour une recherche de variants à hauts risques, sachant que la fréquence de tels variants dans la population a été considérée inférieure à 0,001 (0,1%) dans plusieurs études [276], afin de ne pas risquer d'exercer un filtre trop strict pour l'identification d'un variant candidat.

Enfin, nous avons filtré les variants délétères pour la fonction de la protéine, selon les outils de prédiction SIFT et Polyphen-2. Nous n'avons considéré que les variants ayant un score pathogénique ou probablement pathogénique selon ces deux outils de prédiction (Figure 21).

### 3.3. Analyse des gènes connus de prédisposition au cancer du sein héréditaire

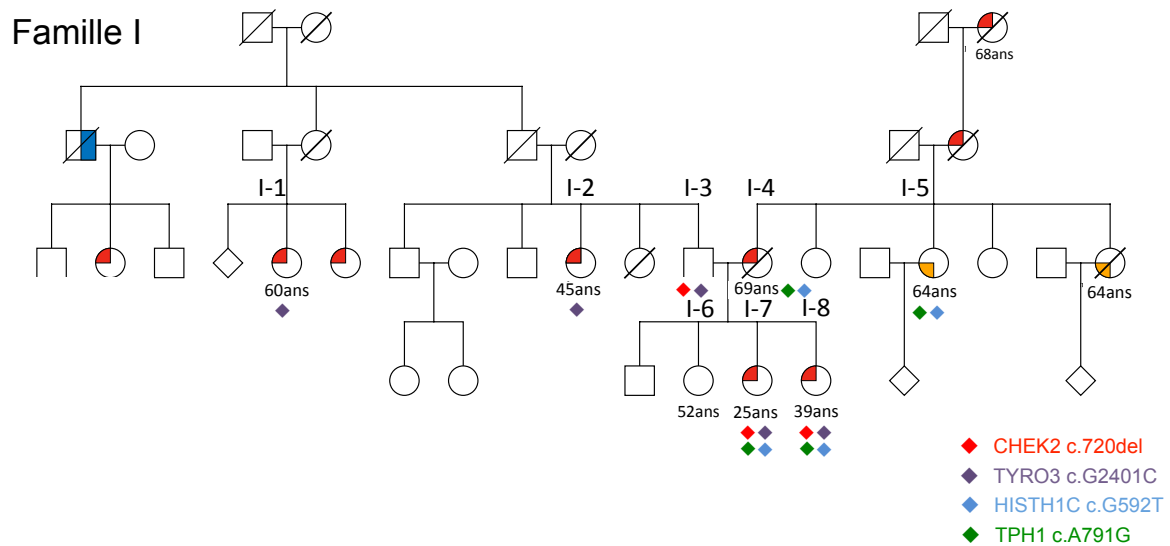
Avant d'appliquer les filtres de sélection des variants, nous avons voulu vérifier l'absence de variants délétères dans les gènes de prédisposition au cancer du sein déjà connus, composé d'un panel de 28 gènes, de pénétrance élevée *BRCA1*, *BRCA2*, *TP53*, et des gènes à pénétrance dite moyenne *ATM*, *BARD1*, *BLM*, *BRIP1*, *CDH1*, *CHEK2*, *FAM175A*, *FANCC*, *FANCM*, *MAPKAP1*, *MLH1*, *MRE11A*, *MSH2*, *NBN*, *NF1*, *PALB2*, *PMS2*, *PTEN*, *RAD50*, *RAD51B*, *RAD51C*, *RAD51D*, *RECQL*, *RINT1*, et *STK11*.

Nous n'avons identifié aucun variant pathogène connu dans les gènes à haut risque *BRCA1*, *BRCA2*, confirmant le caractère BRCAX des familles recrutées. Parmi les gènes à pénétrance modérée, nous avons identifié un variant délétère changeant le cadre de lecture dans le gène *CHEK2* chez 2 individus atteints de la famille I, que nous détaillerons dans la section qui suit. Aucun autre variant délétère ou possiblement délétère dans ces gènes de prédisposition connus n'a été détecté chez les autres individus séquencés, d'après la base de données ClinVar.

### 3.4. Identification de nouveaux gènes candidats

#### 3.4.1 Famille I

La famille I présente une fratrie de deux sœurs ayant développé un CS à un âge jeune I-7 et I-8 (25 et 39 ans, respectivement) et d'une troisième sœur indemne I-6 âgée aujourd'hui de 52 ans (Figure 9). On note la présence d'antécédents familiaux de cancers du sein et de l'ovaire du côté maternel et paternel de la famille. Nous avons pu obtenir des échantillons sanguins des 3 sœurs, de leurs parents et de trois autres individus atteints venant des branches maternelles et paternelles de la famille comme indiqué dans la figure 22, afin de réaliser un séquençage d'exome de leur génome constitutionnel.



| Individu | Age au diagnostic | Sous type histologique                 | Grade | Statuts récepteurs hormonaux | Statut HER2 | Évolution de la maladie    |
|----------|-------------------|--|-------|------------------------------|-------------|----------------------------|
| I-1      | 60                | ND                                     | ND    | ND                           | ND          | ND                         |
| I-2      | 45                | Adénocarcinome infiltrant              | III   | RE+<br>RP-                   | HER2-       | Chimiothérapie             |
| I-3      | Sain              | -                                      | -     | -                            | -           | -                          |
| I-4      | 69                | Carcinome infiltrant                   | I     | RE+<br>RP+                   |             | Chimiothérapie             |
| I-5      | 64                | Adénocarcinome séreux de l'ovaire      | II    | -                            | -           | -                          |
| I-6      | Sain              | -                                      | -     | -                            | -           | -                          |
| I-7      | 25                | Carcinome infiltrant lobulaire         | III   | RE+<br>RP-                   | HER2-       | Chirurgie                  |
| I-8      | 39                | Carcinome canalaire infiltrant luminal | II    | RE+<br>RP-                   | HER2-       | Chirurgie + Chimiothérapie |

**Figure 22** : Arbre généalogique de la famille I. Les individus en rouges sont atteints de CS, en orange de cancer de l'ovaire. Les individus séquencés en exome et leurs caractéristiques cliniques sont présentés dans le tableau

Étant donné la présence de cancer du sein dans les branches paternelles et maternelles de cette famille, nous les avons analysées de façon indépendante : d'une part des individus de la branche paternelle I-1, I-2, I-3, I-6, I-7 et I-8 et d'autre part de la branche maternelle I-4, I-5, I-6, I-7 et I-8. La soeur I-6 n'ayant pas été atteinte et aujourd'hui âgée de 52 ans, nous l'avons considéré comme témoin négatif dans les analyses des branches paternelles et maternelles.

Nous avons identifié respectivement 702 et 455 variants qui co-ségrègent avec les cas de CS du côté maternel et paternel. Comme dit précédemment, nous avons déterminé la présence de la délétion c.720del ; p.Val241fs dans le gène à pénétrance modérée CHEK2 chez les individus I-3, I-7 et I-8. Cette délétion d'un nucléotide, très rare dans la population générale (fréquence de 0,0002 d'après GnomAD), résulte en un décalage du cadre de lecture, et est classée comme pathogénique par ClinVar. Les variants tronquants de *CHEK2* augmentent les risques d'apparition d'un cancer du sein d'un facteur 3,11 [277]. Ce variant co-ségrège avec la maladie étant donné qu'il est transmis aux deux sœurs I-7 et I-8 ayant développé un cancer du sein à 25 et 39 ans par le père sain I-3, et non transmis à la troisième sœur non atteinte I-6. S'il paraît évident que ce variant a un impact sur le développement de la maladie chez I-7 et I-8, il ne peut expliquer la survenue de la maladie chez les tantes du côté paternel (I-1 et I-2), de leur mère et tante maternelle (I-4 et I-5), ni l'âge particulièrement jeune du diagnostic de I-7 et I-8. Nous avons fait l'hypothèse qu'il existe d'autres facteurs de risque dans cette famille, intervenant de façon indépendante des deux branches de la famille et s'accumulant chez les sœurs I-7 et I-8 expliquant la pénétrance très forte de la maladie chez elles.

Après filtrage des variants pour l'identification d'autres facteurs de prédisposition, nous avons identifié des variants pertinents dans le gène *TYRO3* chez tous les membres du côté paternel, et des variants dans les gènes *TPH1* et *HIST1H1C* chez tous les membres de la branche maternelle (Tableau 10). Ces variants sont absents chez le contrôle négatif I-6 et aucun n'est référencé dans la base de données ClinVar. Nous avons confirmé la présence de tous les variants par séquençage Sanger, ainsi que leur absence chez I-6 (Figure supplémentaire 1).

| Gene  | Chromosome-Position | Nucleotide | Acide aminé | Type        | Polyphen Score prediction | SIFT Score prediction | Critère ACMG                        | Membres des familles |   |   |   |   |   |   |   | Fréquence GnomAD | ClinVar    |
|-------|---------------------|------------|-------------|-------------|---------------------------|-----------------------|-------------------------------------|----------------------|---|---|---|---|---|---|---|------------------|------------|
|       |                     |            |             |             |                           |                       |                                     | 1                    | 2 | 3 | 4 | 5 | 6 | 7 | 8 |                  |            |
| CHEK2 | Chr22 : 29120965    | c.720del   | p.V241fs    | Frame shift | -                         | -                     | Pathogénique (PVS1, PM2, PP3)       | -                    | - | + | - | - | - | + | + | 2,00e-04         | Pathogenic |
| TYRO3 | Chr15:4 1578004     | c.2401G>C  | p.D801H     | Faux sens   | Probably damaging (0.979) | Deleterious (0)       | Signification incertaine (PP3, BP1) | +                    | + | + | - | - | - | + | + | 5.572 e-05       | -          |

|                            |                |          |         |           |                           |                    |                                     |   |   |   |   |   |   |   |   |            |   |
|----------------------------|----------------|----------|---------|-----------|---------------------------|--------------------|-------------------------------------|---|---|---|---|---|---|---|---|------------|---|
| HIST1H1C<br>(Histone H1.2) | Chr6:26055837  | c.592G>T | p.A198S | Faux sens | Possibly damaging (0.836) | Deleterious (0.02) | Signification incertaine (PM2, BP4) | - | - | - | + | + | - | + | + | 4.001 e-06 | - |
| TPH1                       | Chr11:18026502 | c.791A>G | p.Y264C | Faux sens | Probably damaging (0.999) | Deleterious (0)    | Signification incertaine (PP3, BS1) | - | - | - | + | + | - | + | + | 3.58 e-05  | - |

**Tableau 10** : SNVs et indels délétères identifiés chez les individus atteints de la famille I.

Nous avons recherché dans la littérature des éléments qui pourraient incriminer ces gènes candidats dans le développement du cancer du sein. Le gène *HIST1H1C* est un candidat particulièrement intéressant du fait de sa fonction. Il encode l'Histone H1.2 qui interagit avec l'ADN entre les nucléosomes et joue un rôle dans la compaction de la chromatine. Récemment, il a été montré qu'il avait un rôle essentiel dans l'initiation de la réparation de l'ADN par RH en réponse aux cassures de l'ADN double brin, en permettant à ATM de se lier spécifiquement sur le site de la cassure pour amorcer la réaction [278].

TYRO3 est un récepteur tyrosine kinase transmembranaire ayant un rôle oncogénique reconnu en activant notamment la voie PI3K en aval. Il est surexprimé dans de nombreux types de cancers et est également impliqué dans la prolifération des cellules de cancer du sein MCF-7 *in vitro* [279, 280].

Le troisième gène candidat de la famille I, *TPH1*, encode une enzyme responsable de la production du neurotransmetteur sérotonine. Elle est surexprimée en particulier dans les cancers du sein triple négatifs, et est impliquée dans la progression tumorale via les effets autocrines de la sécrétion de sérotonine [281]. Afin d'étendre les analyses, il est maintenant nécessaire de réaliser des tests fonctionnels sur ces gènes candidats dans le cadre de collaborations.

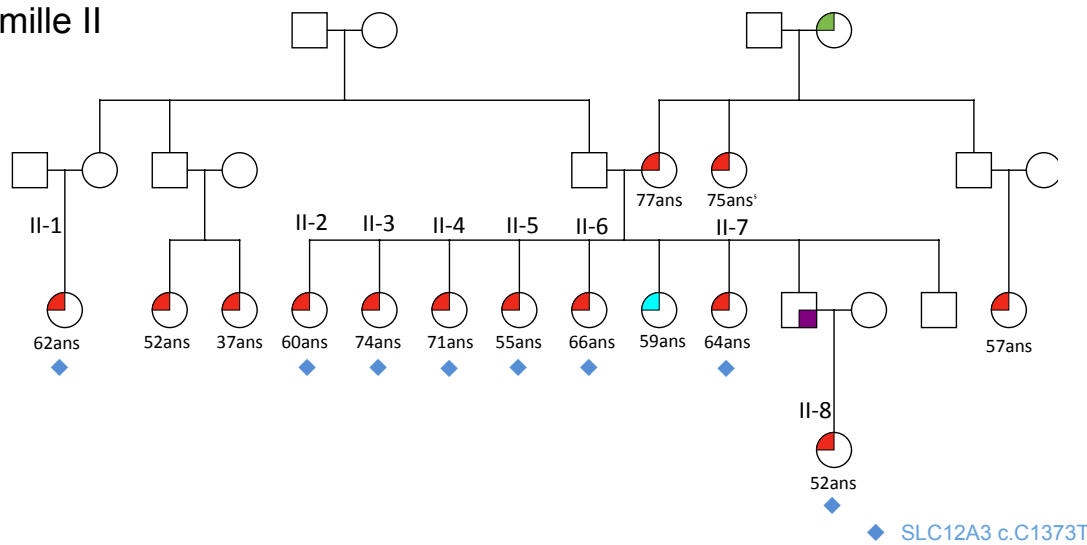
### 3.4.2 Famille II

La famille II est composée d'une grande fratrie de sept sœurs et deux frères, dont six cas de CS et un cas de cancer de la prostate chez un des frères. La fille de ce dernier a aussi développé un CS, tout comme des cousines de cette fratrie. Au total



nous avons effectué le WES de 8 membres atteints de cette famille, identifiés dans la Figure 23.

Famille II



| Individu | Age au diagnostic | Sous type histologique              | Grade | Statuts récepteurs hormonaux | Statut HER2 | Évolution de la maladie           |
|----------|-------------------|-------------------------------------|-------|------------------------------|-------------|-----------------------------------|
| II-1     | 62                | Carcinome lobulaire infiltrant      | I     | ND                           | ND          | Radiothérapie                     |
| II-2     | 60                | Carcinome canalaire invasif         | I     | RE+<br>PR+                   | HER2-       | Radiothérapie                     |
| II-3     | 74                | Carcinome lobulaire invasif         | II    | RE+<br>RP+                   | HER2-       | Ablation + chimio + radiothérapie |
| II-4     | 71                | Adénocarcinome canalaire infiltrant | II    | RE+<br>RP-                   | HER2-       | Ablation + chimio + radiothérapie |
| II-5     | 55                | Carcinome canalaire invasif         | I     | RE+<br>RP-                   | HER2+       | Radiothérapie                     |
| II-6     | 66                | Carcinome canalaire invasif         | II    | RE+<br>RP+                   | HER2-       | Radiothérapie                     |
| II-7     | 64                | Carcinome lobulaire infiltrant      | II    | RE+<br>RP+                   | HER2-       | Ablation des deux seins           |
| II-8     | 52                | Adénocarcinome infiltrant mucineux  | II    | RE+<br>RP+                   | HER2-       | ND                                |

**Figure 23** : Arbre généalogique de la famille II. Les individus en rouges sont atteints de CS, en vert cancer du côlon, et en bleu turquoise de cancer de la vessie. Les individus séquencés en exome et leurs caractéristiques cliniques sont présentés dans le tableau

Comme précédemment, toutes les combinaisons de co-ségrégation de variants ont été soumises aux filtres, et pas seulement les variants partagés par tous les membres. Nous avons identifié au total 7004 variants partagés par l'ensemble des huit membres atteints séquencés.

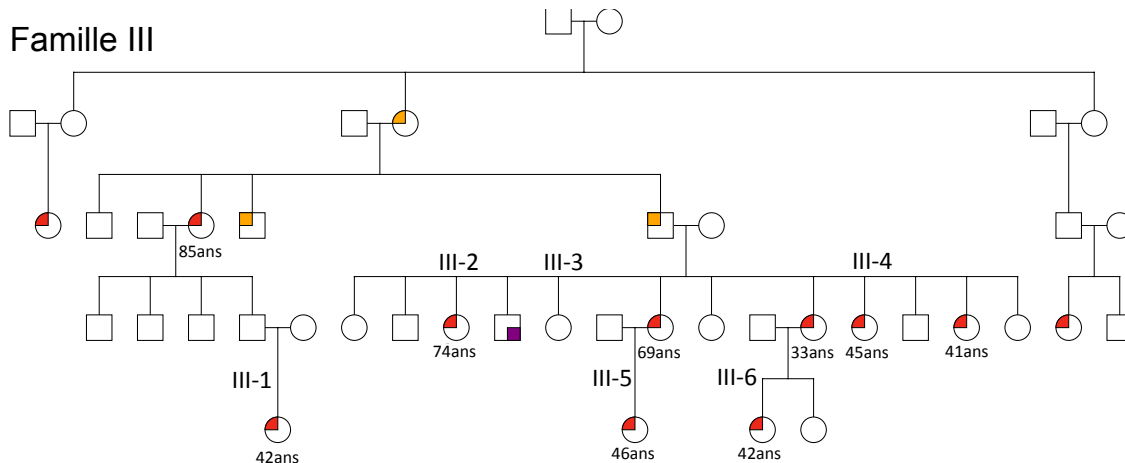
Au final, seul un variant dans le gène *SLC12A3* a passé les filtres, et a été identifié chez tous les individus (Tableau 11). Ce gène code pour un cotransporteur rénal sensible au chlorure de sodium et thiazidique qui est important pour l'homéostasie des électrolytes. Ce cotransporteur assure la médiation de la réabsorption du sodium et du chlorure dans le tubule alvéolaire distal. Des variants délétères de ce gène entraînent le syndrome de Gitelman, une maladie rénale, mais n'a pas de lien connu avec la progression tumorale. La présence de ce variant chez les individus de cette famille a été confirmée par séquençage Sanger (Figure supplémentaire 2).

| Gene    | Chromosome-Position | Nucleotide | Acide aminé | Type      | Polyphen Score prediction | SIFT Score prediction | Critère ACMG                        | Membres des familles |   |   |   |   |   |   |   | Fréquence GnomAD | ClinVar   |   |
|---------|---------------------|------------|-------------|-----------|---------------------------|-----------------------|-------------------------------------|----------------------|---|---|---|---|---|---|---|------------------|-----------|---|
|         |                     |            |             |           |                           |                       |                                     | 1                    | 2 | 3 | 4 | 5 | 6 | 7 | 8 |                  |           |   |
| SLC12A3 | chr16:56879579      | c.1373C>T  | p.T457M     | Faux sens | Probably damaging (0,99)  | Deleterious (0.03)    | Signification incertaine (PM2, BP4) | +                    | + | + | + | + | + | + | + | +                | 1.194e-05 | - |

**Tableau 11:** SNVs et indels délétères identifiés chez les individus atteints de la famille II.

### 3.4.3 Famille III

La famille III se compose d'un grand nombre de cancers du sein déclarés à un âge jeune. Nous avons effectué le WES de deux cousines III-5 et III-6 ayant développé un cancer du sein à 46 et 42 ans, et de leurs tantes III-2 et III-4 l'ayant déclaré respectivement à 74 et 45 ans tandis que III-3 n'en a pas développé et a été considéré comme témoin négatif. Nous avons également séquencé III-1, une cousine au second degré ayant eu un cancer du sein à 42 ans (Figure 24).



| Individu | Age au diagnostic | Sous type histologique         | Grade | Statuts récepteurs hormonaux | Statut HER2 | Évolution de la maladie                        |
|----------|-------------------|--------------------------------|-------|------------------------------|-------------|--|
| III-1    | 42                | ND                             | ND    | ND                           | ND          | ND   |
| III-2    | 74                | Carcinome lobulaire infiltrant | II    | RE+<br>RP+                   | HER2-       | Chirurgie + radiothérapie + hormono            |
| III-3    | Sain              | -                              | -     | -                            | -           | -  |
| III-4    | 45                | Carcinome canalaire infiltrant | II    | ND                           | ND          | Chirurgie + radio + chimio + hormono           |
| III-5    | 46                | Carcinome canalaire infiltrant | II    | RE+<br>RP+                   | HER2+       | Chirurgie partielle + radio + chimio + hormono |
| III-6    | 42                | Carcinome canalaire infiltrant | II    | RE+<br>RP+                   | HER2+       | Chirurgie partielle + radio + chimio + hormono |

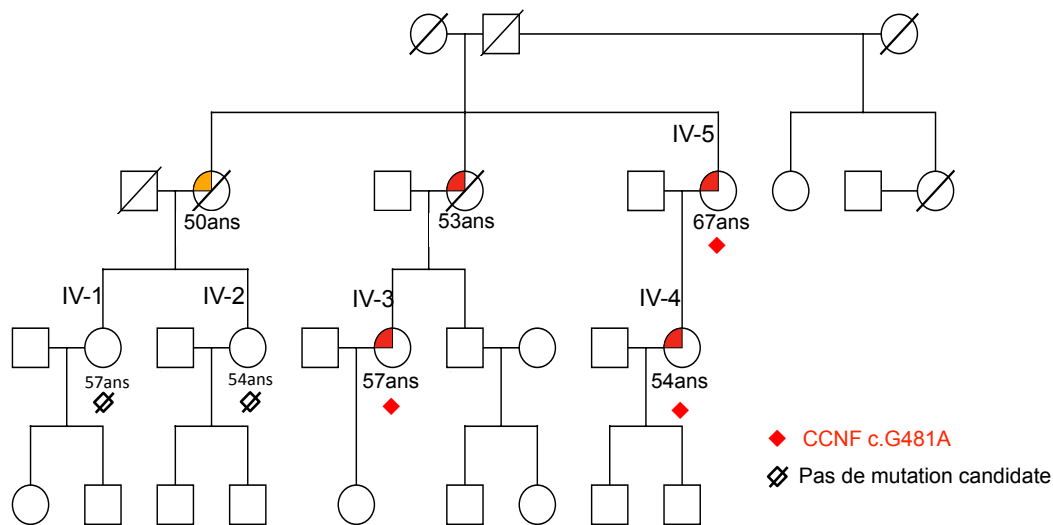
**Figure 24** : Arbre généalogique de la famille III. Les individus en rouges sont atteints de CS, en orange cancer de l'ovaire, et en violet de cancer de la prostate. Les individus séquençés en exome et leurs caractéristiques cliniques sont présentés dans le tableau

Nous avons détecté 353 variants partagés par les membres atteints de la famille. Ce total est faible en raison de l'exclusion des variants non présents chez le contrôle négatif III-3. Malgré le nombre de cas important de cette famille, nous n'avons pas pu identifier de variants passant les filtres.

Afin d'augmenter le nombre de candidats, nous avons refait l'analyse sans le témoin III-3 en faisant l'hypothèse qu'il pouvait être porteur sans développer la maladie (pénétrance incomplète). Nous avons obtenu un total de 6600 variants partagés par les membres atteints mais nous n'avons pas identifié de variant pertinent.

### 3.4.4 Famille IV

Dans une quatrième famille, on dénombre au total 4 cancers du sein et un cancer de l’ovaire. Nous avons séquencé l’exome des 3 membres ayant développé un cancer du sein IV-3, IV-4 et IV-6. On note également dans cette famille IV la présence de deux sœurs IV-1 et IV-2 n’ayant pas développé de cancer du sein. Les caractéristiques phénotypiques des individus séquencés sont présentées dans la Figure 25.



| Individu | Age au diagnostic | Sous type histologique | Grade | Statuts récepteurs hormonaux | Statut HER2 | Évolution de la maladie |
|----------|-------------------|------------------------|-------|------------------------------|-------------|-------------------------|
| IV-1     | Sain              | -                      | -     | -                            | -           | -                       |
| IV-2     | Sain              | -                      | -     | -                            | -           | -                       |
| IV-3     | 57                | Carcinome infiltrant   |       |                              |             |                         |
| IV-4     | 54                | Carcinome infiltrant   |       |                              |             |                         |
| IV-5     | 67                | Carcinome infiltrant   |       |                              |             |                         |

**Figure 25** : Arbre généalogique de la famille IV. Les individus en rouges sont atteints de CS, en orange de cancer de l’ovaire. Les individus séquencés et leurs caractéristiques cliniques sont présentés dans le tableau

Après le filtrage des 9456 variants partagés par les trois membres affectés de cette famille IV-3, IV-4 et IV-5, nous avons identifié le variant candidat c.481G>A ; p.G161R dans le gène CCNF. Il présente strictement tous les critères pour passer les filtres (Tableau 12), c’est à dire qu’il est situé dans la partie codante du gène CCNF, que sa fréquence générale est nettement inférieure à 1% et même 0,1%, et

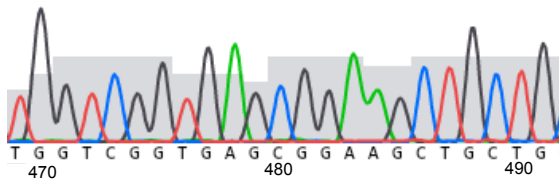
que les outils informatiques SIFT et Polyphen lui prédisent un effet délétère. Cependant il est classé « probablement bénin » selon les critères ACMG en raison de l'application du critère BS1 qui argue en faveur de son caractère bénin. En effet la fréquence allélique dans le population Est Asiatique atteints 0,000642 selon GnomAD, ce qui est supérieur à la fréquence maximale attendue (0,000146) pour ce gène. Nous l'avons tout de même considéré comme pertinent dans le contexte de familles BRCAx, étant donné que le CS n'est pas une maladie rare et que la fréquence de ce variant, même dans la population Est Asiatique, demeure très faible. Aucun autre variant candidat présent chez au moins un des trois membres n'a été identifié.

Dans un deuxième temps, nous avons pu avoir accès à des échantillons salivaires de IV-1 et IV-2, cousines au premier degré de IV-3 et IV-4, respectivement âgées de 57 et 54 ans et n'ayant pas développé de cancer du sein. Le séquençage Sanger de la région d'intérêt sur le gène *CCNF* a montré l'absence de variant chez ces individus sains, et a confirmé sa présence chez les trois individus atteints IV-3, IV-4 et IV-5 (Figure 26). Le variant de *CCNF* co-ségrège donc bien avec les individus atteints de la famille IV.

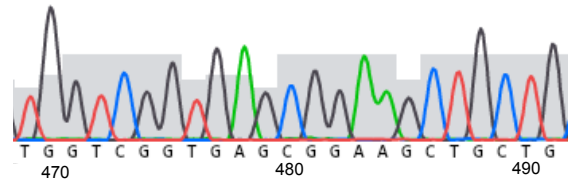
| Gene            | Chromosome-Position | Nucleotide | Acide aminé | Type      | Polyphen Score prediction | SIFT Score prediction | Critère ACMG                         | Membres des familles |   |   |   |   | Fréquence (GnomAD) | ClinVar |
|-----------------|---------------------|------------|-------------|-----------|---------------------------|-----------------------|--------------------------------------|----------------------|---|---|---|---|--------------------|---------|
|                 |                     |            |             |           |                           |                       |                                      | 1                    | 2 | 3 | 4 | 5 |                    |         |
| CCNF (cyclin F) | 16:2437263          | c.481G>A   | p.G161R     | Faux sens | Probably damaging (0.996) | Deleterious (0)       | Probablement bénigne (PP3, BS1, BP1) | -                    | - | + | + | + | 5.229e-05          | -       |

**Tableau 12** : SNVs et indels délétères identifiés chez les individus atteints de la famille IV.

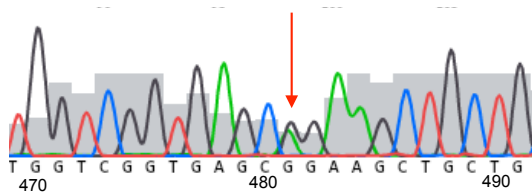
CCNF : c.G481A  
IV-1



IV-2



IV-3



**Figure 26** : Séquençage Sanger de la région d'intérêt du gène *CCNF* confirme la présence du variant candidat c.481G>A dans le gène *CCNF* chez l'individu atteint IV-3 et montre son absence chez les individus sains IV-1 et IV-2.

Le variant c.481G>A ; p.G161R identifié dans le gène *CCNF* a particulièrement retenu notre attention d'un point de vue fonctionnel. Le gène *CCNF* est un gène suppresseur de tumeur encodant la cycline F. Comme toutes les cyclines, elle joue un rôle important dans la régulation du cycle cellulaire, en induisant un arrêt en G2 et en inhibant la mitose [282], mais également dans le maintien de l'intégrité du génome au fil des divisions [283, 284]. Il a été montré *in vitro* qu'une absence de *CCNF* dans des cellules provoque une forte accumulation de variants due à une instabilité génomique [285].

De façon intéressante, un autre variant candidat dans le gène *CCNF* - c.1592T>G / p.L531R - avait été identifié dans une famille BRCAx dans une précédente étude [116]. Sa fréquence faible dans la population ( $2.51e-4$ ), les prédictions de pathogénicité des outils SIFT, Polyphen, et sa ségrégation avec les individus atteints de la famille en faisaient un candidat pertinent. La présence d'une deuxième famille BRCAx comportant un variant candidat dans *CCNF* est un élément important en faveur d'une implication dans la prédisposition au CS.

### 3.5. Implications pour l'entreprise

D'un point de vue entrepreneurial, l'approche d'exome chez les individus BRCAx ne paraît pas pertinente pour améliorer le conseil génétique et le rendre plus documenté. En effet, malgré des familles extensives, nous avons identifié peu de variants rares, dont la pathogénicité est difficilement vérifiable, car ils sont localisés dans des gènes dont l'implication dans le CS n'est pas connue. Sans des études fonctionnelles approfondies, nous ne pouvons faire que des hypothèses sur les variants candidats.

Nous restons donc dans les recommandations actuelles préconisant la réalisation de panels de gènes.

## 4. Discussion

Dans la présente étude, nous nous sommes demandé dans quelle mesure le WES pouvait contribuer à la découverte des prédispositions des familles BRCAx. L'application des filtres de sélection des variants a permis de déterminer au total 6 variants candidats à haut risque co-ségrégeant avec la maladie dans les familles I, II et IV. Les résultats présentés dans cette thèse ne nous permettent pas de démontrer l'implication d'aucun de ces variants dans les prédispositions génétiques existantes dans ces familles, en l'absence de validation par des tests fonctionnels. Or, le cadre de cette thèse CIFRE axée sur les analyses de génétique se prêtait peu à ce type d'expérimentations de biologie cellulaire complexes et longues. Les expérimentations fonctionnelles pourraient toutefois être menées dans le cadre de collaborations.

Cependant, la sélection stricte des variants et leur association avec les individus atteints nous ont permis d'émettre des hypothèses sur l'implication de ces variants sur l'incidence du CS dans ces familles BRCAx.

Dans la famille I, où plusieurs cas de CS sont présents dans les branches paternelles et maternelles, nous avons identifié des variants candidats dans les gènes *CHEK2* et *TYRO3* du côté paternel, et dans les gènes *HISTH1C* et *TPH1* du côté maternel. La délétion dans *CHEK2* identifiée chez les deux membres affectés I-7 et I-8 de la famille I est un facteur de risque intermédiaire au CS connu, augmentant par 3,11 le risque de CS au cours de la vie [277]. Cependant, ce variant ne peut expliquer la totalité de la prédisposition dans cette famille. Il n'est porté que

par les filles atteintes I-7 et I-8 et leur père I-3, mais pas par les tantes paternelles et maternelles également atteintes I-1 et I-2. L'âge précoce d'apparition du CS (25 et 39 ans) chez les sœurs I-7 et I-8 suggère un risque particulièrement accru chez elles. À la lumière de nos résultats, nous pouvons émettre l'hypothèse que ce risque soit dû à une accumulation des 4 variants chez chacune d'elles. En effet l'âge d'apparition du CS chez les 4 membres de la génération précédente était en moyenne de 60 ans, appuyant l'hypothèse de l'agrégation de plusieurs facteurs de risque. La troisième sœur I-6, aujourd'hui âgée de 50 ans, est saine et ne porte aucun des variants à risque, ce qui renforce l'hypothèse de leur pathogénicité.

Nous avons identifié des variants candidats dans les gènes *SLC12A3* et *CCNF* respectivement chez les individus atteints des familles II et IV.

En revanche, aucun variant pertinent n'a été retenu pour expliquer la prédisposition au CS dans la famille III.

Aucun de ces variants n'est répertorié dans la base de données ClinVar. En dépit du fait qu'ils soient considérés comme des VUS par la classification ACMG (sauf *CCNF* qui est « probablement bénin ») certains éléments sont en faveur de leur rôle dans la prédisposition de ces familles BRCAx.

Tout d'abord, ces variants co-ségrègent avec le CS chez au moins trois individus affectés dans chaque famille, tout en étant absents, chez les individus contrôle non atteints. De plus, leur fréquence extrêmement faible dans la population, de l'ordre de  $10^{-6}$  pour le variant *HISTH1C*, et de  $10^{-5}$  pour les variants *TYRO3*, *SLC12A3* et *CCNF*, met en évidence leur potentiel effet délétère. Enfin, leurs scores de pathogénicité calculés par les algorithmes de prédiction Polyphen-2 et SIFT suggèrent fortement leur effet délétère sur le produit du gène pouvant générer une situation d'haploinsuffisance.

Un autre critère qui entre en jeu est la fonction biologique des gènes candidats. Nous savons que la quasi-totalité des gènes de prédisposition au cancer connus se trouvent dans des gènes impliqués dans l'intégrité génomique / réparation de l'ADN [286]. Ainsi, se concentrer en priorité sur les variants délétères dans les gènes de réparation de l'ADN est une approche légitime. Toutefois, les oncogènes peuvent également être pertinents, car ils peuvent être suractivés par des variants activateurs, comme il a été démontré récemment avec l'identification de variants

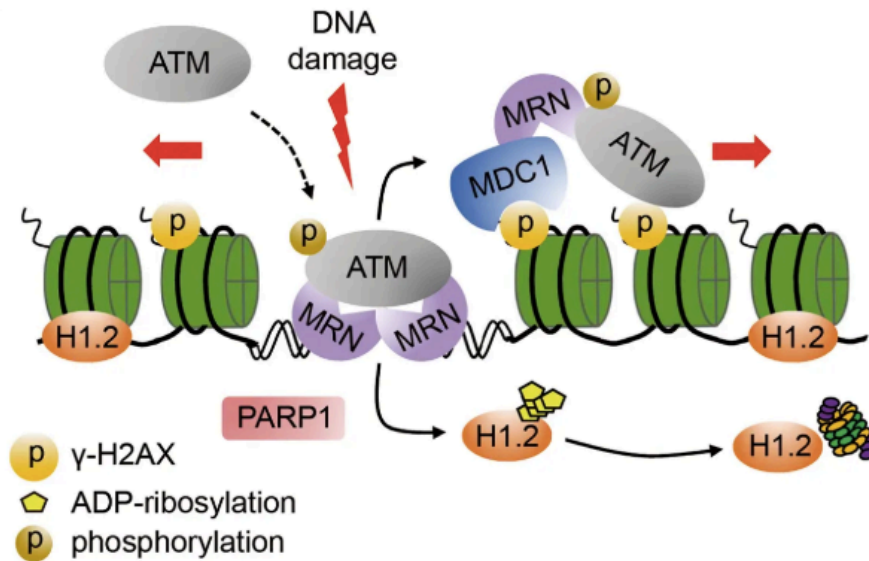


activateurs de l'oncogène *NCK1* prédisposant au BC dans des familles *BRCA1/BRCA2* [287].

Ainsi, le rôle suppresseur de tumeur / réparation de l'ADN des gènes, *HIST1H1C* et *CCNF* et oncogénique de *TYRO3* et *TPH1* en font des candidats de susceptibilité au CS particulièrement intéressants.

Le gène *HIST1H1C* encode l'histone H1.2 qui interagit avec l'ADN entre les nucléosomes et joue un rôle dans la compaction de la chromatine. Comme nous l'avons vu dans la partie résultat, il est impliqué dans la recombinaison homologe, mécanisme essentiel pour la réparation des cassures d'ADN double brin.

En réponse à un dommage à l'ADN, il est instantanément ubiquitinylé par PARP sur son domaine C-terminal pour être dégradé par le protéasome, permettant ainsi le recrutement de *ATM* sur le site de cassure (Figure 27 et Figure 28) [278]. Ces données nous mènent à l'hypothèse que notre variant candidat, c.592G>T ; p.A198S, situé à l'extrémité C-terminal de la protéine, empêche la dégradation de l'histone H1.2 et ainsi affecte la voie RH de réparation de l'ADN. L'altération de la voie RH est un mécanisme fréquent dans les CS héréditaires, notamment par le biais de variants délétères de *BRCA1* et *BRCA2*. L'inactivation de cette protéine chez ces individus pourrait provoquer un défaut de réparation de l'ADN favorisant l'apparition de cancer. La confirmation d'un défaut de la voie RH chez les individus porteurs pourrait être pertinente d'un point de vue thérapeutique, car les patients dont le CS est associé à un dysfonctionnement de la réparation de l'ADN RH peuvent bénéficier de traitements ciblés tels que les inhibiteurs de la PARP.



**Figure 27:** Dynamique de la régulation de ATM par l’histone H1.2. Ce dernier, en réponse aux dommages à l’ADN, est phosphorylé par PARP pour être dégradé par le protéasome, permettant le recrutement d’ATM au site de cassure, qui pourra engager la réparation de l’ADN en phosphorylant un grand nombre de substrats.

| Biological Process (GO) |  |                   |                      |
|-------------------------|--|-------------------|----------------------|
| GO-term                 | description  | count in gene set | false discovery rate |
| GO:0031497              | chromatin assembly   | 11 of 154         | 5.70e-15             |
| GO:0006325              | chromatin organization                                     | 14 of 683         | 1.70e-13             |
| GO:0071158              | positive regulation of cell cycle arrest                   | 8 of 85           | 9.84e-12             |
| GO:0006977              | DNA damage response, signal transduction by p53 class m... | 7 of 59           | 7.11e-11             |
| GO:0071480              | cellular response to gamma radiation                       | 6 of 27           | 8.07e-11             |

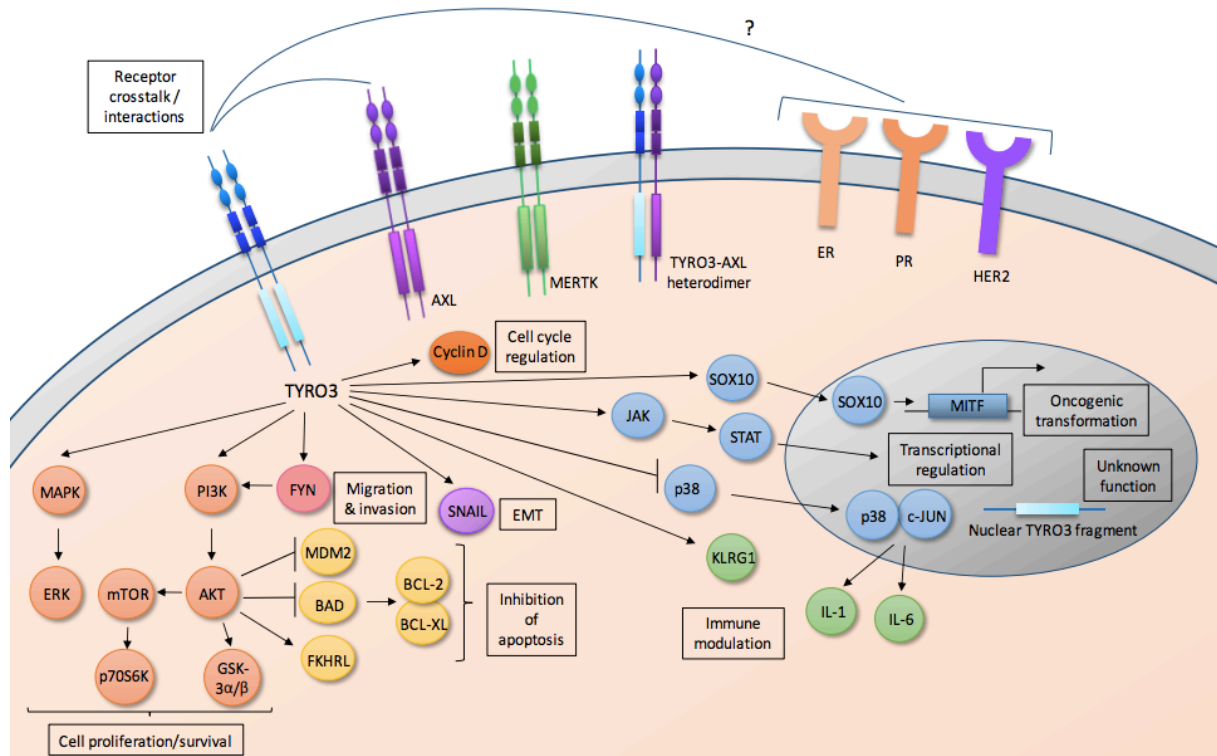
**Figure 28:** Enrichissement fonctionnel des protéines formant le réseau d’interaction de l’histone H1.2. Les processus biologiques significativement associés sont l’organisation de la chromatine et la réparation des dommages à l’ADN.

Le seul variant identifié dans la famille IV est localisé dans le gène *CCNF* et est également impliqué dans le maintien de la stabilité du génome. L’absence de *CCNF* *in vitro* dans des cellules provoque une forte accumulation de variants due à une instabilité génomique [285]. Nous pouvons faire l’hypothèse qu’une perte de fonction de *CCNF* soit à l’origine d’une instabilité génétique progressive chez les individus de la famille IV. Afin de valider cette hypothèse, nous avons entrepris de mesurer le taux mutationnel tumoral des patients de la famille IV par WES mais le séquençage a

malheureusement été trop peu qualitatif en raison de la trop faible quantité de matériel et n'a pas permis de détecter une éventuelle instabilité génomique.

De manière intéressante, un variant du gène du *CCNF* avait déjà été identifié comme facteur de risque candidat dans une famille BRCAX [116]. L'accumulation de deux variants rares et délétères chez le membre affecté de deux familles BRCAX indépendantes est un élément important en faveur de son rôle prédisposant. À ce jour, environ 110 familles BRCAX ont été séquencées, ce qui a permis d'identifier de nombreux variants candidats. Il en résulte que les prédispositions génétiques des familles BRCAX semblent hautement spécifiques à chaque famille [288, 289]. En effet, mis à part notre candidat *CCNF*, déjà identifié dans une précédente étude de famille BRCAX, peu de gènes ont été identifiés dans plus d'une famille BRCAX indépendante. Lorsque cela a été le cas, notamment pour les quatre gènes *XRCC2* [290], *MAPKAP1* [289], *FANCM* [115], *RINT1* [291], il s'en est suivi des études d'association cas témoin à grande échelle qui ont pu démontrer leur association au CS. Une étude d'association plus vaste (de type GWAS), sur un grand nombre d'individus atteints et contrôles, pourrait donc être pertinente pour démontrer l'implication de *CCNF* en tant que facteur de prédisposition au CS.

Quand au gène candidat *TYRO3*, il encode un récepteur tyrosine kinase transmembranaire ayant un rôle oncogénique reconnu en activant les voies PI3K de prolifération et de survie cellulaire. Les données de littérature indiquent qu'il est surexprimé dans de nombreux cancers et impliqué dans la prolifération des cellules de cancer du sein ER+. Le variant identifié c.2401G>C/ p.D801H, identifié chez les membres atteints de la famille I, est localisé dans son domaine kinase intracellulaire dans l'exon 19. Il pourrait avoir comme effet d'augmenter l'activité kinase, entraînant notamment une suractivation des voies PI3K de prolifération tumorale (Figure 29).



**Figure 29** : Les voies de signalisation *TYRO3*. *TYRO3* régule les voies importantes pour la prolifération et la survie des cellules (par exemple, la signalisation PI3K / AKT / mTOR et MAPK / ERK), la migration et l'invasion, la transition épithélio-mésenchymateuse, la régulation du cycle cellulaire, la régulation de la transcription, la modulation immunitaire et la transformation oncogénique. Source : Smart et al., 2018 [279].

Au total, notre approche de séquençage WES sur des familles BRCAx nous a permis d'identifier des candidats pour deux familles sur les quatre étudiées. La méthode WES appliquée sur les familles à risque s'avère efficace dans les cas où la prédisposition est héritée de manière Mendélienne, par un allèle rare autosomique dominant, et ayant une pénétrance modérée à forte, comme c'est probablement le cas dans les familles I et IV.

En revanche, notre approche WES a échoué à identifier des candidats pertinents dans les familles II et III. Plusieurs hypothèses peuvent être envisagées pour expliquer l'absence de candidats dans ces familles, pointant les limites du WES en recherche clinique. L'une des hypothèses est que les altérations en cause soient dans les parties non codantes du génome non couvertes par le WES. En effet, les études GWAS ont mis en évidence qu'une partie des SNV associés aux CS se

trouvent dans les parties non codantes des gènes indétectables par WES. De plus, Une étude a récemment montré que les variants des régions régulatrices non codantes expliquent davantage l'héritabilité de 11 maladies courantes que les variants codants [292]. Les variants des régions régulatrices constitutionnels et somatiques ont été impliqués en tant que *drivers* du cancer [293, 294]. De ce fait le génome non codant ne doit plus être négligé dans la recherche d'une sensibilité génétique au cancer. Par exemple, les variants du promoteur TERT sont parmi les variants oncogéniques non-codants les plus fréquents ; ils entraînent une augmentation de l'activité de la télomérase et des télomères longs, ce qui favorise l'immortalisation des cellules cancéreuses [295]. Des variants des promoteurs TERT somatiques et constitutionnels ont été signalés dans un certain nombre de cancers [296, 297] ; il a été démontré qu'ils peuvent être impliqués dans le risque de cancer familial [297]. D'autres variants non codants associés au cancer ont été découverts ; par exemple, l'allèle mineur de rs55705857A>G, qui est proche du gène *CCDC26*, confère une augmentation du risque relatif d'environ 6 fois de développer un gliome, lorsqu'il est accompagné d'un variant de *IDH1* ou *IDH2* [298].

Malgré ces limitations, nos résultats renforcent l'approche du WES basé sur les familles pour la découverte de nouvelles prédispositions génétiques au CS. L'implication des gènes candidats *TYRO3*, *HISTH1C*, *TPH1* et *CCNF* dans le risque de CS doit être confirmée par des tests fonctionnels. La génération de lignées cellulaires exprimant la protéine mutée permettrait de déterminer leurs effets sur la prolifération des cellules, la stabilité du génome, et sur la réparation de l'ADN par la voie RH pour le mutant *HISTH1C*.

S'ils viennent à être confirmés par des tests fonctionnels, l'identification de ces facteurs de prédisposition génétique pourrait avoir une implication pour le conseil génétique de ces familles, notamment pour permettre une meilleure prévention des risques des futurs cas dans ces familles via un dépistage du CS précoce et régulier chez les individus porteurs sains. Ce travail contribue également à augmenter le spectre des gènes de prédisposition au CS, afin d'améliorer le dépistage et à la prévention des risques de CS familiaux. L'identification de ces candidats apporte également de nouvelles données sur les mécanismes biologiques de la carcinogenèse du sein.

## 5. Conclusion

Le WES appliqué aux familles BRCAx s'est montré un outil efficace pour l'identification de SNPs et indels candidats susceptibles de prédisposer aux CS familiaux. Nous avons pu identifier un facteur de prédisposition connu dans le gène *CHEK2* dans la famille I, et des variants candidats nouveaux dans les gènes *HIST1H1C*, *TYRO3*, *TPH1*, *SLC12A3* et *CCNF*. Ces derniers pourraient avoir un impact sur les produits de ces gènes (par exemple une effet délétère générant une haploinsuffisance), favorisant la prévalence du CS chez les porteurs. L'objectif visé était d'améliorer le conseil génétique de ces familles pour une prise en charge plus personnalisée des risques, notamment pour les individus porteurs mais n'ayant pas encore déclaré la maladie. Cependant, des études fonctionnelles sont nécessaires, dans le cadre d'une future collaboration, afin de démontrer l'implication de ces variants dans la tumorigenèse du sein. Nous n'avons pas pu montrer un avantage direct du WES en clinique pour ces familles.

# Chapitre 4 : Apport du séquençage de l'exome pour la personnalisation des traitements en oncologie

## 1. Introduction

Le NGS donne un nouvel aperçu sur les bases moléculaires du cancer. Le profilage moléculaire souligne qu'un même type de tumeur peut contenir des sous-groupes moléculaires variables avec des propriétés moléculaires différentes. L'association de certains variants et de l'activation de voies de signalisation conduit à l'identification de cibles thérapeutiques. Ces dernières années, sur la base de ce concept, la génomique clinique s'est appliquée au traitement du cancer [299]. Dans de nombreux types de cancer, les variants dits *driver* peuvent être ciblés par des traitements, conduisant à un taux de réponse élevé et à un meilleur taux de survie. Dans certains types de cancer comme le cancer colorectal, le cancer du poumon et le mélanome, le profilage moléculaire est devenu une pratique courante pour rechercher des variants génétiques ciblables [300-302]. Ces découvertes ont mis en pratique concrètement le concept de médecine de précision où des essais cliniques ont commencé à évaluer l'efficacité des thérapies moléculaires ciblées sur différents types de tumeurs en fonction de leurs caractéristiques génétiques spécifiques [303].

Actuellement, la plupart des tests de profilage moléculaire réalisés sont effectués par un grand panel NGS somatique et/ou constitutionnel [304], comme c'est le cas notamment des tests proposés par les entreprises américaines FoundationOne et Caris, leaders mondiaux dans ce domaine. L'examen des profils génétiques somatiques identifie de très nombreux VUS présents dans l'ADN tumoral. Or, la possibilité de cibler de tels variants n'est pas connue à l'heure actuelle. La complexité des analyses de données de NGS a empêché de nombreux hôpitaux et laboratoires de l'utiliser régulièrement à grande échelle. Actuellement, très peu de laboratoires le mettent en œuvre en raison de la complexité et du temps requis pour l'analyse bioinformatique.

Dans le cadre de la R&D de l'entreprise Prenostics, notre travail a tout d'abord eu pour but de déterminer la faisabilité et l'utilité de l'analyse de l'exome tumoral en routine. Nous avons tenté de déterminer si l'exome tumoral peut modifier des

traitements en oncologie par rapport au séquençage ciblé de panel, actuellement utilisé.

Afin de répondre à ces questions, nous avons mené une étude pilote sur une cohorte prospective de 35 individus atteints de cancers avancés, réfractaires à une première ligne de traitement et pour lesquels l'approche classique de séquençage de panels a échoué à personnaliser le traitement. Nous avons donc entrepris l'exome somatique et constitutionnel de ces individus en tentant d'identifier un maximum de marqueurs pouvant influencer sur la stratégie thérapeutique.

## 2. Matériel et Méthodes

### 2.1 Cohorte cancer

Dans cette étude multicentrique, nous avons obtenu des échantillons tumoraux et sanguins de 35 individus présentant des tumeurs malignes avancées provenant de divers organes. Les critères d'inclusion comprenaient un diagnostic histologique de malignité, une tumeur biopsiable et devaient avoir une espérance de vie supérieure à quatre mois. Un consentement éclairé écrit a été obtenu des patients avant le prélèvement des échantillons appariés. Dans cette étude, seules les personnes ayant suffisamment de tumeur à partir d'échantillons fraîchement obtenus étaient éligibles. De plus, ces patients ont pour particularité d'être en situation réfractaire après échappement ou non réponse à au moins une première ligne de traitement.

### 2.2. Analyse des variants somatiques

Nous avons effectué l'analyse d'exome à la fois dans l'ADN des cellules de sang total et l'ADN tumoral pour déterminer les variants somatiques, c'est à dire uniquement présents dans l'échantillon de tumeur et non constitutionnellement. Ainsi, les variants somatiques ont été isolés des variants constitutionnels.

Nous avons utilisé la base de données sur les variants somatiques Cosmic v.64 pour classer chaque variant sélectionné en tant que variants «pathogènes», «pathogénicité probable», «pathogénicité inconnue» ou «bénins». Pour chaque variant détecté et annoté, nous avons retenu pour l'interprétation uniquement les variants annotés comme pathogènes ou probablement pathogènes. Certains variants



somatiques non répertoriés dans Cosmic ont été retenus uniquement s'ils étaient localisés dans un domaine critique de la protéine.

Nous avons filtré les variants candidats ayant une profondeur de couverture minimale de 10X et une fraction allélique dans l'ADN tumoral supérieure à 5%.

### 2.3. Analyse des variants constitutionnels

Nous avons en priorité concentré notre analyse constitutionnelle sur 52 gènes de prédisposition aux formes héréditaires de cancer. Nous avons filtré les variants candidats ayant une profondeur de couverture minimale de 10X.

Pour chaque variant constitutionnel identifié, nous avons déterminé sa fréquence dans la population générale à l'aide des bases de données de population GnomAD, EXAC et dbSNP et sa présence dans la bases de données ClinVar. Nous avons examiné la bibliographie disponible correspondante.

### 2.4. Détermination de la charge mutationnelle tumorale (TMB) par Mb

Les données de séquençage de l'exome entier ont été utilisées pour déterminer la charge de mutations tumorales par Mb. La TMB correspond au nombre total de variants faux sens et indels somatiques, divisé par le nombre de mégabases (Mb) des séquences cibles de la capture d'exome Nimblegen SeqCap EZ MedExome Enrichment (47Mb).

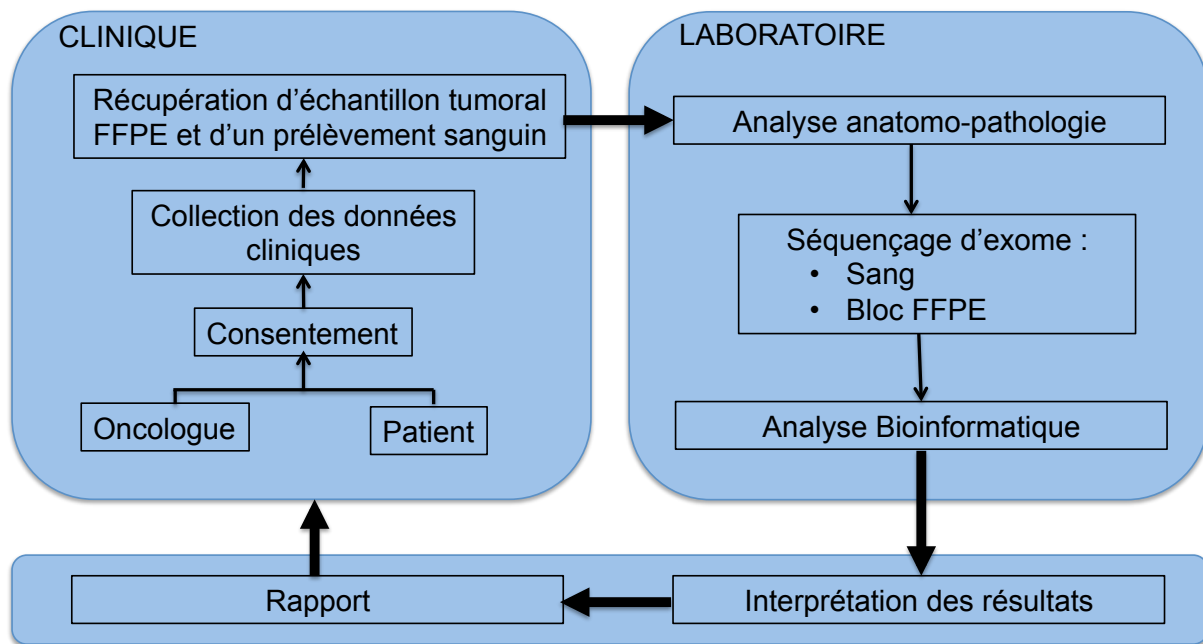
### 2.5. Analyse des signatures mutationnelles tumorales

Nous avons utilisé l'outil en ligne MuSiCa pour la caractérisation somatique des signatures mutationnelles présentes dans la liste des variants du patient [305]. Cet outil est basé sur le module *MutationalPatterns*, construit à partir du framework Shiny en langage R. Le module *MutationalPatterns* extrait les signatures *de novo* avec l'algorithme d'implémentation de la matrice de factorisation non-négative (NMF) qui se base sur plus de 10 000 échantillons tumoraux. À partir des données des SNV somatiques, cet algorithme extrait une série de profils mutationnels et les quantifie. Selon la base de données COSMIC, trente signatures mutationnelles ont été décrites dans quarante types de cancers.

## 3. Résultats

### 3.1. Mise en œuvre du test génétique

Nous avons mis en place les étapes d'un test génétique pour la caractérisation du profil moléculaire de tumeurs. Le test que nous avons développé consiste à séquencer en parallèle l'exome constitutionnel et l'exome tumoral à partir d'échantillons de tissus normaux et tumoraux appariés pour chaque individu. Nous avons extrait l'ADN constitutionnel à partir de prélèvements sanguins et l'ADN tumoral à partir de biopsies FFPE. De cette façon nous avons eu accès à l'exome tumoral et l'exome constitutionnel appariés pour chaque cas. En soustrayant les variants constitutionnels aux variants tumoraux, nous avons pu déterminer les variants somatiques tumoraux. La distinction des variants somatiques et de la lignée germinale à l'échelle de l'exome a permis d'identifier les variants *driver* de la tumeur, mais également les éventuelles prédispositions génétiques héréditaires présentes dans le génome constitutionnel. Cela nous a également permis de calculer la charge mutationnelle des tumeurs ainsi que leur signature moléculaire. L'interprétation des résultats en regard du contexte clinique du patient et des données de la littérature par une équipe pluridisciplinaire a donné lieu à la rédaction de rapport pour chaque cas (Figure 30).



**Figure 30:** Schéma des étapes du test de profilage moléculaire des tumeurs solides

Quarante-trois participants ont été inclus dans l'étude et ont signé un consentement éclairé donnant leur accord pour la réalisation des analyses d'exome.

À la suite de la biopsie de la tumeur, chaque spécimen a fait l'objet d'un examen pathologique afin d'évaluer la cellularité et le pourcentage de cellules tumorales. Les échantillons avec un contenu tumoral insuffisant ont été considérés non éligibles pour le séquençage. Au total, sur les 43 participants inclus dans l'étude, 35 (environ 81%) disposaient de tissu tumoral en quantité suffisante pour le séquençage.

Chacun de ces 35 patients avait un cancer au stade avancé avec un âge médian de 60 ans et une médiane de deux lignes de traitement antérieurs. La cohorte est composée de 21 femmes et 14 hommes. La distribution des tumeurs et les caractéristiques de base sont décrites dans le tableau 13.

Parmi les 35 patients éligibles, une pureté tumorale moyenne de 65% a été atteinte dans les échantillons FFPE et le séquençage d'exome de tous les échantillons a été réalisé avec succès. L'exome a été réalisé à une profondeur médiane de couverture de 158x pour les échantillons sanguins et 104x pour les échantillons FFPE. Chaque cas a généré plus de 8 000 variants de qualité expérimentale suffisante pour être analysés. Le nombre moyen de variants non filtrés détectés était de 8803 dans le

sang et de 9345 dans le tissu FFPE. Seuls les variants dans les gènes en lien avec le cancer ont été reportés, afin de permettre une interprétation clinique.

### 3.2 Profilage moléculaire des tumeurs par séquençage ciblé de panels et étude des CNV

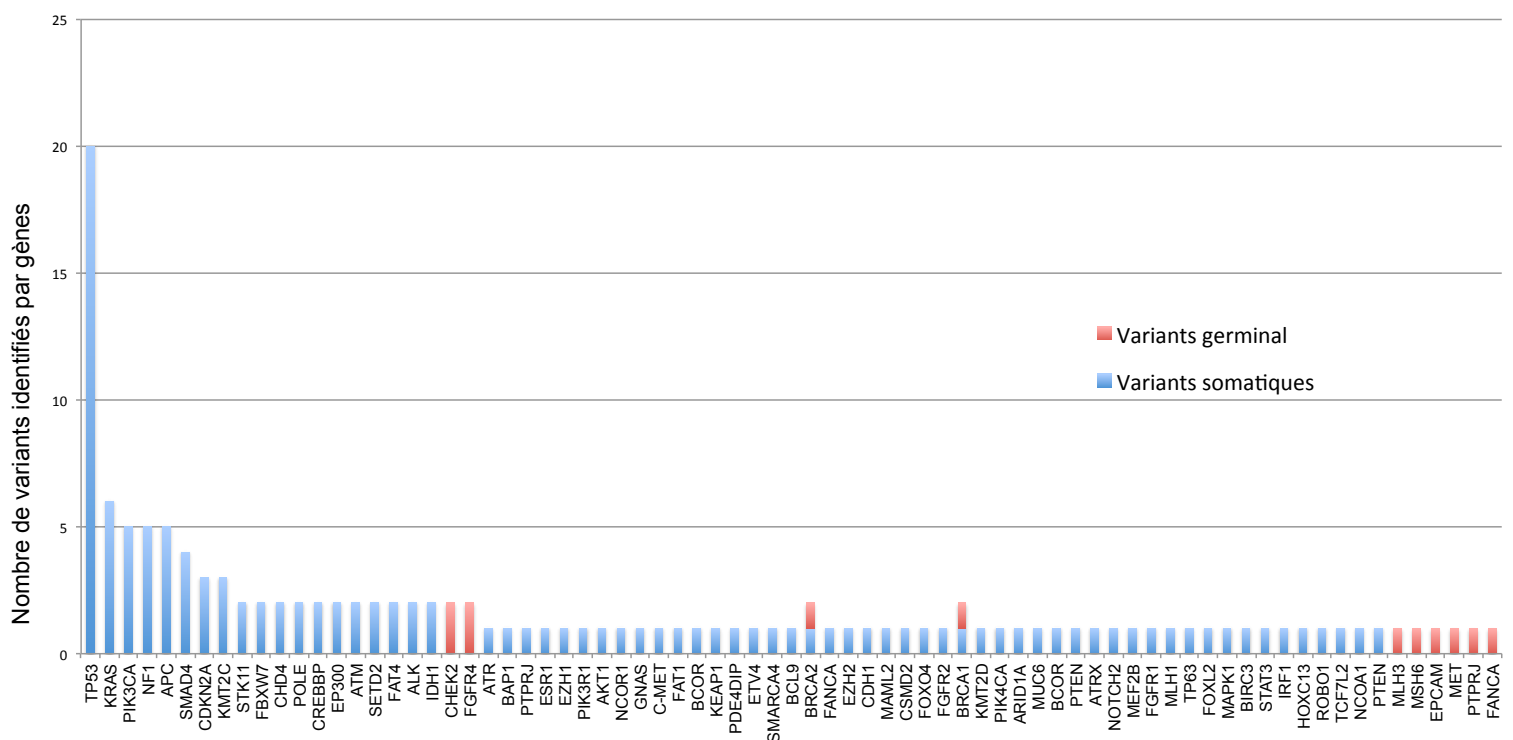
Préalablement aux analyses d'exome, les tumeurs ont été analysées par les méthodes conventionnelles de séquençage de panels de gènes et d'étude des CNV dans des hôpitaux ou dans des laboratoires spécialisés. Ces méthodes ont souvent permis pour certains individus d'identifier des altérations dans les gènes *drivers* les plus communs tels que *TP53*, *KRAS*, *PIK3CA*. Cependant, pour aucun de ces patients ces résultats n'ont permis de modifier le traitement, ni de prédire une réponse ou une résistance à une stratégie thérapeutique particulière.

Après l'échec d'une première ligne de traitement et le peu d'éléments moléculaires susceptibles de guider la stratégie thérapeutique, nous avons reçu pour chaque patient un échantillon de tumeur FFPE et un échantillon de sang afin de réaliser nos analyses d'exome.

### 3.2 Analyse de pattern mutationnels somatiques et constitutionnels

Afin de distinguer les variants potentiellement somatiques, nous avons sélectionné les variants ayant un nombre de *reads* significatif dans le tissu tumoral et non détectés dans le tissu sain. Le séquençage d'exome a identifié un total de 125 variants dans les gènes potentiellement *driver*, comprenant 113 variants somatiques et 12 variants de lignée germinale. Les variants somatiques ont été classés comme pathogènes (n = 51), possiblement pathogènes (n = 19), possiblement bénins (n = 4) et de signification clinique inconnue (n = 19). Le tableau 13 détaille les gènes mutés chez chaque patient. En prenant comme référence les études TCGA relatives à chaque type de cancer, notre étude a identifié des variants somatiques délétères dans des gènes *drivers* chez 33 individus sur 35 (94%). Les 12 variants de lignée germinale ont été identifiés dans 8 des 35 cas (22%) et classés comme pathogéniques (n = 7), possiblement pathogéniques (n = 1), et de signification clinique inconnue (n = 4).

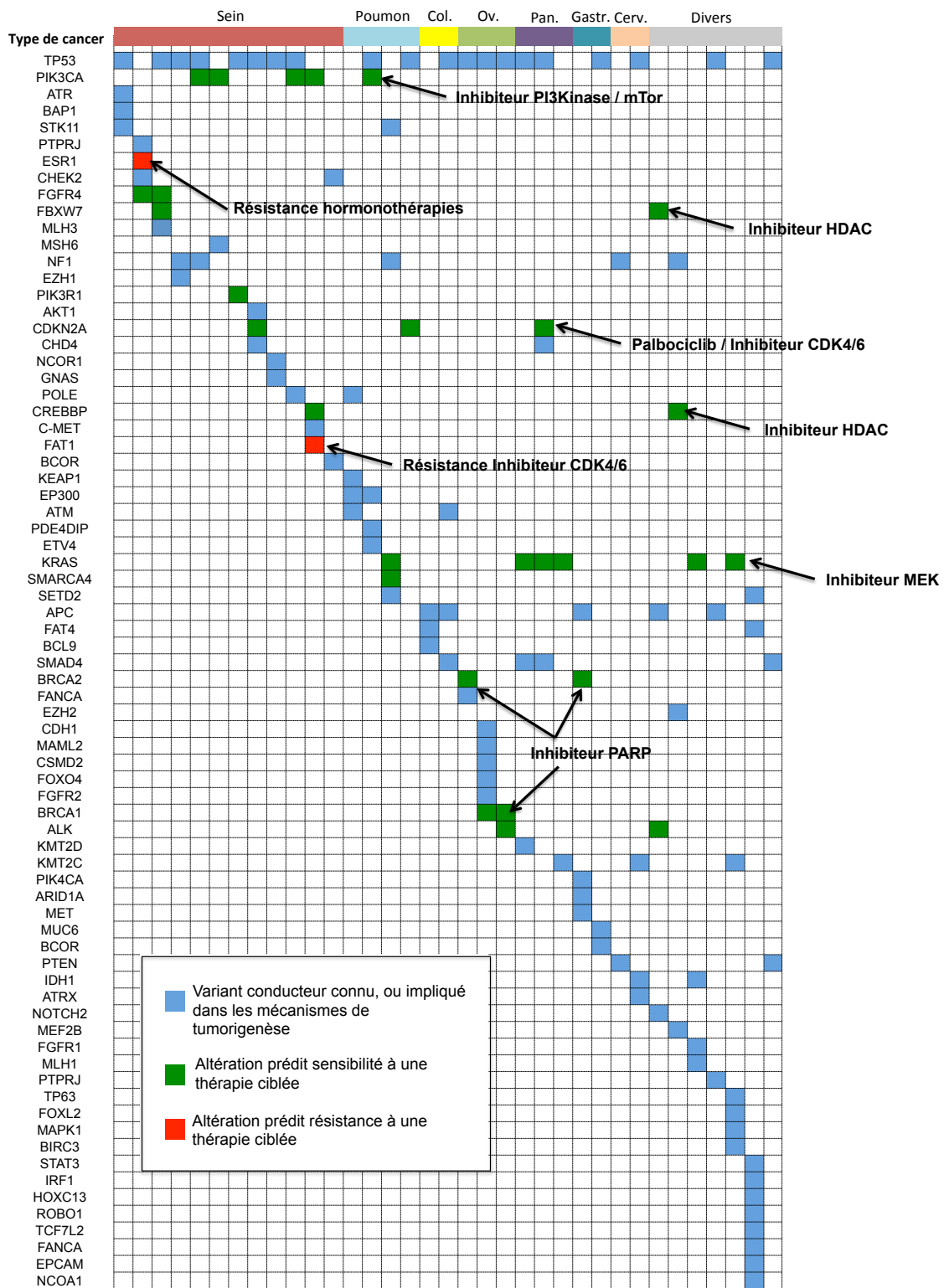
Le gène *TP53* était le gène le plus fréquemment muté, identifié dans 54% des cas (19/35) (Figure 31). Les variants de *PIK3C1* étaient également fréquents (6/35) et particulièrement dans les CS dans lesquels *PIK3C1* était altéré dans un tiers des cas. Des variants activateurs touchant le gène *KRAS* ont été détectés chez six patients (17%), dont trois cancers du pancréas. À l'inverse, les gènes suppresseurs de tumeur *APC* et *CDKN2A* étaient altérés dans respectivement cinq et trois patients (14% et 8% des cas). (Figure 31). Lorsque regroupés par fonction de gène, les voies de réparation des dommages de l'ADN, PI3K et MAPK-RAS étaient les plus souvent affectés.



**Figure 31:** Histogramme montrant les variants de la lignée somatique (en bleu) et germinale (en rouge) par séquençage d'exome de l'ADN tumoral de 35 tumeurs. Seuls les variants dans les gènes associés aux mécanismes du cancer ont été pris en compte.

Les résultats ont été analysés par une équipe pluridisciplinaire de biologistes, bioinformaticiens et médecins, permettant de déterminer les altérations ciblables de la tumeur, en prenant en compte la biologie de la tumeur et le contexte clinique, pouvant être partagés avec l'oncologue des patients (Figure 32). L'identification de ces variants et leur annotation avec la base de données oncoKB a permis d'identifier

des variants prédictifs indicateurs d'une sensibilité (*PIK3CA*, *CDKN2A*, *KRAS*, *FBXW7*, *CREBBP*, *SMARCA4*) ou d'une résistance (*FAT1*, *FBXW7*, *ESR1*) à certains traitements anticancéreux d'après l'étude de la littérature. Au total, la présence de SNVs permettant de donner des indications sur la réponse à un traitement approuvé par la FDA a été observée dans 18 cas (51% des cas).



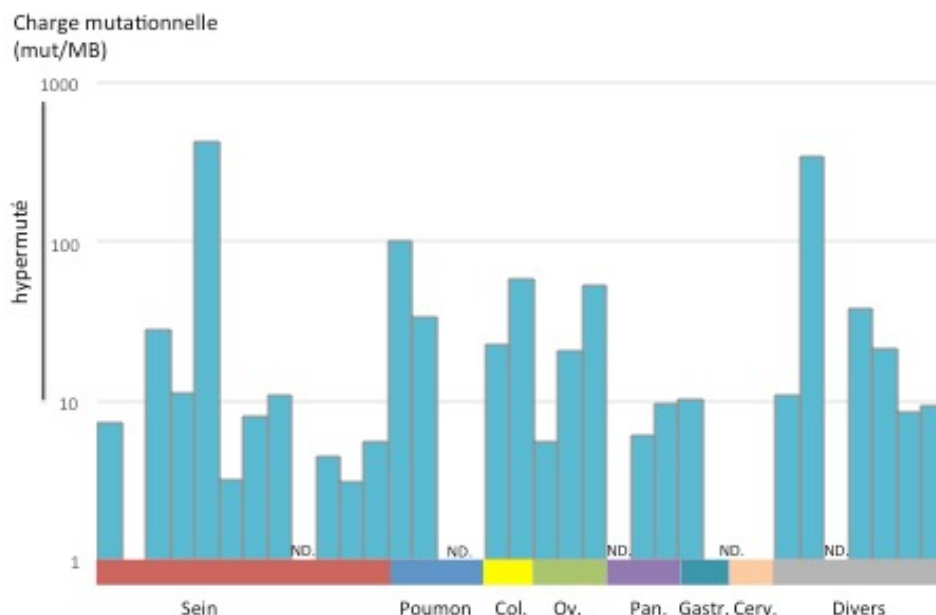
**Figure 32:** Analyse des variants actionnables et potentiellement ciblables, et identification des options thérapeutiques envisageables

### 3.3. Le calcul de la charge mutationnelle permet de prédire la sensibilité aux immunothérapies

Pour chacune des tumeurs, nous avons déterminé la charge mutationnelle, qui s'est avérée efficace pour prédire les réponses du traitement à l'immunothérapie du cancer et notamment par les anticorps anti-PD1 et PD-L1 [153].

Nous avons pris en compte l'ensemble des variants somatiques détectés passant les filtres de qualité, ramenés au nombre de mégabase (MB) que comporte un exome, pour obtenir une valeur en nombre de variants somatiques par MB. Campbell et al. ont établi un seuil de 10 mut/MB pour définir les tumeurs hypermutées. Cette valeur coïncide avec la charge mutationnelle médiane des individus ayant répondu aux immunothérapies par inhibition du checkpoint [306].

Sur les 35 tumeurs de notre cohorte, nous avons déterminé 15 tumeurs hypermutées (41%), dont un tiers des cancers du sein, 2/4 des cancers du poumon, 2/2 des cancers du côlon, 2/3 des cancers de l'ovaire, un cancer gastrique et un lymphome (figure 33). Ces patients sont donc éligibles à un traitement par immunothérapie.



**Figure 33:** Charge mutationnelle des 35 cas en nombre de mutations par mégabase (mut/MB). Une tumeur est considérée comme hypermutée à partir de 10 mut/MB.



Des variants somatiques ou constitutionnels délétères dans des gènes de réparation de l'ADN peuvent générer de l'instabilité génétique. Dans notre cohorte, les variants constitutionnels dans les gènes BRCA1 et BRCA2 identifiés respectivement dans des cancers de l'ovaire et de l'estomac sont associés à des tumeurs hypermutées (20,5 et 10,4 mut/MB).

Le variant c.713A>C ; p.Tyr238Ser de MLH3 associé au syndrome de Lynch et à une instabilité génomique est associé à une tumeur du sein hypermutée à 28,3 Mut/MB. En revanche, les variants de *MSH6* et *EPCAM* identifiés ne sont pas associés à des tumeurs hypermutées.

Enfin, les cinq tumeurs présentant un variant somatique délétère d'*APC* sont également hypermutées, ce qui pourrait être dû à la déstabilisation du fuseau mitotique et à l'instabilité génomique induite par la perte d'*APC*.

### 3.4. Identification des signatures mutationnelles et du profil BRCAness

Tous les cancers sont causés par des variants somatiques. Nous nous sommes intéressés à l'identification des processus biologiques générant ces variants dans les tumeurs de notre cohorte. En 2013, Alexandrov et al. ont établi que l'analyse de l'ensemble des variants somatiques de tumeurs permet d'identifier des signatures moléculaires des processus de mutagenèse qui ont été à l'œuvre durant le développement de la tumeur [307]. Les mutations somatiques trouvées dans les génomes de cancer peuvent par exemple être la conséquence d'erreurs intrinsèques de la machinerie de réplication de l'ADN, d'expositions mutagènes, de modifications enzymatiques de l'ADN ou d'anomalies des mécanismes de réparation.

Différents processus mutationnels génèrent souvent des combinaisons de mutations-types appelées «signatures». Les progrès récents dans la technologie de séquençage ont permis de déchiffrer ces signatures à l'échelle du génome entier. Le WGS, prenant en compte l'ensemble des variants somatiques, est la technique optimale pour la recherche de signatures. D'autres études ont démontré que la méthode peut également être appliquée, mais avec moins d'efficacité, à partir des données d'exome. L'intérêt d'établir les signatures moléculaires en clinique est que la présence de certaines signatures peut être liée à la sensibilité ou à la résistance à certains traitements et donc guider le choix thérapeutique (Figure 34).

Nous avons appliqué à nos 35 cas l'algorithme NMF accessible en version web sur l'application MuSiCa pour extraire des signatures mutationnelles à partir des données de l'exome tumoral. Nous avons pu déterminer la présence d'une signature connue de manière significative dans douze tumeurs. En particulier, nous avons déterminé la présence significative de la signature 3 dans un CS. Elle indique un profil de tumeur BRCAness marqué par un défaut de RH et donc sensible à l'olaparib. La signature 6 a été retrouvée dans deux cancers du sein et signe un défaut de *Mismatch Repair* ciblable par l'immunothérapie, tout comme la signature 20 identifiée dans un cancer du sein et un astrocytome. La signature 8 a été identifiée dans un cancer du poumon et peut indiquer un défaut de réparation par excision de nucléotide (NER) suggérant une sensibilité au cisplatine. Nous avons également observé la signature 13 dans un cancer du sein, qui prédit une résistance au tamoxifène. Enfin, les signatures 12, 17 et 18 ont été identifiées dans d'autres tumeurs de notre cohorte mais n'ont pas d'impact sur le choix thérapeutique.

| A Mutational signatures useful in analysis |            |           |       |            |           | B Underlying mutational process                | C Relevant genes                    | D Predisposition syndrome                     | E Proposed therapy choice   |
|--|------------|-----------|-------|------------|-----------|--|-------------------------------------|---|---|
| CS-3                                       | CS-8       | MH-indels | RS-3  | RS-5       | HRD index | Homologous Recombination Repair Deficiency     | <i>BRCA1, BRCA2, RAD51C, PALB2</i>  | Hereditary Breast and Ovarian Cancer Syndrome | PARP inhibition <sup>32-34</sup> , Platinum-based chemotherapy <sup>35-37</sup> |
| CS-6                                       | CS-15      | CS-20     | CS-26 | STR-indels |           | Mismatch Repair Deficiency                     | <i>MLH1, MSH2, MSH6, PMS1, PMS2</i> | Lynch, CMMRD, BMMR-D, HNPCC                   | PD1-immunotherapy <sup>48-49,52</sup>   |
| CS-5                                       | CS-8       | TSB-sign  |       |            |           | Nucleotide Excision Repair Deficiency          | <i>ERCC1, ERCC2, XPC</i>            | Xeroderma Pigmentosum                         | Cisplatin <sup>63-65</sup>  |
| CS-18                                      | CS-30      | TSB-sign  | C>A*  | G>T*       | C>T*      | Base excision Repair Deficiency                | <i>MUTYH, OGG1, NTHL1, SMUG1</i>    | MAP<br>NAP                                    |   |
| CS-10                                      | STR-indels |           |       |            |           | Deficient DNA polymerase proofreading activity | <i>POLE, POLD1</i>                  | PPAP  | PD1-immunotherapy <sup>48-49,52</sup>   |
| ?  |            |           |       |            |           | Non-Homologous End Joining Deficiency          |                                     | Nijmegen Breakage Syndrome                    |   |
| CS-2                                       | CS-13      | Kataegis  |       |            |           | APOBEC Over-activity                           | <i>APOBEC1, APOBEC3A, APOBEC3B</i>  |   | Tamoxifen Resistance <sup>70,71</sup>   |

**Figure 34:** Les signatures mutationnelles peuvent être associées à une sensibilité ou à une résistance à certains traitements anticancéreux. Les signatures mutationnelles dans les génomes tumoraux peuvent refléter l'activité de processus de mutation spécifiques et ainsi contribuer à la sélection d'un traitement. L'interprétation diagnostique des signatures caractéristiques peut contribuer au choix du traitement et indiquer une sensibilité (en vert) ou une résistance (en rouge) à un traitement. D'après Van Hoeck et al., 2019 [308].

### 3.5 Impact sur les modifications de traitement et devenir des patients

Nous avons émis des possibilités de stratégies thérapeutiques lorsqu'un profil génétique ciblable était identifié. Au total, environ la moitié des individus comportait des altérations pouvant indiquer une modification thérapeutique (immunothérapie, profil BRCAness, variants *PIK3CA*). Pour certains d'entre eux, nous avons pu avoir accès à des informations sur les modifications de traitements apportées par le WES. Au moins 16 individus ont vu leur prise en charge modifiée suite à notre analyse et sont rapportées dans le tableau 13. Par exemple, l'individu 17 atteint d'un cancer du côlon répond à son nouveau traitement par immunothérapie que nous avons suggéré en raison de sa charge mutationnelle relativement élevée (23 mut/Mb). D'autre part, la patiente 19 atteinte d'un cancer de l'ovaire a partiellement répondu à l'olaparib pendant 1 an, médicament que nous avons recommandé en raison d'une altération somatique de *BRCA2*.

| Patient | Âge | Tumeur primaire | Variants constitutionnels | Variants somatiques               | Charge mutationnelle de la tumeur (mut/MB) | Signature moléculaire de la tumeur | Profil génétique ciblable (Oui/Non) | Modification de traitement  |
|---------|-----|-----------------|---------------------------|-----------------------------------|--|------------------------------------|-------------------------------------|-----------------------------|
| 1       | 40  | Sein            |                           | TP53<br>ATR<br>BAP1<br>STK11/LKB1 | 7,3  | 3 (déficience voie BRCA)           | O                                   | Olaparib                    |
| 2       | 75  | Sein            | FGFR4<br>CHEK2            | PTPRJ<br>ESR1                     | -  | ND                                 | N                                   | -                           |
| 3       | 40  | Sein            | FGFR4<br>MLH3             | TP53<br>FBXW7                     | 28   | 6                                  | O                                   | Modification chimiothérapie |
| 4       | 60  | Sein            |                           | TP53<br>NF1<br>EZH1               | 11   | 12                                 | O                                   | -                           |
| 5       | 75  | Sein            |                           | TP53<br>NF1<br>PIK3CA             | 422  | 20                                 | O                                   | Immunothérapie              |
| 6       | 51  | Sein            |                           | PIK3CA                            | 3,2  | ND                                 | N                                   | -                           |
| 7       | 65  | Sein            |                           | TP53<br>PIK3R1                    | 8,1  | 17                                 | N                                   | -                           |

|    |    |        |       |  |     |             |   |                              |
|----|----|--------|-------|--|-----|-------------|---|------------------------------|
| 8  | 60 | Sein   |       | TP53<br>AKT1<br>CDKN2A<br>CHD4           | 11  | 13          | N | Hecreptine,<br>Perjeta       |
| 9  | 47 | Sein   |       | TP53<br>NCOR1<br>GNAS                    | 9,4 | ND          | O | ?                            |
| 10 | 48 | Sein   | MSH6  | TP53<br>PIK3CA<br>POLE                   | 4,5 | 6           | O | Modification<br>surveillance |
| 11 | 43 | Sein   |       | PIK3CA<br>CREBBP<br>C-Met<br>FAT1        | 3,1 | ND          | O | -                            |
| 12 | 70 | Sein   | CHEK2 | BCOR                                     | 5,5 | ND          | N | -                            |
| 13 | 90 | Poumon |       | KEAP1<br>ATM<br>EP300<br>POLE            | 102 | ND          | O | Immunothé-<br>rapie          |
| 14 | 68 | Poumon |       | TP53<br>EP300<br>PIK3CA<br>ETV4          | 34  | 8           | O | Immunothé-<br>rapie          |
| 15 | 55 | Poumon |       | KRAS<br>STK11<br>NF1<br>SMARCA4<br>SETD2 |     | ND          | N | -                            |
| 16 | 50 | Poumon |       | TP53<br>CDKN2A                           |     | ND          | O | -                            |
| 17 | 57 | Côlon  |       | APC<br>FAT4<br>BCL9                      | 23  | 1, 12, 6, 3 | O | Immunothé-<br>rapie          |
| 18 | 67 | Côlon  |       | TP53<br>APC<br>ATM<br>SMAD4              | 59  | ND          | O | -                            |
| 19 | 59 | Ovaire |       | TP53<br>BRCA2<br>FANCA                   | 5,5 | ND          | O | Olaparib                     |
| 20 | 54 | Ovaire | BRCA1 | TP53<br>CDH1<br>MAML2                    | 21  | ND          | O | Olaparib                     |

|    |    |   |              |   |      |         |   |                                |
|----|----|---|--------------|---|------|---------|---|--------------------------------|
|    |    |   |              | FOXO4<br>FGFR2                          |      |         |   |                                |
| 21 | 73 | Ovaire  |              | TP53<br>BRCA1<br>ALK                    | 54   | ND      | O | Olaparib                       |
| 22 | 54 | Pancréas                                      |              | TP53<br>KRAS<br>SMAD4                   |      | ND      | N | Avastin                        |
| 23 | 62 | Pancréas                                      |              | TP53<br>KRAS<br>SMAD4<br>CHD4<br>CDKN2A | 6,2  | 1       | N | Avastin                        |
| 24 | 61 | Pancréas                                      |              | KRAS<br>KMT2C                           | 9,6  | ND      | N | -                              |
| 25 | 63 | Estomac                                       | BRCA2<br>MET | APC<br>ARID1A<br>PIK4CA                 | 10,4 | 12 et 1 | O | Modification<br>chimiothérapie |
| 26 | 46 | Estomac                                       |              | TP53<br>BCOR                            |      | ND      | N | -                              |
| 27 | 52 | Cerveau                                       |              | PTEN<br>NF1                             |      | 18      | N | -                              |
| 28 | 40 | Cerveau                                       |              | TP53<br>IDH1<br>ATRX<br>KMT2C           |      | 20      | N | Avastin                        |
| 29 | 47 | Maxillaire<br>(indifférencié)                 |              | ALK<br>APC<br>FBXW7<br>NOTCH2           | 11   | ND      | N | Inhibiteur Alk                 |
| 30 | 77 | Lymphome<br>diffus à<br>grandes<br>cellules B |              | CREBBP<br>EZH2<br>MEF2B<br>NF1          | 347  | ND      | O | -                              |
| 31 | 48 | Cholangiocar<br>cinome                        |              | KRAS<br>IDH1<br>FGFR1<br>MLH1           |      | ND      | N | -                              |
| 32 | -  | Glandes<br>salivaires<br>(parotide)           |              | PIK3CA<br>APC                           | 38   | 12      | O | -                              |
| 33 | -  | Amygdales                                     |              | KRAS<br>TP63                            | 22   | 12      | N |                                |

|    |    |                                      |                |                                |     |    |   |   |
|----|----|--------------------------------------|----------------|--------------------------------|-----|----|---|---|
|    |    |                                      |                | KMT2C<br>MAPK1<br>BIRC3        |     |    |   | - |
| 34 | 75 | Lymphome<br>NK/T nasal               | FGFR4<br>EPCAM | STAT3<br>FAT4<br>IRF1<br>SETD2 | 8,5 | 1  | N | - |
| 35 | 80 | Zone<br>périanale<br>(indifférencié) |                | TP53<br>SMAD4<br>PTEN          | 9,5 | ND | N | - |

**Tableau 13:** Résultats du profilage génétique des tumeurs des 35 individus de notre cohorte. Le tableau résume les patterns mutationnels constitutionnels et somatiques de chacun, la charge mutationnelle de la tumeur, et les éventuelles signatures moléculaires identifiées. Il est également indiqué les traitements suggérés, et les éventuelles modifications du traitement ayant eu lieu.

### 3.6. Mise en place du pipeline d'analyse en entreprise et construction d'un rapport

Toutes les étapes du test ont été développées par l'entreprise Prenostics. Le service de séquençage a été réalisé dans un laboratoire accrédité et certifié CAP (College of American Pathologist) et CLIA (Clinical Laboratory Improvement Amendments). Le traitement bioinformatique des données brutes incluant le contrôle qualité des *reads* est effectué en interne.

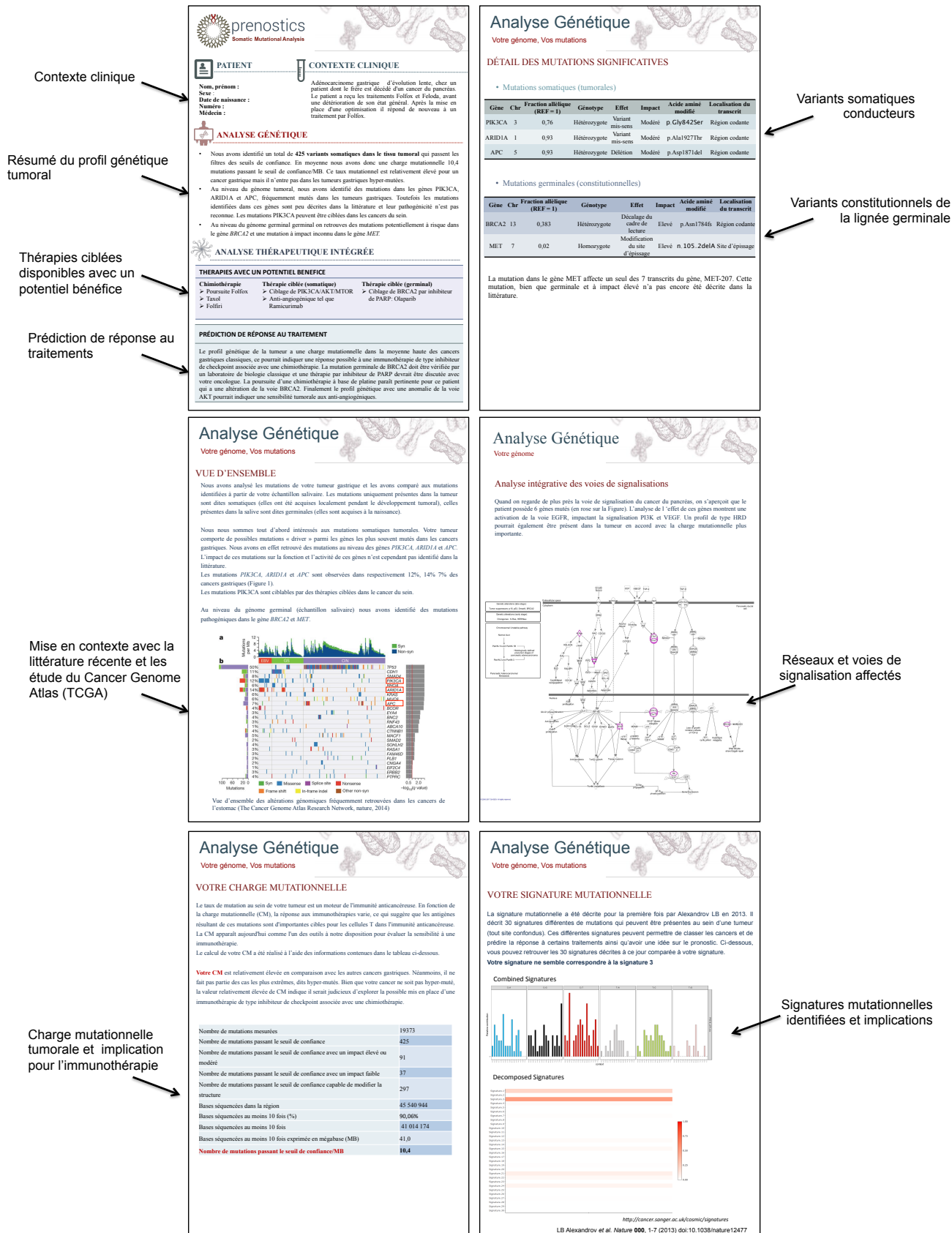
Les pipelines d'analyse des données pour l'établissement du profil génétique des tumeurs ont été développés en entreprise, permettant d'établir :

- La liste des variants somatiques tumoraux, comprenant les variants *driver* de la tumeur,
- les variants constitutionnels de prédisposition au cancer dans la lignée germinale,
- l'identification des voies de signalisation affectées dans la tumeur,
- la charge mutationnelle tumorale,
- la signature mutationnelle de la tumeur comprenant un score de stabilité des microsatellites.

Toutes ces données sont interprétées manuellement, en regard des données de la littérature, permettant la mise en évidence de marqueurs prédictifs d'un bénéfice aux thérapies ciblées existantes ou à l'inclusion dans un essai clinique en cours.

La construction du rapport final décrivant le profil génétique de la tumeur s'est donc articulée autour de ces cinq points. La première page contient un résumé des principales caractéristiques génétiques identifiées et des éventuelles altérations actionnables et les pages suivantes présentent en détail les résultats, une mise en contexte de la littérature, des thérapies ciblées disponibles et des essais cliniques en cours (Figure 35). Un exemple de rapport complet est présenté en annexe 2.

La durée de rendu des résultats était en moyenne de 10 semaines sur l'ensemble de la cohorte. L'optimisation progressive des étapes nous a permis d'atteindre un délai de quatre semaines pour les dernières analyses.





## 4. Discussion

Le traitement du cancer est entré dans une nouvelle ère dans laquelle des altérations génétiques propres à chaque tumeur sont identifiables, offrant ainsi la possibilité de choisir un traitement ciblé adapté au profil génétique de chaque patient. Les technologies de séquençage ont progressé, offrant la possibilité du séquençage tumoral à l'échelle de l'exome, laissant entrevoir des applications prometteuses en oncologie de précision. Mais les performances de cet outil restent encore à être évaluées en milieu clinique.

Cette étude pilote avait pour objectif d'évaluer la faisabilité de réaliser le WES en temps réel chez des individus atteints d'un cancer avancé et d'explorer son potentiel d'utilité clinique supplémentaire par rapport au séquençage de panels.

La faisabilité a été évaluée sur les critères de réussite de biopsie des échantillons de tumeur (pureté tumorale suffisante pour le séquençage pour plus de 80% des individus biopsiés), l'analyse de séquençage réussie (exome effectué avec succès pour plus de 90% des individus séquencés), les délais de traitement (résultats en moins de 12 semaines médianes) et l'utilité clinique (résultats génomiques donnant lieu à une modification du traitement identifiés chez plus de 30% des individus séquencés). Sur la base de critères prédéfinis, notre étude a atteint tous les critères de faisabilité et a démontré que le WES est réalisable en clinique et qu'il peut apporter des avantages supplémentaires aux patients.

Comme attendu, nous n'avons pas identifié de variants directement ciblables (par exemple *EGFR* T790M ou *BRAF* V600E) parmi les tumeurs de notre cohorte. Cela s'explique par le fait que les individus recrutés dans notre étude ont des cancers réfractaires, ayant été préalablement testés pour les variants directement actionnables (faux sens, tronquant, fusion de gènes) par la réalisation de séquençage ciblé de gènes actionnables par des thérapies ciblées.

L'analyse de l'exome a montré que près de la moitié des cas présentaient des variants somatiques et/ou constitutionnels indiquant une potentielle réponse ou une résistance à un type de traitement. Toutefois, avec un panel de gènes adapté, nous aurions obtenu les mêmes résultats. Le séquençage d'exome n'est donc pas beaucoup plus avantageux que le séquençage de gènes ciblés sur le plan de l'identification de SNV d'intérêt clinique. Notons toutefois que le WES a permis

l'identification d'altérations de certains gènes moins connus des oncologues car non directement actionnables par les thérapies ciblées actuellement disponibles, mais potentiellement ciblables par de futures molécules en cours d'essais cliniques. Ainsi, avec l'exome entier, ce test ne nécessitera pas de re-séquençage si de nouveaux gènes doivent être analysés par la suite. Les données de l'exome pourront faire l'objet d'une relecture annuelle en considérant les mises à jour de traitements et d'essais cliniques disponibles.

En revanche, il est des critères où le WES a été nettement plus contributif que le panel de gènes. Grâce à sa couverture nettement plus large, nous avons été capables de déterminer avec précision d'autres paramètres, indéfinissables par séquençage de panels, qui se sont avérés pertinents pour influencer la stratégie thérapeutique.

Parmi ces critères caractérisés, la charge mutationnelle de la tumeur a contribué de façon importante au profilage des tumeurs. Elle se définit par le nombre total de variants somatiques présents dans l'ensemble des régions exoniques du génome et a donc pu être très précisément calculée par notre approche. La détermination de cette charge mutationnelle est un indicateur de la réponse des cellules tumorales aux immunothérapies, en particulier aux inhibiteurs du checkpoint immunitaire via les anticorps anti-PDL1 [309]. Les tumeurs présentant une charge mutationnelle supérieure à 10 mut/MB sont considérées comme hypermutées et sensibles aux immunothérapies [306]. Au total, dans notre cohorte, 42% des individus avaient une charge mutationnelle élevée, supérieure à 10, et donc susceptibles d'être sensibles à ce traitement.

L'instabilité des microsatellites élevée est également un marqueur de réponse aux immunothérapies. Cependant, des études récentes tendent à montrer que la charge mutationnelle est un indicateur plus fiable de réponse aux immunothérapies, et de plus, les tumeurs présentant un profil hypermuté sont 5 fois plus nombreuses que celles présentant une instabilité microsatellite élevée [310]. De ce fait, le calcul de la charge mutationnelle semble être un meilleur marqueur prédictif et inclut plus de patients pour l'immunothérapie que l'instabilité microsatellite ; c'est pourquoi nous n'avons pas systématiquement déterminé cette dernière dans notre étude.

Un autre avantage intéressant de l'exome sur le panel de gène est la possibilité de caractériser la signature moléculaire de chaque tumeur, décrite par Alexandrov *et al.* en 2013 [307]. Elle ne peut être calculée à partir du séquençage d'un panel de gène car il est essentiel d'avoir une vision la plus complète possible du génome pour la calculer précisément. La signature moléculaire est révélatrice des processus biologiques de mutagenèse en œuvre dans la tumeur, à l'origine de l'accumulation de variants somatiques. Comme les mécanismes moléculaires sous-jacents aux différentes signatures spécifiques peuvent être couplés à des stratégies de traitement, nous les avons calculées pour chaque tumeur. L'analyse de signature mutationnelle nécessite une analyse NGS à grande échelle de type WGS ou WES réalisée en comparaison à un échantillon sain apparié..

Dans notre cohorte, nous avons pu identifier une signature spécifique se détachant clairement dans douze tumeurs et trois étaient des mélanges de plusieurs signatures. Parmi les signatures identifiées, nous avons identifié une grande majorité de signatures signant un défaut de *Mismatch Repair* (signatures 6, 8, 13, 20) associées à une meilleure réponse aux immunothérapies. Le profil le plus intéressant du point de vue thérapeutique est la signature 3 qui correspond au profil BRCAness et qui marque un dysfonctionnement de la RH. Sa présence dans la tumeur mammaire de l'individu 1, par ailleurs sans altération de *BRCA1/BRCA2*, est en effet évocatrice d'une sensibilité à l'olaparib.

Au niveau de l'ADN constitutionnel, la présence d'un facteur de prédisposition a été identifiée dans 4 cas sur 35. Nous avons identifié chez ces individus des variants dans les gènes de prédispositions aux cancers du sein et de l'ovaire *BRCA1* et *BRCA2*, ainsi que dans les gènes du syndrome de Lynch *MLH3*, *MSH6*. Ces gènes sont tous impliqués dans la réparation de l'ADN, et constituent, dans les cas de *BRCA1* et *BRCA2*, une marque de sensibilité aux inhibiteurs de PARP tel que l'olaparib [311]. Au-delà d'aider à la prise de décision thérapeutique, la présence de ces variants à risque chez ces individus suggère une probable prédisposition familiale au cancer et implique de tester les membres de sa famille et de mettre en place un conseil génétique dans une optique préventive.

Nous avons rencontré plusieurs difficultés techniques lors de la mise en œuvre de ce test. Premièrement, les échantillons tumoraux utilisés étaient des tissus FFPE, ce qui

peut endommager l'ADN, ayant nécessité des méthodologies optimisées pour permettre une extraction suffisante et un dosage précis de l'ADN. Un des problèmes récurrent a été la quantité et la qualité d'ADN extraite à partir des tissus FFPE qui était souvent à la limite requise pour le séquençage d'exome. Huit individus soit 18% des individus de la cohorte de départ n'ont pas pu être séquençés en raison de la quantité insuffisante d'ADN tumoral extrait et de la non disponibilité de matériel supplémentaire. La qualité de l'ADN a également été un facteur limitant, puisque deux séquençages ont échoué en raison d'une contamination chimique de l'échantillon, probablement due à la présence résiduelle de paraffine. En plus du faible rendement d'extraction des tissus FFPE, les biopsies de tumeurs avancées, associées au plus grand risque de morbidité, deviennent de plus en plus petites. Nous avons pour chaque individu sélectionné des biopsies tumorales présentant le pourcentage le plus élevé de noyaux tumoraux. Cependant pour un certain nombre d'échantillons, le pourcentage de tissu tumoral était faible (taux élevé de contamination cellulaire normale), ce qui a nécessité une sensibilité élevée du test. Une spécificité élevée a été également requise car les rapports de faux positifs peuvent conduire à un choix thérapeutique non optimal. C'est pour cette raison qu'il a été nécessaire d'augmenter la profondeur de séquençage à 130X, pour détecter avec précision le mosaïcisme des variants somatiques tumoraux.

Une autre difficulté importante à laquelle nous avons été confrontés a été de raccourcir au maximum les délais. Il est généralement admis dans la communauté des oncologues que les résultats du profilage moléculaire devraient être produits dans un délai de trois semaines. Ce laps de temps permet de prendre une décision de traitement avant la détérioration du patient et est la cible de la plupart des services de profilage privés et publics. Dans notre étude, la réalisation de l'exome a nécessité une durée moyenne de 10 semaines. Les expériences tirées de ce projet pilote nous ont permis de générer des protocoles et des procédures standards pour atteindre l'objectif de trois semaines. Le délai s'est progressivement réduit au fur et à mesure des analyses, passant de plus de douze semaines pour les premières à quatre semaines pour les dernières.

Démontrer la faisabilité du WES de l'ADN tumoral d'un patient en temps réel est une première étape indispensable pour la démocratisation de la médecine personnalisée

en cancérologie. L'objectif ultime est de pouvoir associer les profils génétiques de chaque patient à des réponses aux traitements.

Finalement, cette étude démontre que le WES du génome tumoral est techniquement réalisable et qu'il complète la prise de décision clinique. Toutefois, il est nécessaire de continuer à développer davantage les processus de travail de laboratoire afin de réduire le délai d'obtention des résultats, du moins dans le cadre de notre programme. Au fil du temps, la collecte continue de données WES améliorera notre compréhension de la génétique du cancer et fera pencher la balance en faveur de l'utilisation du séquençage génomique à grande échelle en oncologie clinique.

## 5. Conclusion

Dans des cas où l'approche classique (séquençage de panel et identification des CNV) était en échec pour proposer des thérapies personnalisées, l'analyse des données de l'exome a éclairé la gestion du cancer dans 51% des cas notamment grâce au calcul de la charge mutationnelle et l'identification des profils BRCAness. À notre connaissance, au moins huit individus ont pu bénéficier d'un changement thérapeutique. Trois d'entre eux, ayant une déficience de la voie RH, ont pu débiter un traitement par inhibiteur de PARP (olaparib). Quinze individus ayant une charge mutationnelle élevée pour leur type tumoral ont été orientés vers l'immunothérapie par inhibition de checkpoint. De plus, quatre individus pour lesquels nous avons identifié des variants pathogéniques constitutionnels dans les gènes de prédisposition au cancer et au syndrome de Lynch (*BRCA1*, *BRCA2*, *MSH6*, *MLH3*) ont pu bénéficier de conseils sur la gestion du risque familial de cancer. Les principales limitations de notre test ont été l'absence de données sur les CNV de la tumeur et la présence de réarrangements oncogéniques, qui peuvent dans certains cas être ciblables, à l'image des réarrangements de *ALK* ciblables par crizotinib. Des tests complémentaires comme le FISH ou le CGH array doivent encore être effectués afin d'avoir une vue complète des altérations moléculaires de la tumeur.

## Conclusions générales et perspectives

Nous avons démontré l'utilité clinique du WES dans nos différentes cohortes pour le diagnostic des individus atteints de maladies monogéniques et de cancers avancés.

L'impact de nos analyses WES sur la prise en charge des patients est important :

- concernant les cas pédiatriques, notre approche par trio a permis de diagnostiquer 35% d'individus non diagnostiqués par séquençage de panel de gènes, mettant fin à l'errance diagnostique de ces familles et améliorant potentiellement la prise en charge et le traitement ;
- L'analyse de l'exome de tumeurs avancées réfractaires de traitement a amélioré significativement la prise de décision thérapeutique par rapport au séquençage de panel, notamment grâce au calcul précis de la charge mutationnelle tumorale et des signatures moléculaires.
- En revanche l'intérêt de l'exome par rapport au séquençage ciblé la recherche de de prédisposition au CS héréditaire dans les familles BRCAx est moins évident. Malgré l'identification des variant candidats dans des gènes non associés au CS, l'impact et la pathogénicité de ces derniers sont difficiles à valider fonctionnellement. L'identification des facteurs de prédisposition dans chaque famille permettrait une meilleure évaluation du risque de leurs membres non atteints.

Ces résultats démontrent l'intérêt et de la faisabilité du séquençage d'exome en clinique. Son rendement, bien supérieur au séquençage ciblé de panels, est en faveur de son application en routine clinique, et notamment en première instance de tests.

Cependant, ces projets ont mis en évidence des limitations dans l'utilisation du WES. Bien qu'il soit performant pour l'identification de SNV et indels, sa couverture restreinte du génome et ses lectures courtes ne sont pas optimales pour la détection des variants structuraux tels que les réarrangements chromosomiques et les CNVs. Des efforts ont été réalisés pour détecter ce type d'altérations de manière fiable à partir des données de WES, comme le développement de techniques bioinformatiques pour la détection de réarrangements (Breakdancer ou ClusterFAST) et le développement de captures augmentées (CoDE-seq) pour la détection de

CNVs. Dans nos cohortes d'individus atteints de maladies génétiques ou de cancer, ces anomalies ont été recherchées en première intention par CGH array et le WES a été réalisé seulement dans un deuxième temps lorsque le CGH array était négatif. Nous nous sommes donc limités intentionnellement aux SNVs et aux petits indels. La mise au point de tests WES plus complets, incluant tous les types d'altération en première intention, permettrait d'optimiser son rendement.

La seconde limitation importante du WES est l'absence de couverture des régions non codantes. Même si la majorité des altérations impliquées dans les maladies sont situées dans les parties codantes, il est estimé qu'environ 20% résident dans les régions non codantes. Ces variants non détectables dans les régions non codantes par WES représentent probablement une part des cas de maladies génétiques que nous n'avons pas pu diagnostiquer. Le coût du WGS, environ trois à quatre fois supérieur que celui du WES, reste pour le moment prohibitif pour une application de routine dans la plupart des applications cliniques. Cependant, les coûts diminuent régulièrement et un rapport coût/efficacité satisfaisant du WGS pourrait être atteint dans les années à venir et pourrait alors remplacer le WES.

De plus, de nombreuses questions éthiques doivent être examinées et traitées [312, 313]. Par exemple, le séquençage à grande échelle peut mener à l'identification d'anomalies par hasard, appelées découvertes fortuites. Elles peuvent avoir un effet important sur les prédispositions des personnes à certaines pathologies non suspectées à la base et pourraient engendrer des prises de décision médicales [312, 314]. La Société Française de Médecine Prédictive et Personnalisée (SFMPP) a récemment publié des directives sur la manière de traiter ces découvertes fortuites, en précisant lesquelles présentent un bénéfice à être communiquées aux patients [315].

Le NGS doit continuer à fournir de nouvelles connaissances sur les mécanismes moléculaires des maladies mais il doit également se démocratiser dans le système de soin actuel en tant qu'outil d'investigation à la fois diagnostic, prédictif, et thérapeutique.

L'accès au WES demeure très inégalitaire en France étant donné que ce type de séquençage n'est pas remboursé par la sécurité sociale. Par exemple, dans le contexte du diagnostic de maladies rares, il revient aux hôpitaux de prendre en

charge la moitié du coût du WES, qui est de 2200€ en moyenne pour chaque patient, et certains établissements y sont plus enclins que d'autres.

C'est dans ce contexte que la France a mis en place le plan France médecine génomique en 2016. Celui-ci vise à faire intégrer d'ici 2025 la médecine génomique dans le parcours de soins courant et la prise en charge des pathologies et à s'assurer que chacun puisse y accéder de manière équitable sur tout le territoire. Pour ce faire, le développement d'une filière nationale de médecine génomique au service des patients est en cours, dont l'objectif est de favoriser l'innovation scientifique et technologique, la valorisation industrielle et la croissance économique. Le plan France génomique prévoit de placer la France dans le peloton de tête des grands pays engagés dans la médecine personnalisée, avec une capacité d'exportation du savoir-faire de notre filière médicale et industrielle en médecine génomique. L'enjeu économique est en effet majeur. D'une part en termes de coût pour notre système de soin, car la médecine de précision devrait supprimer des coûts du système de santé actuel (notamment par la diminution du nombre de bilans inadaptés, imprécis et onéreux, la suppression ou limitation de médicaments inutiles, et l'élimination de certains effets secondaires handicapants). D'autre part, la médecine génomique représente une opportunité de développement d'une nouvelle filière industrielle. Cependant, la France concède du retard sur ce point par rapport aux autres pays développés, dans ce domaine où la concurrence est devenue particulièrement accrue. Bien que les prévisions de croissance du marché des tests génétiques et de la génomique médicale soient bonnes (10% de croissance par an d'ici 2025) dû notamment à l'essor de la médecine de précision, il existe très peu de sociétés françaises dans ce secteur. La France a raté cette vague, à la fois parce que le contexte médical n'y était pas propice mais aussi à cause du manque de financements nationaux disponibles. Quelques entreprises s'aventurent dans le séquençage, la majorité se focalise dans le domaine du génotypage ciblé sur certaines pathologies, majoritairement dans l'oncologie comme IntegraGen. Elles font face à une rude concurrence de plate-formes hospitalières publiques et d'entreprises américaines du même domaine comme FoundationMedicine (Roche), Abbott Molecular, 23AndMe, Genomic Health (OncotypeDX), Myriad, ou européennes comme Sophia Genetics et DNA vision. Très peu de start-up se sont aventurées dans la pratique de tests génomiques à grande échelle de type exome ou génome entier en partie à cause de la législation française particulièrement stricte



concernant l'utilisation de ce type de tests génétiques. En effet les lois successives de bioéthique françaises encadrent très étroitement à la fois leur prescription, leur réalisation et le mode de rendu du résultat au patient. Un test de génomique médicale peut être proposé au patient seulement s'il est réalisé dans un laboratoire de biologie médicale accrédité et que sa valeur médicale est importante et complètement démontrée (prescription médicale nécessaire).

Nous sommes donc encore loin, en France, d'une médecine génomique prédictive, permettant à chacun de séquencer son génome entier à des fins de dépistage (cancer, cardiovasculaires, diabète, obésité, Alzheimer ...). Bien qu'une telle pratique puisse améliorer la prévention en diminuant les comportements à risques chez les personnes susceptibles de développer des maladies, la divulgation et le stockage de cette masse de données génomiques confidentielles pose des problèmes éthiques importantes qui font débat.

Dans un avenir proche, alors que les coûts continuent de baisser et que le débit continue à augmenter, les technologies telles que le séquençage de l'exome deviendront encore plus accessibles en recherche et en clinique, de sorte que des projets tels que celui réalisé ici puissent continuer à identifier de nouveaux variants contribuant aux maladies. L'accumulation croissante de données de séquençage pourra continuer à alimenter les bases de connaissance et favoriser l'annotation des VUS et de leurs impacts sur les maladies humaines.

Plusieurs études pilotes réalisées ces dernières années, dont la nôtre, ont montré la faisabilité et le bénéfice de l'utilisation du séquençage de l'exome complet dans la prise en charge des patients atteints de maladie rares ou de cancers. Il est donc légitime de penser qu'elles deviendront probablement incontournables à court ou moyen terme dans le parcours de soin des patients.

## Références

1. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase*. J Mol Biol, 1975. **94**(3): p. 441-8.
2. Adams, M.D., et al., *Complementary DNA sequencing: expressed sequence tags and human genome project*. Science, 1991. **252**(5013): p. 1651-6.
3. Collins, F.S., M. Morgan, and A. Patrinos, *The Human Genome Project: lessons from large-scale biology*. Science, 2003. **300**(5617): p. 286-90.
4. Schmutz, J., et al., *Quality assessment of the human genome sequence*. Nature, 2004. **429**(6990): p. 365-8.
5. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors*. Nature, 2005. **437**(7057): p. 376-80.
6. Mardis, E.R., *A decade's perspective on DNA sequencing technology*. Nature, 2011. **470**(7333): p. 198-203.
7. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies*. Nat Rev Genet, 2016. **17**(6): p. 333-51.
8. Cock, P.J., et al., *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants*. Nucleic Acids Res, 2010. **38**(6): p. 1767-71.
9. Fonseca, N.A., et al., *Tools for mapping high-throughput sequencing data*. Bioinformatics, 2012. **28**(24): p. 3169-77.
10. Li, H. and R. Durbin, *Fast and accurate long-read alignment with Burrows-Wheeler transform*. Bioinformatics, 2010. **26**(5): p. 589-95.
11. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
12. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
13. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 2016. **536**(7616): p. 285-91.
14. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data*. Nat Genet, 2011. **43**(5): p. 491-8.
15. Regier, A.A., et al., *Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects*. Nat Commun, 2018. **9**(1): p. 4038.
16. Zook, J.M., et al., *Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls*. Nat Biotechnol, 2014. **32**(3): p. 246-51.
17. Chiang, C., et al., *The impact of structural variation on human gene expression*. Nat Genet, 2017. **49**(5): p. 692-699.
18. Chen, X., et al., *Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications*. Bioinformatics, 2016. **32**(8): p. 1220-2.
19. Handsaker, R.E., et al., *Discovery and genotyping of genome structural polymorphism by sequencing on a population scale*. Nat Genet, 2011. **43**(3): p. 269-76.
20. Layer, R.M., et al., *LUMPY: a probabilistic framework for structural variant discovery*. Genome Biol, 2014. **15**(6): p. R84.
21. Rausch, T., et al., *DELLY: structural variant discovery by integrated paired-end and split-read analysis*. Bioinformatics, 2012. **28**(18): p. i333-i339.

22. Montagne, L., et al., *CoDE-seq, an augmented whole-exome sequencing, enables the accurate detection of CNVs and mutations in Mendelian obesity and intellectual disability*. Mol Metab, 2018. **13**: p. 1-9.
23. 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, 2015. **17**(5): p. 405-24.
24. Genomes Project, C., et al., *A global reference for human genetic variation*. Nature, 2015. **526**(7571): p. 68-74.
25. Tennesen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes*. Science, 2012. **337**(6090): p. 64-9.
26. Kobayashi, Y., et al., *Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation*. Genome Med, 2017. **9**(1): p. 13.
27. Whiffin, N., et al., *Using high-resolution variant frequencies to empower clinical genome interpretation*. Genet Med, 2017. **19**(10): p. 1151-1158.
28. Jang, M.A., et al., *Frequency of the moyamoya-related RNF213 p.Arg4810Lys variant in 1,516 Korean individuals*. BMC Med Genet, 2015. **16**: p. 109.
29. Tabor, H.K., et al., *Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results*. Am J Hum Genet, 2014. **95**(2): p. 183-93.
30. Jagadeesh, K.A., et al., *M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity*. Nat Genet, 2016. **48**(12): p. 1581-1586.
31. Ionita-Laza, I., et al., *A spectral approach integrating functional genomic annotations for coding and noncoding variants*. Nat Genet, 2016. **48**(2): p. 214-20.
32. Ioannidis, N.M., et al., *REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants*. Am J Hum Genet, 2016. **99**(4): p. 877-885.
33. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nat Methods, 2010. **7**(4): p. 248-9.
34. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function*. Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
35. Choi, Y., et al., *Predicting the functional effect of amino acid substitutions and indels*. PLoS One, 2012. **7**(10): p. e46688.
36. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants*. Nat Genet, 2014. **46**(3): p. 310-5.
37. Amendola, L.M., et al., *Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium*. Am J Hum Genet, 2016. **99**(1): p. 247.
38. Nykamp, K., et al., *Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria*. Genet Med, 2017. **19**(10): p. 1105-1117.
39. Jarvik, G.P. and B.L. Browning, *Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants*. Am J Hum Genet, 2016. **98**(6): p. 1077-1081.
40. Jian, X., E. Boerwinkle, and X. Liu, *In silico prediction of splice-altering single nucleotide variants in the human genome*. Nucleic Acids Res, 2014. **42**(22): p. 13534-44.
41. Castellsague, E., et al., *Allele-specific expression of APC in adenomatous polyposis families*. Gastroenterology, 2010. **139**(2): p. 439-47, 447 e1.

42. Gaildrat, P., et al., *The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements*. J Med Genet, 2010. **47**(6): p. 398-403.
43. Borrás, E., et al., *Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants*. J Med Genet, 2013. **50**(8): p. 552-63.
44. Kizhakeyil, A., et al., *Isolation of Human Peripheral Blood T-Lymphocytes*. Methods Mol Biol, 2019. **1930**: p. 11-17.
45. Rath, A., et al., *Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells*. Hum Mutat, 2019.
46. Worthey, E.A., et al., *Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease*. Genet Med, 2011. **13**(3): p. 255-62.
47. Amberger, J.S., et al., *OMIM.org: leveraging knowledge across phenotype-gene relationships*. Nucleic Acids Res, 2019. **47**(D1): p. D1038-D1043.
48. Hirschhorn, J.N., *Genetic approaches to studying common diseases and complex traits*. Pediatr Res, 2005. **57**(5 Pt 2): p. 74R-77R.
49. Johnson, G.C. and J.A. Todd, *Strategies in complex disease mapping*. Curr Opin Genet Dev, 2000. **10**(3): p. 330-4.
50. Ghosh, S. and F.S. Collins, *The geneticist's approach to complex disease*. Annu Rev Med, 1996. **47**: p. 333-53.
51. Wood, A.R., et al., *Defining the role of common variation in the genomic and biological architecture of adult human height*. Nat Genet, 2014. **46**(11): p. 1173-86.
52. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration*. Science, 2005. **308**(5720): p. 385-9.
53. Edwards, A.O., et al., *Complement factor H polymorphism and age-related macular degeneration*. Science, 2005. **308**(5720): p. 421-4.
54. Haines, J.L., et al., *Complement factor H variant increases the risk of age-related macular degeneration*. Science, 2005. **308**(5720): p. 419-21.
55. Evangelou, E. and J.P. Ioannidis, *Meta-analysis methods for genome-wide association studies and beyond*. Nat Rev Genet, 2013. **14**(6): p. 379-89.
56. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations*. Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
57. Gallagher, M.D. and A.S. Chen-Plotkin, *The Post-GWAS Era: From Association to Function*. Am J Hum Genet, 2018. **102**(5): p. 717-730.
58. Cook, E.H., Jr. and S.W. Scherer, *Copy-number variations associated with neuropsychiatric conditions*. Nature, 2008. **455**(7215): p. 919-23.
59. Vissers, L.E., B.B. de Vries, and J.A. Veltman, *Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis*. J Med Genet, 2010. **47**(5): p. 289-97.
60. Koolen, D.A., et al., *Genomic microarrays in mental retardation: a practical workflow for diagnostic applications*. Hum Mutat, 2009. **30**(3): p. 283-92.
61. Sebat, J., et al., *Strong association of de novo copy number mutations with autism*. Science, 2007. **316**(5823): p. 445-9.
62. Xu, B., et al., *Strong association of de novo copy number mutations with sporadic schizophrenia*. Nat Genet, 2008. **40**(7): p. 880-5.

63. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder*. Nat Genet, 2010. **42**(1): p. 30-5.
64. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome*. Nat Genet, 2010. **42**(9): p. 790-3.
65. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome*. Nat Genet, 2010. **42**(6): p. 483-5.
66. Gilissen, C., et al., *Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome*. Am J Hum Genet, 2010. **87**(3): p. 418-23.
67. Need, A.C., et al., *Clinical application of exome sequencing in undiagnosed genetic conditions*. J Med Genet, 2012. **49**(6): p. 353-61.
68. Gahl, W.A., et al., *The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases*. Genet Med, 2012. **14**(1): p. 51-9.
69. Yang, Y., et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders*. N Engl J Med, 2013. **369**(16): p. 1502-11.
70. Yang, Y., et al., *Molecular findings among patients referred for clinical whole-exome sequencing*. JAMA, 2014. **312**(18): p. 1870-9.
71. Farwell, K.D., et al., *Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions*. Genet Med, 2015. **17**(7): p. 578-86.
72. 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, 2013. **34**(12): p. 1721-6.
73. Vissers, L.E., et al., *A de novo paradigm for mental retardation*. Nat Genet, 2010. **42**(12): p. 1109-12.
74. de Ligt, J., et al., *Diagnostic exome sequencing in persons with severe intellectual disability*. N Engl J Med, 2012. **367**(20): p. 1921-9.
75. Harripaul, R., et al., *Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families*. Mol Psychiatry, 2018. **23**(4): p. 973-984.
76. Euro, E.-R.E.S.C.E.a.e.-R.E.S.u.a.b., et al., *De Novo Mutations in Synaptic Transmission Genes Including DNMT1 Cause Epileptic Encephalopathies*. Am J Hum Genet, 2017. **100**(1): p. 179.
77. Epi, K.C., et al., *De novo mutations in epileptic encephalopathies*. Nature, 2013. **501**(7466): p. 217-21.
78. O'Roak, B.J., et al., *Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations*. Nat Genet, 2011. **43**(6): p. 585-9.
79. Jiang, Y.H., et al., *Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing*. Am J Hum Genet, 2013. **93**(2): p. 249-63.
80. Sanders, S.J., et al., *De novo mutations revealed by whole-exome sequencing are strongly associated with autism*. Nature, 2012. **485**(7397): p. 237-41.
81. O'Roak, B.J., et al., *Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations*. Nature, 2012. **485**(7397): p. 246-50.
82. Lescai, F., et al., *Whole-exome sequencing of individuals from an isolated population implicates rare risk variants in bipolar disorder*. Transl Psychiatry, 2017. **7**(2): p. e1034.
83. Chardon, J.W., et al., *LIMS2 mutations are associated with a novel muscular dystrophy, severe cardiomyopathy and triangular tongues*. Clin Genet, 2015. **88**(6): p. 558-64.

84. Schindler, R.F., et al., *POPDC1(S201F) causes muscular dystrophy and arrhythmia by affecting protein trafficking*. J Clin Invest, 2016. **126**(1): p. 239-53.
85. Kayman-Kurekci, G., et al., *Mutation in TOR1AIP1 encoding LAP1B in a form of muscular dystrophy: a novel gene related to nuclear envelopathies*. Neuromuscul Disord, 2014. **24**(7): p. 624-33.
86. Basmanav, F.B., et al., *Mutations in POGlut1, encoding protein O-glucosyltransferase 1, cause autosomal-dominant Dowling-Degos disease*. Am J Hum Genet, 2014. **94**(1): p. 135-43.
87. Straussberg, R., et al., *Kyphoscoliosis peptidase (KY) mutation causes a novel congenital myopathy with core targetoid defects*. Acta Neuropathol, 2016. **132**(3): p. 475-8.
88. O'Grady, G.L., et al., *Variants in the Oxidoreductase PYROXD1 Cause Early-Onset Myopathy with Internalized Nuclei and Myofibrillar Disorganization*. Am J Hum Genet, 2016. **99**(5): p. 1086-1105.
89. Majczenko, K., et al., *Dominant mutation of CCDC78 in a unique congenital myopathy with prominent internal nuclei and atypical cores*. Am J Hum Genet, 2012. **91**(2): p. 365-71.
90. Agrawal, P.B., et al., *SPEG interacts with myotubularin, and its deficiency causes centronuclear myopathy with dilated cardiomyopathy*. Am J Hum Genet, 2014. **95**(2): p. 218-26.
91. O'Connor, E., et al., *Identification of mutations in the MYO9A gene in patients with congenital myasthenic syndrome*. Brain, 2016. **139**(Pt 8): p. 2143-53.
92. Oliveira, J., et al., *The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms ASCC1 involvement*. Clin Genet, 2017. **92**(4): p. 434-439.
93. Chi, B., et al., *The neurodegenerative diseases ALS and SMA are linked at the molecular level via the ASC-1 complex*. Nucleic Acids Res, 2018. **46**(22): p. 11939-11951.
94. Knierim, E., et al., *Mutations in Subunits of the Activating Signal Cointegrator 1 Complex Are Associated with Prenatal Spinal Muscular Atrophy and Congenital Bone Fractures*. Am J Hum Genet, 2016. **98**(3): p. 473-489.
95. Bohm, J., et al., *Novel ASCC1 mutations causing prenatal-onset muscle weakness with arthrogyrosis and congenital bone fractures*. J Med Genet, 2019. **56**(9): p. 617-621.
96. Guelly, C., et al., *Targeted high-throughput sequencing identifies mutations in atlastin-1 as a cause of hereditary sensory neuropathy type I*. Am J Hum Genet, 2011. **88**(1): p. 99-105.
97. Klein, C.J., et al., *Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss*. Nat Genet, 2011. **43**(6): p. 595-600.
98. Kornak, U., et al., *Sensory neuropathy with bone destruction due to a mutation in the membrane-shaping atlastin GTPase 3*. Brain, 2014. **137**(Pt 3): p. 683-92.
99. Helbig, K.L., et al., *Biallelic loss-of-function variants in DOCK3 cause muscle hypotonia, ataxia, and intellectual disability*. Clin Genet, 2017. **92**(4): p. 430-433.
100. Guzman, Y.F., et al., *A gain-of-function mutation in the GRIK2 gene causes neurodevelopmental deficits*. Neurol Genet, 2017. **3**(1): p. e129.
101. Perez, Y., et al., *SLC30A9 mutation affecting intracellular zinc homeostasis causes a novel cerebro-renal syndrome*. Brain, 2017. **140**(4): p. 928-939.
102. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome*. Nature, 2009. **458**(7239): p. 719-24.

103. Garraway, L.A. and E.S. Lander, *Lessons from the cancer genome*. Cell, 2013. **153**(1): p. 17-37.
104. Meyerson, M., S. Gabriel, and G. Getz, *Advances in understanding cancer genomes through second-generation sequencing*. Nat Rev Genet, 2010. **11**(10): p. 685-96.
105. Nakagawa, H., et al., *Cancer whole-genome sequencing: present and future*. Oncogene, 2015. **34**(49): p. 5943-50.
106. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours*. Nature, 2012. **490**(7418): p. 61-70.
107. International Cancer Genome, C., et al., *International network of cancer genome projects*. Nature, 2010. **464**(7291): p. 993-8.
108. Nakagawa, H. and M. Fujita, *Whole genome sequencing analysis for cancer genomics and precision medicine*. Cancer Sci, 2018. **109**(3): p. 513-522.
109. King, C.R., M.H. Kraus, and S.A. Aaronson, *Amplification of a novel v-erbB-related gene in a human mammary carcinoma*. Science, 1985. **229**(4717): p. 974-6.
110. Ilyas, M., *Next-Generation Sequencing in Diagnostic Pathology*. Pathobiology, 2017. **84**(6): p. 292-305.
111. Jegu, M., et al., [*Breast and ovarian cancer due to BRCA1&2 hereditary cancer predisposition syndrome and reproduction: literature review*]. J Gynecol Obstet Biol Reprod (Paris), 2015. **44**(1): p. 10-7.
112. Easton, D.F., *How many more breast cancer predisposition genes are there?* Breast Cancer Res, 1999. **1**(1): p. 14-7.
113. Ambrosone, C.B., *The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer*. Breast Cancer Res, 2007. **9**(6): p. 114.
114. Thompson, E.R., et al., *Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles*. PLoS Genet, 2012. **8**(9): p. e1002894.
115. Kiiski, J.I., et al., *Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer*. Proc Natl Acad Sci U S A, 2014. **111**(42): p. 15172-7.
116. Noh, J.M., et al., *Exome sequencing in a breast cancer family without BRCA mutation*. Radiat Oncol J, 2015. **33**(2): p. 149-54.
117. Meijers-Heijboer, H., et al., *Low-penetrance susceptibility to breast cancer due to CHEK2(\*)1100delC in noncarriers of BRCA1 or BRCA2 mutations*. Nat Genet, 2002. **31**(1): p. 55-9.
118. Vahteristo, P., et al., *A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer*. Am J Hum Genet, 2002. **71**(2): p. 432-8.
119. Rahman, N., et al., *PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene*. Nat Genet, 2007. **39**(2): p. 165-7.
120. Thompson, D., et al., *Cancer risks and mortality in heterozygous ATM mutation carriers*. J Natl Cancer Inst, 2005. **97**(11): p. 813-22.
121. Seal, S., et al., *Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles*. Nat Genet, 2006. **38**(11): p. 1239-41.
122. Mizzi, C., et al., *A European Spectrum of Pharmacogenomic Biomarkers: Implications for Clinical Pharmacogenomics*. PLoS One, 2016. **11**(9): p. e0162866.
123. Del Re, M., et al., *Pharmacogenetics and Metabolism from Science to Implementation in Clinical Practice: The Example of Dihydropyrimidine Dehydrogenase*. Curr Pharm Des, 2017. **23**(14): p. 2028-2034.

124. Johnson, J.A., et al., *Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update*. Clin Pharmacol Ther, 2017. **102**(3): p. 397-404.
125. Vaclova, T., et al., *Germline missense pathogenic variants in the BRCA1 BRCT domain, p.Gly1706Glu and p.Ala1708Glu, increase cellular sensitivity to PARP inhibitor olaparib by a dominant negative effect*. Hum Mol Genet, 2016. **25**(24): p. 5287-5299.
126. Zhao, P., et al., *Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy*. J Hematol Oncol, 2019. **12**(1): p. 54.
127. Shimada, S., et al., *The value of MDM2 and CDK4 amplification levels using real-time polymerase chain reaction for the differential diagnosis of liposarcomas and their histologic mimickers*. Hum Pathol, 2006. **37**(9): p. 1123-9.
128. Vogelstein, B., et al., *Cancer genome landscapes*. Science, 2013. **339**(6127): p. 1546-58.
129. Huang, J., et al., *Variant Profiling of Candidate Genes in Pancreatic Ductal Adenocarcinoma*. Clin Chem, 2015. **61**(11): p. 1408-16.
130. Cleary, S.P., et al., *Identification of driver genes in hepatocellular carcinoma by exome sequencing*. Hepatology, 2013. **58**(5): p. 1693-702.
131. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types*. Nature, 2013. **502**(7471): p. 333-339.
132. Churi, C.R., et al., *Mutation profiling in cholangiocarcinoma: prognostic and therapeutic implications*. PLoS One, 2014. **9**(12): p. e115383.
133. Tirode, F., et al., *Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations*. Cancer Discov, 2014. **4**(11): p. 1342-53.
134. Moreira, A.L., et al., *Massively parallel sequencing identifies recurrent mutations in TP53 in thymic carcinoma associated with poor prognosis*. J Thorac Oncol, 2015. **10**(2): p. 373-80.
135. Parry, M., et al., *Genetics and Prognostication in Splenic Marginal Zone Lymphoma: Revelations from Deep Sequencing*. Clin Cancer Res, 2015. **21**(18): p. 4174-4183.
136. Parkin, B., et al., *Integrated genomic profiling, therapy response, and survival in adult acute myelogenous leukemia*. Clin Cancer Res, 2015. **21**(9): p. 2045-56.
137. Kriegsmann, M., et al., *Mutational profiles in triple-negative breast cancer defined by ultradeep multigene sequencing show high rates of PI3K pathway alterations and clinically relevant entity subgroup specific differences*. Oncotarget, 2014. **5**(20): p. 9952-65.
138. de Albuquerque, A., et al., *Prognostic and predictive value of circulating tumor cell analysis in colorectal cancer patients*. J Transl Med, 2012. **10**: p. 222.
139. Hench, I.B., J. Hench, and M. Tolnay, *Liquid Biopsy in Clinical Management of Breast, Lung, and Colorectal Cancer*. Front Med (Lausanne), 2018. **5**: p. 9.
140. Walther, A., et al., *Genetic prognostic and predictive markers in colorectal cancer*. Nat Rev Cancer, 2009. **9**(7): p. 489-99.
141. Plimack, E.R., et al., *Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer*. Eur Urol, 2015. **68**(6): p. 959-67.
142. Li, Q., et al., *ERCC2 Helicase Domain Mutations Confer Nucleotide Excision Repair Deficiency and Drive Cisplatin Sensitivity in Muscle-Invasive Bladder Cancer*. Clin Cancer Res, 2019. **25**(3): p. 977-988.



143. Konstantinopoulos, P.A., et al., *Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer*. *Cancer Discov*, 2015. **5**(11): p. 1137-54.
144. Chila, R., F. Guffanti, and G. Damia, *Role and therapeutic potential of CDK12 in human cancers*. *Cancer Treat Rev*, 2016. **50**: p. 83-88.
145. Lui, G.Y.L., C. Grandori, and C.J. Kemp, *CDK12: an emerging therapeutic target for cancer*. *J Clin Pathol*, 2018. **71**(11): p. 957-962.
146. Ceccaldi, R., et al., *A unique subset of epithelial ovarian cancers with platinum sensitivity and PARP inhibitor resistance*. *Cancer Res*, 2015. **75**(4): p. 628-34.
147. Li, S., et al., *Genetic variants in RPA1 associated with the response to oxaliplatin-based chemotherapy in colorectal cancer*. *J Gastroenterol*, 2019.
148. Zhu, C., et al., *Somatic mutation of DNAH genes implicated higher chemotherapy response rate in gastric adenocarcinoma patients*. *J Transl Med*, 2019. **17**(1): p. 109.
149. Dhillon, K.K., E.M. Swisher, and T. Taniguchi, *Secondary mutations of BRCA1/2 and drug resistance*. *Cancer Sci*, 2011. **102**(4): p. 663-9.
150. Weigelt, B., et al., *Diverse BRCA1 and BRCA2 Reversion Mutations in Circulating Cell-Free DNA of Therapy-Resistant Breast or Ovarian Cancer*. *Clin Cancer Res*, 2017. **23**(21): p. 6708-6720.
151. Liu, D., et al., *Mutational patterns in chemotherapy resistant muscle-invasive bladder cancer*. *Nat Commun*, 2017. **8**(1): p. 2193.
152. Catucci, I., et al., *Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility*. *Genet Med*, 2018. **20**(4): p. 452-457.
153. Rizvi, N.A., et al., *Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer*. *Science*, 2015. **348**(6230): p. 124-8.
154. Van Allen, E.M., et al., *Genomic correlates of response to CTLA-4 blockade in metastatic melanoma*. *Science*, 2015. **350**(6257): p. 207-211.
155. Harris, L.N., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline*. *J Clin Oncol*, 2016. **34**(10): p. 1134-50.
156. Krop, I., et al., *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update*. *J Clin Oncol*, 2017. **35**(24): p. 2838-2847.
157. Hyman, D.M., B.S. Taylor, and J. Baselga, *Implementing Genome-Driven Oncology*. *Cell*, 2017. **168**(4): p. 584-599.
158. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2*. *N Engl J Med*, 2001. **344**(11): p. 783-92.
159. Druker, B.J., et al., *Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia*. *N Engl J Med*, 2001. **344**(14): p. 1031-7.
160. Peters, S., et al., *Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer*. *N Engl J Med*, 2017. **377**(9): p. 829-838.
161. Camidge, D.R., et al., *Exploratory Analysis of Brigatinib Activity in Patients With Anaplastic Lymphoma Kinase-Positive Non-Small-Cell Lung Cancer and Brain Metastases in Two Clinical Trials*. *J Clin Oncol*, 2018. **36**(26): p. 2693-2701.

162. Soria, J.C., et al., *First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study*. Lancet, 2017. **389**(10072): p. 917-929.
163. Kazandjian, D., et al., *FDA approval summary: crizotinib for the treatment of metastatic non-small cell lung cancer with anaplastic lymphoma kinase rearrangements*. Oncologist, 2014. **19**(10): p. e5-11.
164. Cortes, J.E., et al., *Bosutinib Versus Imatinib for Newly Diagnosed Chronic Myeloid Leukemia: Results From the Randomized BFORE Trial*. J Clin Oncol, 2018. **36**(3): p. 231-237.
165. Muller, M.C., et al., *Dasatinib treatment of chronic-phase chronic myeloid leukemia: analysis of responses according to preexisting BCR-ABL mutations*. Blood, 2009. **114**(24): p. 4944-53.
166. Johnson, J.R., et al., *Approval summary: imatinib mesylate capsules for treatment of adult patients with newly diagnosed philadelphia chromosome-positive chronic myelogenous leukemia in chronic phase*. Clin Cancer Res, 2003. **9**(6): p. 1972-9.
167. Schultz, K.R., et al., *Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study*. J Clin Oncol, 2009. **27**(31): p. 5175-81.
168. Kantarjian, H.M., et al., *Nilotinib versus imatinib for the treatment of patients with newly diagnosed chronic phase, Philadelphia chromosome-positive, chronic myeloid leukaemia: 24-month minimum follow-up of the phase 3 randomised ENESTnd trial*. Lancet Oncol, 2011. **12**(9): p. 841-51.
169. Cortes, J.E., et al., *A phase 2 trial of ponatinib in Philadelphia chromosome-positive leukemias*. N Engl J Med, 2013. **369**(19): p. 1783-96.
170. Long, G.V., et al., *Dabrafenib plus trametinib versus dabrafenib monotherapy in patients with metastatic BRAF V600E/K-mutant melanoma: long-term survival and safety analysis of a phase 3 study*. Ann Oncol, 2017. **28**(7): p. 1631-1639.
171. Planchard, D., et al., *Dabrafenib plus trametinib in patients with previously untreated BRAF(V600E)-mutant metastatic non-small-cell lung cancer: an open-label, phase 2 trial*. Lancet Oncol, 2017. **18**(10): p. 1307-1316.
172. Subbiah, V., et al., *Dabrafenib and Trametinib Treatment in Patients With Locally Advanced or Metastatic BRAF V600-Mutant Anaplastic Thyroid Cancer*. J Clin Oncol, 2018. **36**(1): p. 7-13.
173. Goetz, M.P., et al., *MONARCH 3: Abemaciclib As Initial Therapy for Advanced Breast Cancer*. J Clin Oncol, 2017. **35**(32): p. 3638-3646.
174. Turner, N.C., et al., *Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer*. N Engl J Med, 2015. **373**(3): p. 209-19.
175. Janni, W., et al., *First-line ribociclib plus letrozole in postmenopausal women with HR+ , HER2- advanced breast cancer: Tumor response and pain reduction in the phase 3 MONALEESA-2 trial*. Breast Cancer Res Treat, 2018. **169**(3): p. 469-479.
176. Hortobagyi, G.N., et al., *Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer*. Ann Oncol, 2018. **29**(7): p. 1541-1547.
177. Licitra, L., et al., *Predictive value of epidermal growth factor receptor expression for first-line chemotherapy plus cetuximab in patients with head and neck and colorectal cancer: analysis of data from the EXTREME and CRYSTAL studies*. Eur J Cancer, 2013. **49**(6): p. 1161-8.

178. Wang, J.P., et al., *Erlotinib is effective in pancreatic cancer with epidermal growth factor receptor mutations: a randomized, open-label, prospective trial*. *Oncotarget*, 2015. **6**(20): p. 18162-73.
179. Cicens, S., et al., *Maintenance erlotinib versus erlotinib at disease progression in patients with advanced non-small-cell lung cancer who have not progressed following platinum-based chemotherapy (IUNO study)*. *Lung Cancer*, 2016. **102**: p. 30-37.
180. Shin, S., et al., *Erlotinib plus gemcitabine versus gemcitabine for pancreatic cancer: real-world analysis of Korean national database*. *BMC Cancer*, 2016. **16**: p. 443.
181. Douillard, J.Y., et al., *First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study*. *Br J Cancer*, 2014. **110**(1): p. 55-62.
182. Reck, M., et al., *The Effect of Necitumumab in Combination with Gemcitabine plus Cisplatin on Tolerability and on Quality of Life: Results from the Phase 3 SQUIRE Trial*. *J Thorac Oncol*, 2016. **11**(6): p. 808-18.
183. Akamatsu, H., et al., *Osimertinib in Japanese patients with EGFR T790M mutation-positive advanced non-small-cell lung cancer: AURA3 trial*. *Cancer Sci*, 2018. **109**(6): p. 1930-1938.
184. Mok, T.S., et al., *Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer*. *N Engl J Med*, 2017. **376**(7): p. 629-640.
185. Soria, J.C., et al., *Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer*. *N Engl J Med*, 2018. **378**(2): p. 113-125.
186. Douillard, J.Y., et al., *Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer*. *Ann Oncol*, 2014. **25**(7): p. 1346-55.
187. Price, T.J., et al., *Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority phase 3 study*. *Lancet Oncol*, 2014. **15**(6): p. 569-79.
188. Yang, J.C., et al., *Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6*. *Lancet Oncol*, 2015. **16**(7): p. 830-8.
189. Dieras, V., et al., *Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial*. *Lancet Oncol*, 2017. **18**(6): p. 732-742.
190. Krop, I.E., et al., *Trastuzumab emtansine versus treatment of physician's choice in patients with previously treated HER2-positive metastatic breast cancer (TH3RESA): final overall survival results from a randomised open-label phase 3 trial*. *Lancet Oncol*, 2017. **18**(6): p. 743-754.
191. Bang, Y.J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial*. *Lancet*, 2010. **376**(9742): p. 687-97.
192. Cameron, D., et al., *11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial*. *Lancet*, 2017. **389**(10075): p. 1195-1205.

193. Gianni, L., et al., *Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial*. *Lancet Oncol*, 2011. **12**(3): p. 236-44.
194. von Minckwitz, G., et al., *Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer*. *N Engl J Med*, 2017. **377**(2): p. 122-131.
195. Ryan, Q., et al., *FDA drug approval summary: lapatinib in combination with capecitabine for previously treated metastatic breast cancer that overexpresses HER-2*. *Oncologist*, 2008. **13**(10): p. 1114-9.
196. Chan, A., et al., *Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial*. *Lancet Oncol*, 2016. **17**(3): p. 367-77.
197. Martin, M., et al., *Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial*. *Lancet Oncol*, 2017. **18**(12): p. 1688-1700.
198. Droogendijk, H.J., et al., *Imatinib mesylate in the treatment of systemic mastocytosis: a phase II trial*. *Cancer*, 2006. **107**(2): p. 345-51.
199. Cohen, M.H., et al., *Approval summary: imatinib mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors*. *Oncologist*, 2009. **14**(2): p. 174-80.
200. Yardley, D.A., et al., *Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis*. *Adv Ther*, 2013. **30**(10): p. 870-84.
201. Yao, J.C., et al., *Everolimus for advanced pancreatic neuroendocrine tumors*. *N Engl J Med*, 2011. **364**(6): p. 514-23.
202. Yao, J.C., et al., *Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study*. *Lancet*, 2016. **387**(10022): p. 968-977.
203. Motzer, R.J., et al., *Phase 3 trial of everolimus for metastatic renal cell carcinoma : final results and analysis of prognostic factors*. *Cancer*, 2010. **116**(18): p. 4256-65.
204. Franz, D.N., et al., *Efficacy and safety of everolimus for subependymal giant cell astrocytomas associated with tuberous sclerosis complex (EXIST-1): a multicentre, randomised, placebo-controlled phase 3 trial*. *Lancet*, 2013. **381**(9861): p. 125-32.
205. Apperley, J.F., et al., *Response to imatinib mesylate in patients with chronic myeloproliferative diseases with rearrangements of the platelet-derived growth factor receptor beta*. *N Engl J Med*, 2002. **347**(7): p. 481-7.
206. Rutkowski, P., et al., *Imatinib mesylate in advanced dermatofibrosarcoma protuberans: pooled analysis of two phase II clinical trials*. *J Clin Oncol*, 2010. **28**(10): p. 1772-9.
207. Ji, H., et al., *The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies*. *Cancer Cell*, 2006. **9**(6): p. 485-95.
208. Ahmadzada, T., et al., *An Update on Predictive Biomarkers for Treatment Selection in Non-Small Cell Lung Cancer*. *J Clin Med*, 2018. **7**(6).
209. Russo, A., et al., *A decade of EGFR inhibition in EGFR-mutated non small cell lung cancer (NSCLC): Old successes and future perspectives*. *Oncotarget*, 2015. **6**(29): p. 26814-25.
210. Jackman, D.M., et al., *Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer*

- patients: results of an online tumor registry of clinical trials. *Clin Cancer Res*, 2009. **15**(16): p. 5267-73.
211. Yang, J.C., et al., *Epidermal growth factor receptor mutation analysis in previously unanalyzed histology samples and cytology samples from the phase III Iressa Pan-ASia Study (IPASS)*. *Lung Cancer*, 2014. **83**(2): p. 174-81.
  212. Rosell, R., et al., *Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial*. *Lancet Oncol*, 2012. **13**(3): p. 239-46.
  213. De Roock, W., et al., *Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis*. *Lancet Oncol*, 2010. **11**(8): p. 753-62.
  214. Cheng, L., et al., *Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine*. *Mod Pathol*, 2018. **31**(1): p. 24-38.
  215. Hauschild, A., et al., *Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial*. *Lancet*, 2012. **380**(9839): p. 358-65.
  216. Chapman, P.B., et al., *Improved survival with vemurafenib in melanoma with BRAF V600E mutation*. *N Engl J Med*, 2011. **364**(26): p. 2507-16.
  217. Manzano, J.L., et al., *Resistant mechanisms to BRAF inhibitors in melanoma*. *Ann Transl Med*, 2016. **4**(12): p. 237.
  218. Rugo, H.S., et al., *Impact of palbociclib plus letrozole on patient-reported health-related quality of life: results from the PALOMA-2 trial*. *Ann Oncol*, 2018. **29**(4): p. 888-894.
  219. Le Tourneau, C., et al., *Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial*. *Lancet Oncol*, 2015. **16**(13): p. 1324-34.
  220. Tsimberidou, A.M., et al., *Personalized medicine in a phase I clinical trials program: the MD Anderson Cancer Center initiative*. *Clin Cancer Res*, 2012. **18**(22): p. 6373-83.
  221. Tsimberidou, A.M., et al., *Personalized medicine for patients with advanced cancer in the phase I program at MD Anderson: validation and landmark analyses*. *Clin Cancer Res*, 2014. **20**(18): p. 4827-36.
  222. Juric, D., et al., *Phase I Dose-Escalation Study of Taselisib, an Oral PI3K Inhibitor, in Patients with Advanced Solid Tumors*. *Cancer Discov*, 2017. **7**(7): p. 704-715.
  223. Jhaveri, K.L., et al., *Ado-trastuzumab emtansine (T-DM1) in patients with HER2-amplified tumors excluding breast and gastric/gastroesophageal junction (GEJ) adenocarcinomas: results from the NCI-MATCH trial (EAY131) subprotocol Q*. *Ann Oncol*, 2019. **30**(11): p. 1821-1830.
  224. Chae, Y.K., et al., *Phase II Study of AZD4547 in Patients With Tumors Harboring Aberrations in the FGFR Pathway: Results From the NCI-MATCH Trial (EAY131) Subprotocol W*. *J Clin Oncol*, 2020. **38**(21): p. 2407-2417.
  225. de Leng, W.W., et al., *Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material*. *PLoS One*, 2016. **11**(2): p. e0149405.

226. Froyen, G., et al., *Validation and Application of a Custom-Designed Targeted Next-Generation Sequencing Panel for the Diagnostic Mutational Profiling of Solid Tumors*. PLoS One, 2016. **11**(4): p. e0154038.
227. D'Haene, N., et al., *Clinical Validation of Targeted Next Generation Sequencing for Colon and Lung Cancers*. PLoS One, 2015. **10**(9): p. e0138245.
228. Spencer, D.H., et al., *Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens*. J Mol Diagn, 2013. **15**(5): p. 623-33.
229. Wall, J.D., et al., *Estimating genotype error rates from high-coverage next-generation sequence data*. Genome Res, 2014. **24**(11): p. 1734-9.
230. Shin, S. and J. Park, *Characterization of sequence-specific errors in various next-generation sequencing systems*. Mol Biosyst, 2016. **12**(3): p. 914-22.
231. Pritchard, C.C., et al., *Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens*. J Mol Diagn, 2014. **16**(1): p. 56-67.
232. Mandelker, D., et al., *Comprehensive diagnostic testing for stereocilin: an approach for analyzing medically important genes with high homology*. J Mol Diagn, 2014. **16**(6): p. 639-47.
233. Li, J., et al., *A Comprehensive Strategy for Accurate Mutation Detection of the Highly Homologous PMS2*. J Mol Diagn, 2015. **17**(5): p. 545-53.
234. Yohe, S.L., et al., *Standards for Clinical Grade Genomic Databases*. Arch Pathol Lab Med, 2015. **139**(11): p. 1400-12.
235. Pugh, T.J., et al., *VisCap: inference and visualization of germ-line copy-number variants from targeted clinical sequencing data*. Genet Med, 2016. **18**(7): p. 712-9.
236. Zhao, M., et al., *Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives*. BMC Bioinformatics, 2013. **14 Suppl 11**: p. S1.
237. Xu, J., et al., *Mortality in the United States, 2015*. NCHS Data Brief, 2016(267): p. 1-8.
238. Yoon, P.W., et al., *Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study*. Arch Pediatr Adolesc Med, 1997. **151**(11): p. 1096-103.
239. O'Malley, M. and R.G. Hutcheon, *Genetic disorders and congenital malformations in pediatric long-term care*. J Am Med Dir Assoc, 2007. **8**(5): p. 332-4.
240. Stevenson, D.A. and J.C. Carey, *Contribution of malformations and genetic disorders to mortality in a children's hospital*. Am J Med Genet A, 2004. **126A**(4): p. 393-7.
241. Petrikin, J.E., et al., *Rapid whole genome sequencing and precision neonatology*. Semin Perinatol, 2015. **39**(8): p. 623-31.
242. Miller, N.A., et al., *A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases*. Genome Med, 2015. **7**: p. 100.
243. South, S.T., et al., *ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013*. Genet Med, 2013. **15**(11): p. 901-9.
244. Miller, D.T., et al., *Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies*. Am J Hum Genet, 2010. **86**(5): p. 749-64.

245. Lionel, A.C., et al., *Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test*. *Genet Med*, 2018. **20**(4): p. 435-443.
246. van Egmond, M.E., et al., *A post hoc study on gene panel analysis for the diagnosis of dystonia*. *Mov Disord*, 2017. **32**(4): p. 569-575.
247. Montaut, S., et al., *Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders*. *JAMA Neurol*, 2018. **75**(10): p. 1234-1245.
248. Han, J.Y., et al., *Targeted Next-Generation Sequencing of Korean Patients With Developmental Delay and/or Intellectual Disability*. *Front Pediatr*, 2018. **6**: p. 391.
249. Retterer, K., et al., *Clinical application of whole-exome sequencing across clinical indications*. *Genet Med*, 2016. **18**(7): p. 696-704.
250. Vissers, L., et al., *A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology*. *Genet Med*, 2017. **19**(9): p. 1055-1063.
251. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2009. **25**(14): p. 1754-60.
252. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. *Genome Res*, 2010. **20**(9): p. 1297-303.
253. Liu, X., et al., *dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs*. *Hum Mutat*, 2016. **37**(3): p. 235-41.
254. Sarkar, A., et al., *Novel de novo mutations in EFTUD2 detected by exome sequencing in mandibulofacial dysostosis with Microcephaly syndrome*. *Am J Med Genet A*, 2015. **167A**(4): p. 914-8.
255. Rylaarsdam, L. and A. Guemez-Gamboa, *Genetic Causes and Modifiers of Autism Spectrum Disorder*. *Front Cell Neurosci*, 2019. **13**: p. 385.
256. Varga, E.A., et al., *The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly*. *Genet Med*, 2009. **11**(2): p. 111-7.
257. McMillan, H.J., et al., *Recessive mutations in ATP8A2 cause severe hypotonia, cognitive impairment, hyperkinetic movement disorders and progressive optic atrophy*. *Orphanet J Rare Dis*, 2018. **13**(1): p. 86.
258. Hartman, P., et al., *Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory*. *Mol Genet Metab Rep*, 2019. **19**: p. 100464.
259. Graziola, F., et al., *Diagnostic Yield of a Targeted Next-Generation Sequencing Gene Panel for Pediatric-Onset Movement Disorders: A 3-Year Cohort Study*. *Front Genet*, 2019. **10**: p. 1026.
260. Walsh, M., et al., *Diagnostic and cost utility of whole exome sequencing in peripheral neuropathy*. *Ann Clin Transl Neurol*, 2017. **4**(5): p. 318-325.
261. Meng, L., et al., *Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management*. *JAMA Pediatr*, 2017. **171**(12): p. e173438.
262. Feliciano, P., et al., *Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes*. *NPJ Genom Med*, 2019. **4**: p. 19.
263. Caceres, E.F. and L.D. Hurst, *The evolution, impact and properties of exonic splice enhancers*. *Genome Biol*, 2013. **14**(12): p. R143.
264. Savisaar, R. and L.D. Hurst, *Exonic splice regulation imposes strong selection at synonymous sites*. *Genome Res*, 2018. **28**(10): p. 1442-1454.

265. Zhao, L., et al., *Comparative study of whole exome sequencing-based copy number variation detection tools*. BMC Bioinformatics, 2020. **21**(1): p. 97.
266. Gerdes, A.M., et al., *Evaluation of two different models to predict BRCA1 and BRCA2 mutations in a cohort of Danish hereditary breast and/or ovarian cancer families*. Clin Genet, 2006. **69**(2): p. 171-8.
267. Melchor, L. and J. Benitez, *The complex genetic landscape of familial breast cancer*. Hum Genet, 2013. **132**(8): p. 845-63.
268. Hollestelle, A., et al., *Discovering moderate-risk breast cancer susceptibility genes*. Curr Opin Genet Dev, 2010. **20**(3): p. 268-76.
269. Renwick, A., et al., *ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles*. Nat Genet, 2006. **38**(8): p. 873-5.
270. Steffen, J., et al., *Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland*. Int J Cancer, 2006. **119**(2): p. 472-5.
271. Michailidou, K., et al., *Association analysis identifies 65 new breast cancer risk loci*. Nature, 2017. **551**(7678): p. 92-94.
272. Shendure, J. and H. Ji, *Next-generation DNA sequencing*. Nat Biotechnol, 2008. **26**(10): p. 1135-45.
273. Kim, Y.C., et al., *Unique Features of Germline Variation in Five Egyptian Familial Breast Cancer Families Revealed by Exome Sequencing*. PLoS One, 2017. **12**(1): p. e0167581.
274. Wen, H., et al., *Family-specific, novel, deleterious germline variants provide a rich resource to identify genetic predispositions for BRCAx familial breast cancer*. BMC Cancer, 2014. **14**: p. 470.
275. Lynch, H., et al., *Can unknown predisposition in familial breast cancer be family-specific?* Breast J, 2013. **19**(5): p. 520-8.
276. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery*. Nat Rev Genet, 2011. **12**(11): p. 745-55.
277. Decker, B., et al., *Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks*. J Med Genet, 2017. **54**(11): p. 732-741.
278. Li, Z., et al., *Destabilization of linker histone H1.2 is essential for ATM activation and DNA damage repair*. Cell Res, 2018. **28**(7): p. 756-770.
279. Smart, S.K., et al., *The Emerging Role of TYRO3 as a Therapeutic Target in Cancer*. Cancers (Basel), 2018. **10**(12).
280. Ekyalongo, R.C., et al., *TYRO3 as a potential therapeutic target in breast cancer*. Anticancer Res, 2014. **34**(7): p. 3337-45.
281. Gautam, J., et al., *Tryptophan hydroxylase 1 and 5-HT7 receptor preferentially expressed in triple-negative breast cancer promote cancer progression through autocrine serotonin signaling*. Mol Cancer, 2016. **15**(1): p. 75.
282. Wang, Z., et al., *Roles of F-box proteins in cancer*. Nat Rev Cancer, 2014. **14**(4): p. 233-47.
283. D'Angiolella, V., et al., *SCF(Cyclin F) controls centrosome homeostasis and mitotic fidelity through CP110 degradation*. Nature, 2010. **466**(7302): p. 138-42.
284. D'Angiolella, V., M. Esencay, and M. Pagano, *A cyclin without cyclin-dependent kinases: cyclin F controls genome stability through ubiquitin-mediated proteolysis*. Trends Cell Biol, 2013. **23**(3): p. 135-40.
285. D'Angiolella, V., et al., *Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair*. Cell, 2012. **149**(5): p. 1023-34.



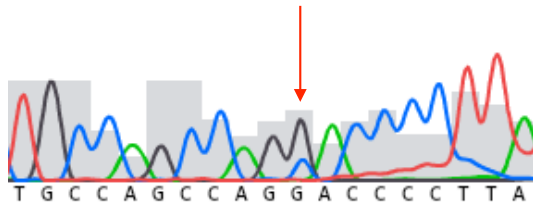
286. Shiovitz, S. and L.A. Korde, *Genetics of breast cancer: a topic in evolution*. Ann Oncol, 2015. **26**(7): p. 1291-9.
287. Yin, J., et al., *Revisiting Non-BRCA1/2 Familial Whole Exome Sequencing Datasets Implicates NCK1 as a Cancer Gene*. Front Genet, 2019. **10**: p. 527.
288. Hamdi, Y., et al., *Family specific genetic predisposition to breast cancer: results from Tunisian whole exome sequenced breast cancer cases*. J Transl Med, 2018. **16**(1): p. 158.
289. Gracia-Aznarez, F.J., et al., *Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles*. PLoS One, 2013. **8**(2): p. e55681.
290. Park, D.J., et al., *Rare mutations in XRCC2 increase the risk of breast cancer*. Am J Hum Genet, 2012. **90**(4): p. 734-9.
291. Park, D.J., et al., *Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers*. Cancer Discov, 2014. **4**(7): p. 804-15.
292. Gusev, A., et al., *Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases*. Am J Hum Genet, 2014. **95**(5): p. 535-52.
293. Koch, L., *Cancer genomics: Non-coding mutations in the driver seat*. Nat Rev Genet, 2014. **15**(9): p. 574-5.
294. Ongen, H., et al., *Putative cis-regulatory drivers in colorectal cancer*. Nature, 2014. **512**(7512): p. 87-90.
295. Chiba, K., et al., *Cancer-associated TERT promoter mutations abrogate telomerase silencing*. Elife, 2015. **4**.
296. Bojesen, S.E., et al., *Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer*. Nat Genet, 2013. **45**(4): p. 371-84, 384e1-2.
297. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma*. Science, 2013. **339**(6122): p. 959-61.
298. Jenkins, R.B., et al., *A low-frequency variant at 8q24.21 is strongly associated with risk of oligodendroglial tumors and astrocytomas with IDH1 or IDH2 mutation*. Nat Genet, 2012. **44**(10): p. 1122-5.
299. Zehir, A., et al., *Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients*. Nat Med, 2017. **23**(6): p. 703-713.
300. Salvatore, L., et al., *Management of metastatic colorectal cancer patients: guidelines of the Italian Medical Oncology Association (AIOM)*. ESMO Open, 2017. **2**(1): p. e000147.
301. Kerr, K.M., et al., *Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer*. Ann Oncol, 2014. **25**(9): p. 1681-90.
302. Dummer, R., et al., *Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Ann Oncol, 2015. **26 Suppl 5**: p. v126-32.
303. Manolio, T.A., et al., *Opportunities, resources, and techniques for implementing genomics in clinical care*. Lancet, 2019. **394**(10197): p. 511-520.
304. Berger, M.F. and E.R. Mardis, *The emerging clinical relevance of genomics in cancer medicine*. Nat Rev Clin Oncol, 2018. **15**(6): p. 353-365.
305. Diaz-Gay, M., et al., *Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples*. BMC Bioinformatics, 2018. **19**(1): p. 224.

306. Campbell, B.B., et al., *Comprehensive Analysis of Hypermutation in Human Cancer*. Cell, 2017. **171**(5): p. 1042-1056 e10.
307. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer*. Nature, 2013. **500**(7463): p. 415-21.
308. Van Hoeck, A., et al., *Portrait of a cancer: mutational signature analyses for cancer diagnostics*. BMC Cancer, 2019. **19**(1): p. 457.
309. Zhu, J., et al., *Association Between Tumor Mutation Burden (TMB) and Outcomes of Cancer Patients Treated With PD-1/PD-L1 Inhibitions: A Meta-Analysis*. Front Pharmacol, 2019. **10**: p. 673.
310. Goodman, A.M., et al., *Microsatellite-Stable Tumors with High Mutational Burden Benefit from Immunotherapy*. Cancer Immunol Res, 2019. **7**(10): p. 1570-1573.
311. Fong, P.C., et al., *Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers*. N Engl J Med, 2009. **361**(2): p. 123-34.
312. Hillman, S.C., et al., *Prenatal exome sequencing for fetuses with structural abnormalities: the next step*. Ultrasound Obstet Gynecol, 2015. **45**(1): p. 4-9.
313. Jelin, A.C. and N. Vora, *Whole Exome Sequencing: Applications in Prenatal Genetics*. Obstet Gynecol Clin North Am, 2018. **45**(1): p. 69-81.
314. van den Veyver, I.B. and C.M. Eng, *Genome-Wide Sequencing for Prenatal Detection of Fetal Single-Gene Disorders*. Cold Spring Harb Perspect Med, 2015. **5**(10).
315. Pujol, P., et al., *Guidelines for reporting secondary findings of genome sequencing in cancer genes: the SFMPP recommendations*. Eur J Hum Genet, 2018. **26**(12): p. 1732-1742.

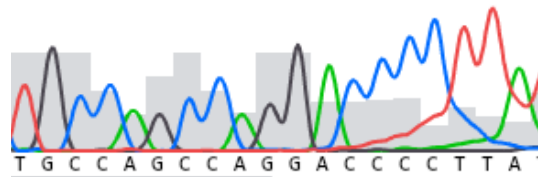
## Figures supplémentaires

TYRO3 : c.G2401C

I-7

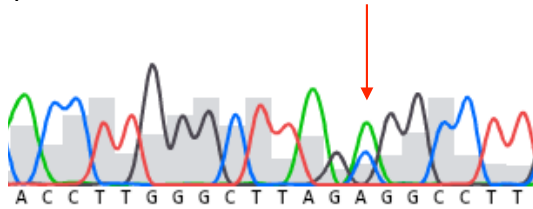


I-6

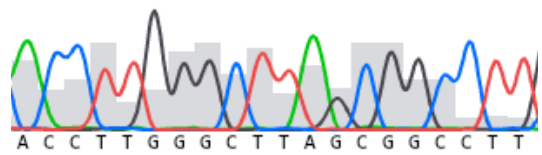


HIST1H1C : c.G592T

I-7

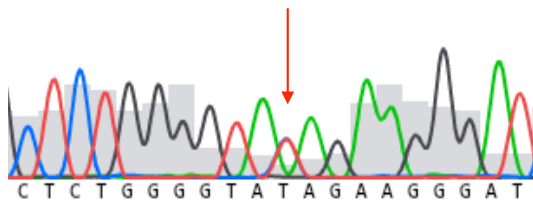


I-6

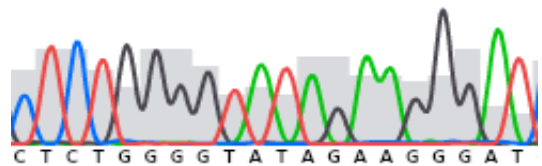


TPH1 : c.A791G

I-7

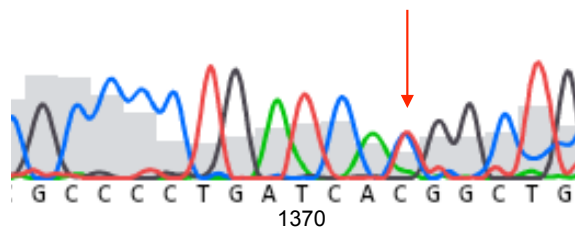


I-6



**Figure 1** : Confirmation des variants d'intérêt de la famille I par séquençage Sanger de l'individu porteur I-7 et du contrôle négatif I-6.

SLC12A3 : c.C1373T



**Figure 2** : Confirmation du variant candidat dans le gène *SLC12A3* chez les individus de la famille II

# Annexe 1: Exemple de rapport d'analyse des résultats de WES en trio pour le diagnostic de maladies génétiques rares



Predictive medicine impacting **Life**



GEMA: Germline Mutational Analysis

prenostics

PRENOSTICS SAS  
RCS 812 943 660 00017 - TVA: FR53 812 943 660  
E-Mail: [contact@prenostics.com](mailto:contact@prenostics.com)  
Web: [www.prenostics.com](http://www.prenostics.com)

Main Office: 14 Avenue de  
l'Observatoire 75006 Paris  
FRANCE, Phone.: +33781232791



prenostics

**PATIENT :**

Nom, prénom :

Date de naissance :

Sexe :

Numéro :

Médecin :

**Contexte clinique:**

L'étude a consisté en l'analyse d'exome en trio du fœtus et de ses parents M. et Mme .

Le fœtus présente les caractéristiques suivantes :

- Fœtus de sexe masculin ayant un poids et une taille comprise entre les 10ème et 25ème percentiles et un âge statural osseux en faveur d'un terme de 22SA.
- Microcéphalie avec front étroit et fuyant, racine du nez large et narines antéversées, discret hypertélorisme et oreilles basses implantées.
- Discrètes anomalies des extrémités.
- Isomérisme pulmonaire gauche sans autre malformation.
- L'examen neuropathologique montre une microlissencéphalie avec cortex à densité neuronale diminuée et microcolonnes de la couche II (type microcéphalie vera).
- Il s'y associe une héli-atrophie cérébelleuse droite, sans cause vasculaire hémorragique identifiée.
- Caryotype et ACPA normaux.

Concernant la grossesse:

L'échographie T2 à 22SA+2 met en évidence une microcéphalie sévère avec front fuyant et profil inhabituel. Hypotrophie cérébelleuse et cérébrale. RCIU < 5ème percentile. Doppler ombilical, Ductus normaux.

**Echantillons :**

Echantillons du fœtus fournis par le service de génétique de l'hôpital Américain.

**Séquençage:**

Le séquençage de l'exome a été réalisé sur plateforme Illumina. L'analyse thérapeutique a été réalisée en utilisant la matrice de donnée Prenostics, regroupant les données dans les bases de type OMIM, pubmed... Une analyse des voies de signalisations ainsi qu'un recoupement avec les données de la littérature ont été réalisées et contrôlées par Prenostics manuellement.

**Phénotype: Fœtus présentant un retard de croissance et une microcéphalie**

**Test réalisés:** Séquençage de l'exome du fœtus

**Résultats:**

Les résultats de l'analyse d'exome ont identifié au total 6059 variants chez le foetus.

Nous avons recoupé ces gènes avec des gènes de prédisposition aux différents troubles observés.

En particulier nous avons spécifiquement regardé les gènes impliqués dans les retards de croissance intra-utérin et les microcéphalies et en particulier la microcéphalie vera. Nous avons filtré les mutations pour ne garder que les mutations rares et avons exclu les mutations bénignes avérées

**Mutations hétérozygotes composites et *de novo***

Dans ce contexte où les parents sont non apparentés et sains, et au vu de la récurrence de microcéphalie chez ce couple, nous avons en priorité étudié les mutations hétérozygotes composites. Nous avons également étudié les mutations *de novo* chez l'enfant.

Nous avons identifié des mutations hétérozygotes composites dans le gène *RTTN*. Cela signifie que ce gène comporte 2 mutations hétérozygotes différentes chez le fœtus, l'une provenant de la mère et l'autre du père, altérant ainsi les deux allèles du gène à la manière des mutations homozygotes. Les mutations identifiées sont les suivantes :

- NM\_173630.4: c.1917G>T ; p.T639T héritée du père
- NM\_173630.4: c.2284C>T ; p.R762\* héritée de la mère

| Name                             | Gene Symbol | Clinical Significance  | Position (hg19) | Fréquence     | Chr | Ref allele fraction<br>50 = hétérozygote<br>100 = homozygote | Read count                          | Phenotype   |
|----------------------------------|-------------|------------------------|-----------------|---------------|-----|--|-------------------------------------|---|
| NM_173630.4: c.1917G>T ; p.T639T | <i>RTTN</i> | Uncertain significance | chr18:67741142  | 0% ou <0,001% | 18  | Fœtus : 35<br>Père : 42<br>Mère : -                          | Fœtus : 51<br>Père : 35<br>Mère : - | Microcéphalie, petite taille et polymicrogyrie avec convulsions |
| NM_173630.4: c.2284C>T ; p.R762* | <i>RTTN</i> | Uncertain significance | chr18:67816162  | 0,003%        | 18  | Fœtus : 43<br>Père : -<br>Mère : 50                          | Fœtus : 78<br>Père : -<br>Mère : 71 | Microcéphalie, petite taille et polymicrogyrie avec convulsions |

Ces deux mutations sont prédites pour être délétères pour la fonction du gène *RTTN*. La première provenant du père est une mutation synonyme mais fortement susceptible d'affecter un site d'épissage et de provoquer un changement du cadre de lecture du gène.

La seconde provenant de la mère est une mutation non sens qui provoque l'apparition d'un codon stop prématuré qui résulte en une protéine tronquée non fonctionnelle.

Des mutations bi-alléliques du gène *RTTN* sont bien connues pour causer dès la vie fœtale une microcéphalie sévère primaire, des malformations du cerveau incluant polymicrogyrie et lissencéphalie, et un retard de croissance important (voir citations).

Ce phénotype induit par des mutations dans le gène *RTTN* se transmet de manière autosomique récessive, ce qui signifie qu'il s'exprime que lorsque les deux allèles sont mutés, soit par une mutation homozygote, soit par des mutations hétérozygotes composites comme c'est le cas ici.

### **Conclusion**

**Nous avons identifié deux mutations hétérozygotes probablement délétères dans le gène *RTTN* chez le fœtus :**

- NM\_173630.4: c.1917G>T ; p.T639T héritée du père, et
- NM\_173630.4: c.2284C>T ; p.R762\* héritée de la mère.

**Ces mutations résultent en une mutation hétérozygote composite chez le fœtus chez qui les deux copies du gène sont mutées. D'après la littérature, les mutations bi-alléliques de *RTTN* causent les mêmes symptômes que ceux décrits chez le patient (microcéphalie primaire, lissencéphalie, polymicrogyrie, retard de croissance, voir citations).**

**Cette mutations bi-allélique de *RTTN* est fort probablement responsable du phénotype observé chez le fœtus, et explique son mode de transmission au sein de cette famille.**

**Citations sélectionnées :**

**Primary microcephaly, primordial dwarfism and brachydactyly in adult cases with bi-allelic skipping of RTTN exon 42.**

Zakaria M, Fatima A, Klar J, Wikström J, Abdullah U, Ali Z, Akram T, Tariq M, Ahmad H, Schuster J, Baig SM, Dahl N.

Hum Mutat. 2019 Mar 30. doi: 10.1002/humu.23755.

**Heterogeneous clinical phenotypes and cerebral malformations reflected by rotatin cellular dynamics.**

Vandervore LV, Schot R, Kasteleijn E, Oegema R, Stouffs K, Gheldof A, Grochowska MM, van der Sterre MLT, van Unen LMA, Wilke M, Elfferich P, van der Spek PJ, Heijsman D, Grandone A, Demmers JAA, Dekkers DHW, Slotman JA, Kremers GJ, Schaaf GJ, Masius RG, van Essen AJ, Rump P, van Haeringen A, Peeters E, Altunoglu U, Kalayci T, Poot RA, Dobyns WB, Bahi-Buisson N, Verheijen FW, Jansen AC, Mancini GMS.

Brain. 2019 Apr 1;142(4):867-884. doi: 10.1093/brain/awz045.

**Recurrent RTTN mutation leading to severe microcephaly, polymicrogyria and growth restriction.**

Cavallin M, Bery A, Maillard C, Salomon LJ, Bole C, Reilly ML, Nitschké P, Boddaert N, Bahi-Buisson N.

Eur J Med Genet. 2018 Dec;61(12):755-758. doi: 10.1016/j.ejmg.2018.08.001. Epub 2018 Aug 16.

**Functional characterization of biallelic RTTN variants identified in an infant with microcephaly, simplified gyral pattern, pontocerebellar hypoplasia, and seizures.**

Wambach JA, Wegner DJ, Yang P, Shinawi M, Baldrige D, Betleja E, Shimony JS, Spencer D, Hackett BP, Andrews MV, Ferkol T, Dutcher SK, Mahjoub MR, Cole FS.

Pediatr Res. 2018 Sep;84(3):435-441. doi: 10.1038/s41390-018-0083-z.

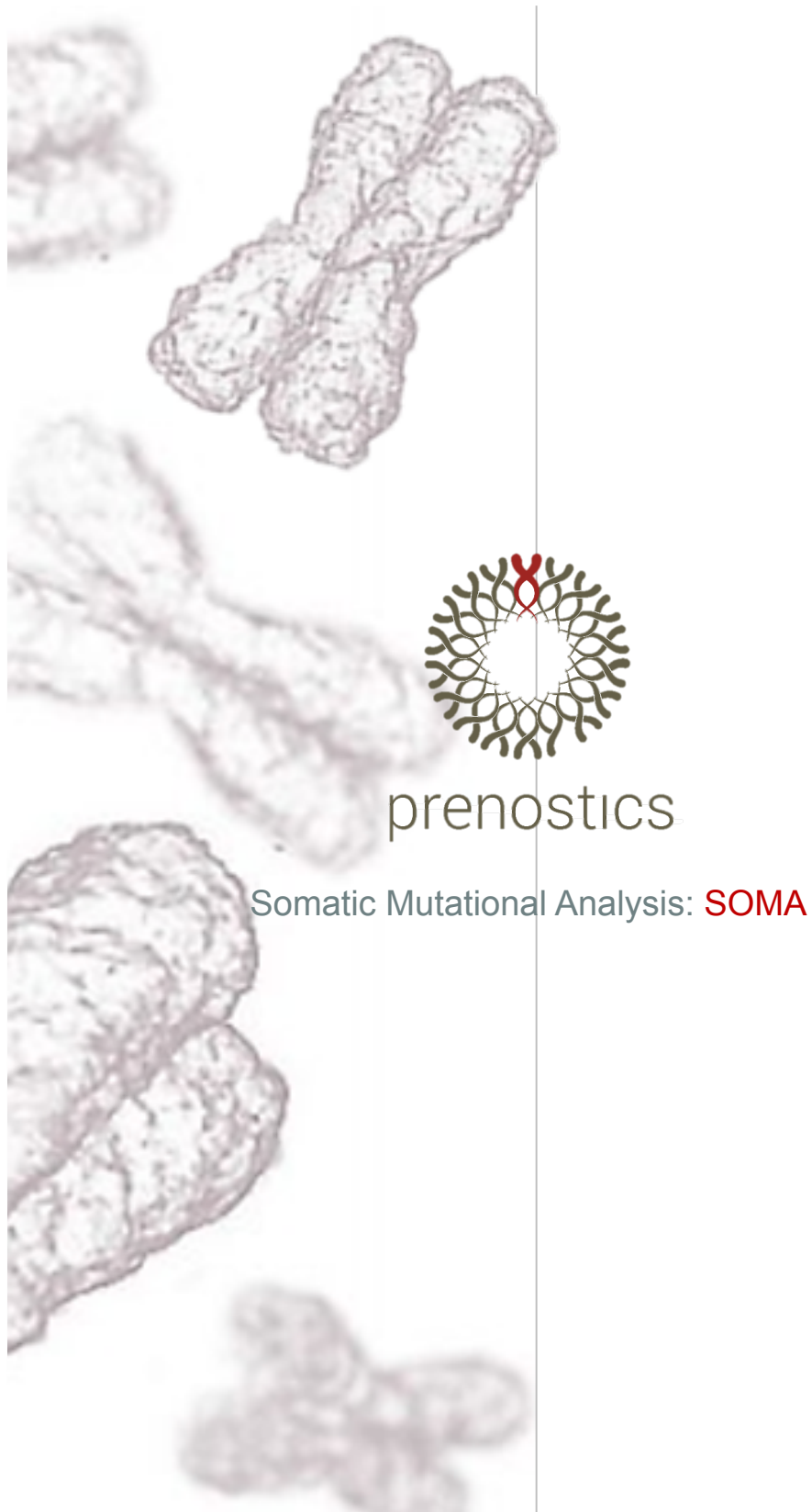
**Biallelic mutations in RTTN are associated with microcephaly, short stature and a wide range of brain malformations.**

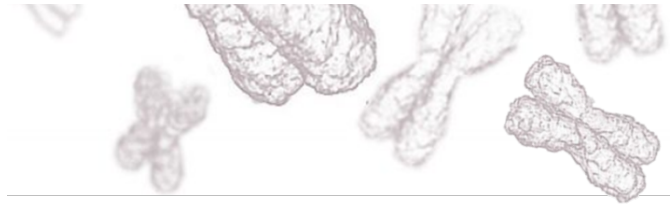
Stouffs K, Moortgat S, Vanderhasselt T, Vandervore L, Dica A, Mathot M, Keymolen K, Seneca S, Gheldof A, De Meirleir L, Jansen AC.

Eur J Med Genet. 2018 Dec;61(12):733-737. doi: 10.1016/j.ejmg.2018.06.001.



## Annexe 2 : Exemple de rapport d'analyse des résultats de WES du génome somatique et constitutionnel pour le profilage moléculaire de tumeurs





## PATIENT

**Nom, prénom :**  
**Sexe :**  
**Date de naissance :**  
**Numéro :**  
**Médecin :**



## CONTEXTE CLINIQUE

Adénocarcinome gastrique d'évolution lente, chez un patient dont le frère est décédé d'un cancer du pancréas. Le patient a reçu les traitements Folfox et Feloda, avant une détérioration de son état général. Après la mise en place d'une optimisation il répond de nouveau à un traitement par Folfox.



## ANALYSE GÉNÉTIQUE

- Nous avons identifié un total de **425 variants somatiques dans le tissu tumoral** qui passent les filtres des seuils de confiance. En moyenne nous avons donc une charge mutationnelle 10,4 mutations passant le seuil de confiance/MB. Ce taux mutationnel est relativement élevé pour un cancer gastrique mais il n'entre pas dans les tumeurs gastriques hyper-mutées.
- Au niveau du génome tumoral, nous avons identifié des mutations dans les gènes PIK3CA, ARID1A et APC, fréquemment mutés dans les tumeurs gastriques. Toutefois les mutations identifiées dans ces gènes sont peu décrites dans la littérature et leur pathogénicité n'est pas reconnue. Les mutations PIK3CA peuvent être ciblées dans les cancers du sein.
- Au niveau du génome germinale on retrouve des mutations potentiellement à risque dans le gène *BRCA2* et une mutation à impact inconnu dans le gène *MET*.



## ANALYSE THÉRAPEUTIQUE INTÉGRÉE

### THÉRAPIES AVEC UN POTENTIEL BÉNÉFICÉ

#### Chimiothérapie

- Poursuite Folfox
- Taxol
- Folfiri

#### Thérapie ciblée (somatique)

- Ciblage de PIK3CA/AKT/MTOR
- Anti-angiogénique tel que Ramicurimab

#### Thérapie ciblée (germinale)

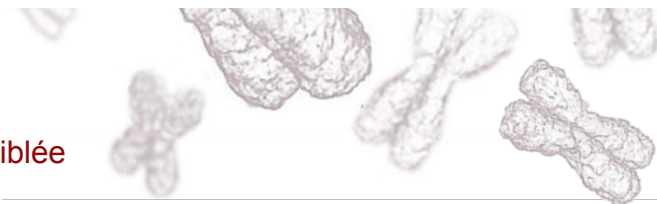
- Ciblage de BRCA2 par inhibiteur de PARP: Olaparib

### PRÉDICTION DE RÉPONSE AU TRAITEMENT

Le profil génétique de la tumeur a une charge mutationnelle dans la moyenne haute des cancers gastriques classiques, ce pourrait indiquer une réponse possible à une immunothérapie de type inhibiteur de checkpoint associée avec une chimiothérapie. La mutation germinale de BRCA2 doit être vérifiée par un laboratoire de biologie classique et une thérapie par inhibiteur de PARP devrait être discutée avec votre oncologue. La poursuite d'une chimiothérapie à base de platine paraît pertinente pour ce patient qui a une altération de la voie BRCA2. Finalement le profil génétique avec une anomalie de la voie AKT pourrait indiquer une sensibilité tumorale aux anti-angiogéniques.

# Avant-propos

De votre génome à votre thérapie ciblée



## PATIENT :

Nom, prénom :

Sexe :

Numéro :

Docteur :

## ÉCHANTILLON :

Site de la tumeur primaire :

Situation clinique :

Date du prélèvement :

Site du prélèvement :

Date du test génétique :

## SÉQUENCAGE :

Nous avons procédé à l'analyse par séquençage des exons après carottage tumoral et nous avons comparé le tissu tumoral avec le génome constitutionnel (échantillon salive). Le séquençage des exons a été réalisé sur plateforme Illumina. Le tableau ci-dessous montre les statistiques d'efficacité de votre séquençage ce qui nous permet une analyse pertinente de votre génome.

| Echantillon | Nombre total de lectures<br>non alignées | Lecture finale<br>alignée | Couverture moyenne dans les<br>régions « MedExome » |
|-------------|--|---------------------------|---|
| Salive      | 102 752 132                              | 79 891 464                | 146   |
| Tumeur      | 55 059 170                               | 39 553 024                | 69  |

L'analyse thérapeutique a été réalisée en utilisant la matrice de données PRENOSTICS. Les mutations pouvant constituer des cibles ont été analysées par un algorithme de prédiction et par un groupe d'experts prenant en compte l'histoire complète de votre maladie (localisation de la tumeur et traitements déjà reçus). L'analyse complète des voies de signalisations ainsi qu'un recoupement avec les données de la littérature ont été réalisées et contrôlées manuellement et vous seront présentés dans les pages de ce rapport.

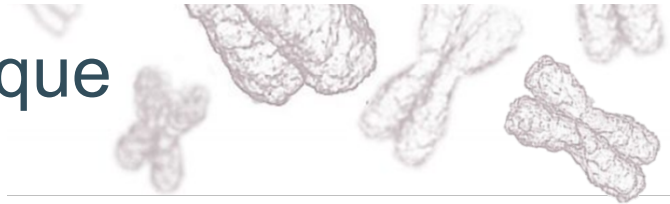
## Nota Bene:

Les thérapies associées avec un bénéfice potentiel ou une absence de bénéfice résultent de l'analyse génétique de votre tumeur confrontée à l'analyse de la littérature médicale. Les études peuvent concerner votre type tumoral ou un autre type tumoral présentant les mêmes caractéristiques biologiques. La découverte de mutations doit être considérée en fonction du contexte clinique. La spécificité de PRENOSTICS est de faire appel à des experts de thérapie ciblée pour analyser le profil des mutations de votre tumeur en fonction du contexte clinique pour prédire les traitements les plus appropriés. Le plus souvent plusieurs mutations sont découvertes au sein d'une tumeur.

L'avis d'expert PRENOSTICS est un avis consultatif donné par un expert de votre tumeur en fonction des données de votre analyse génétique. La décision thérapeutique finale appartient à votre médecin traitant et doit prendre en compte en plus de ce rapport votre état général, l'évolution de votre maladie et les protocoles thérapeutiques standards.

# Analyse Génétique

Votre génome, Vos mutations



## VUE D'ENSEMBLE

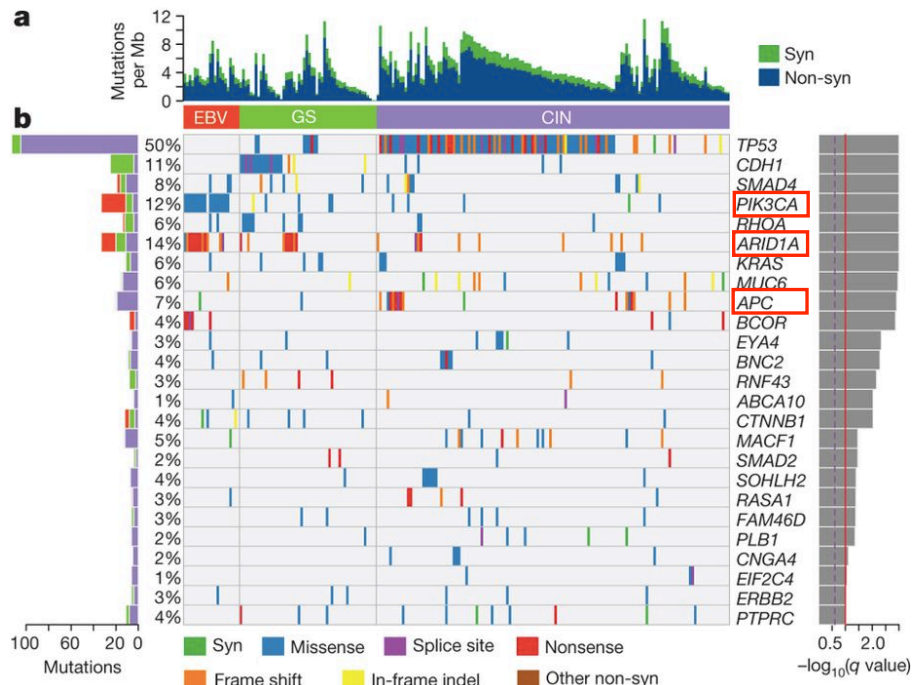
Nous avons analysé les mutations de votre tumeur gastrique et les avons comparé aux mutations identifiées à partir de votre échantillon salivaire. Les mutations uniquement présentes dans la tumeur sont dites somatiques (elles ont été acquises localement pendant le développement tumoral), celles présentes dans la salive sont dites germinales (elles sont acquises à la naissance).

Nous nous sommes tout d'abord intéressés aux mutations somatiques tumorales. Votre tumeur comporte de possibles mutations « driver » parmi les gènes les plus souvent mutés dans les cancers gastriques. Nous avons en effet retrouvé des mutations au niveau des gènes *PIK3CA*, *ARID1A* et *APC*. L'impact de ces mutations sur la fonction et l'activité de ces gènes n'est cependant pas identifié dans la littérature.

Les mutations *PIK3CA*, *ARID1A* et *APC* sont observées dans respectivement 12%, 14% 7% des cancers gastriques (Figure 1).

Les mutations *PIK3CA* sont ciblables par des thérapies ciblées dans le cancer du sein.

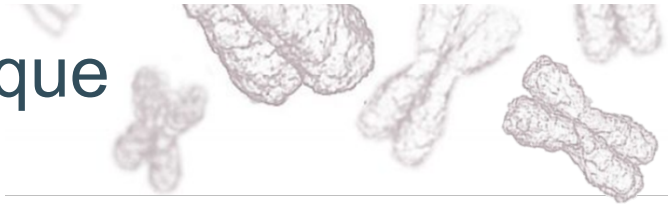
Au niveau du génome germlinal (échantillon salivaire) nous avons identifié des mutations pathogéniques dans le gène *BRCA2* et *MET*.



Vue d'ensemble des altérations génomiques fréquemment retrouvées dans les cancers de l'estomac (The Cancer Genome Atlas Research Network, nature, 2014)

# Analyse Génétique

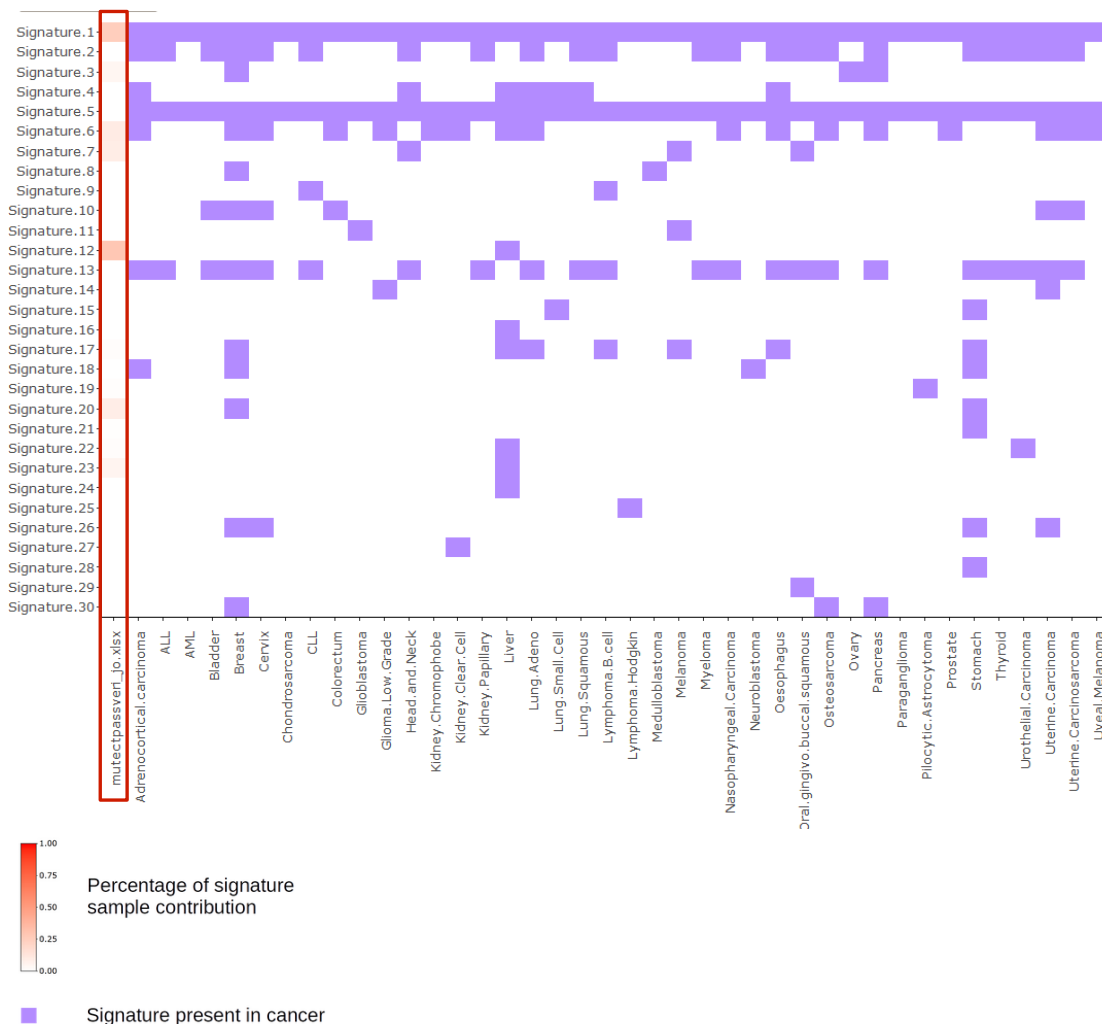
Votre génome, Vos mutations



## VOTRE SIGNATURE MUTATIONNELLE

La signature mutationnelle a été décrite pour la première fois par Alexandrov LB en 2013. Il décrit 30 signatures différentes de mutations qui peuvent être présentes au sein d'une tumeur (tout site confondus). Ces différentes signatures peuvent permettre de classer les cancers et de prédire la réponse à certains traitements ainsi qu'avoir une idée sur le pronostic. Ci-dessous, vous pouvez retrouver les 30 signatures décrites à ce jour comparée à votre signature.

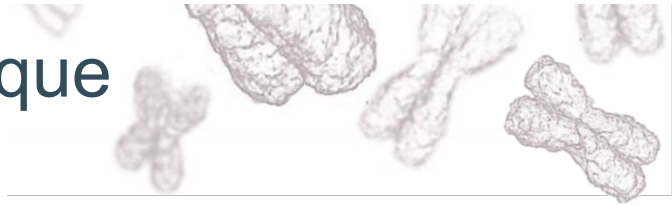
**Votre signature** (à l'extrême gauche) semble correspondre à un mélange entre la signature 12 et la signature 1.



<http://cancer.sanger.ac.uk/cosmic/signatures>

# Analyse Génétique

Votre génome, Vos mutations



## VOTRE SIGNATURE MUTATIONNELLE

| Signature | Possible Etiologie                                 |
|-----------|--|
| 12        | Unknown (liver cancer)                             |
| 1         | Age  |
| 6         | Defective DNA MMR / MSI (small INDELs)             |
| 7         | UV light   |
| 20        | Defective DNA MMR (small INDELs)                   |
| 23        | Unknown (liver cancer)                             |
| 3         | BRCA1 / BRCA2 (failure of DNA DSBR / large INDELs) |
| 17        | Unknown (different cancers)                        |
| 22        | Aristolochic acid                                  |
| 19        | Unknown (pilocytic astrocytoma)                    |
| 21        | Unknown (stomach cancer / MSI)                     |

# Annexes

Pour tout comprendre



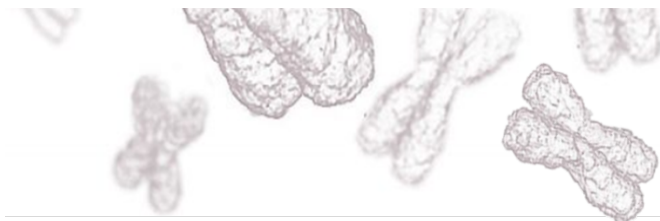
## Annexe 1 : Méthodologie bio-informatique

### Method Details:

- Sample Sources: Normal from Saliva and Tumor from biopsy
- Exome Library preparation: Truseq exome kit (Illumina)
- Sequencing: Illumina NextSeq500 Paired End 2x100
- FastQ Generation: Casava 1.8
- Reference: GATK bundle 2.8, b37
- Alignment: BWA 0.7.12
- Indexing: Samtools 1.2
- Duplicate removal: PICARD 1.110
- Indel Realignment and base correction: GATK 3.8
- Variant Calling: GATK HaplotypeCaller
- Somatic mutation calling: MUTECT2
- Variant Annotation: SNPEFF/SNPSIFT 4.3p
- Mutational Signature: SomaticSignatures 2.10.00
- Alignment Statistics: FASTQC (0.11.5) / Qualimap (2.2)

# Annexes

Pour tout comprendre



## Annexe 2 : Références

### Cancer de l'estomac (étude génétique TCGA) :

Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., ... Shen, H. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, 513(7517), 202–209. doi:10.1038/nature13480

### Mutation MET :

Paik PK, Drilon A, Fan PD, et al. Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. *Cancer Discov.* 2015;5(8):842-9.

Awad, M. M., Oxnard, G. R., Jackman, D. M., Savukoski, D. O., Hall, D., Shivdasani, P., ... Sholl, L. M. (2016). MET Exon 14 Mutations in Non-Small-Cell Lung Cancer Are Associated With Advanced Age and Stage-Dependent MET Genomic Amplification and c-Met Overexpression. *Journal of Clinical Oncology*, 34(7), 721–730. doi:10.1200/jco.2015.63.4600

Frampton, G. M., Ali, S. M., Rosenzweig, M., Chmielecki, J., Lu, X., Bauer, T. M., ... Miller, V. A. (2015). Activation of MET via Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical Sensitivity to MET Inhibitors. *Cancer Discovery*, 5(8), 850–859. doi: 10.1158/2159-8290.cd-15-0285

Liu, X., Jia, Y., Stoopler, M. B., Shen, Y., Cheng, H., Chen, J., ... Borczuk, A. C. (2016). Next-Generation Sequencing of Pulmonary Sarcomatoid Carcinoma Reveals High Frequency of Actionable MET Gene Mutations. *Journal of Clinical Oncology*, 34(8), 794–802. doi:10.1200/jco.2015.62.0674

### Mutation PIK3CA :

Bosch, A., Li, Z., Bergamaschi, A., Ellis, H., Toska, E., Prat, A., ... Baselga, J. (2015). PI3K inhibition results in enhanced estrogen receptor function and dependence in hormone receptor-positive breast cancer. *Science Translational Medicine*, 7(283), 283ra51–283ra51. doi:10.1126/scitranslmed.aaa4442

Juric D, Rodon J, Tabernero J, et al. Phosphatidylinositol 3-Kinase  $\alpha$ -Selective Inhibition With Alpelisib (BYL719) in PIK3CA-Altered Solid Tumors: Results From the First-in-Human Study. *J Clin Oncol.* 2018;36(13):1291-1299.

Juric, D., de Bono, J. S., LoRusso, P. M., Nemunaitis, J., Heath, E. I., Kwak, E. L., ... Tabernero, J. (2017). A First-in-Human, Phase I, Dose-Escalation Study of TAK-117, a Selective PI3K $\alpha$  Isoform Inhibitor, in Patients with Advanced Solid Malignancies. *Clinical Cancer Research*, 23(17), 5015–5023. doi: 10.1158/1078-0432.ccr-16-2888

Dickler, M. N., Saura, C., Richards, D. A., Krop, I. E., Cervantes, A., Bedard, P. L., ... Baselga, J. (2018). Phase II Study of Taselisib (GDC-0032) in Combination with Fulvestrant in Patients with HER2-Negative, Hormone Receptor-Positive Advanced Breast Cancer. *Clinical Cancer Research*. doi: 10.1158/1078-0432.ccr-18-0613



# Annexes

Pour tout comprendre



## Suite

Juric, D., Janku, F., Rodón, J., Burris, H. A., Mayer, I. A., Schuler, M., ... Rugo, H. S. (2018). Alpelisib Plus Fulvestrant in PIK3CA-Altered and PIK3CA-Wild-Type Estrogen Receptor-Positive Advanced Breast Cancer. *JAMA Oncology*, e184475. doi:10.1001/jamaoncol.2018.4475

Rodon, J., Braña, I., Siu, L. L., De Jonge, M. J., Homji, N., Mills, D., ... Bendell, J. C. (2014). Phase I dose-escalation and -expansion study of buparlisib (BKM120), an oral pan-Class I PI3K inhibitor, in patients with advanced solid tumors. *Investigational New Drugs*, 32(4), 670–681. doi:10.1007/s10637-014-0082-9

Fritsch, C., Huang, A., Chatenay-Rivauday, C., Schnell, C., Reddy, A., Liu, M., ... Sellers, W. R. (2014). Characterization of the Novel and Specific PI3K Inhibitor NVP-BYL719 and Development of the Patient Stratification Strategy for Clinical Trials. *Molecular Cancer Therapeutics*, 13(5), 1117–1129. doi:10.1158/1535-7163.mct-13-0865

Ndubaku, C. O., Heffron, T. P., Staben, S. T., Baumgardner, M., Blaquiere, N., Bradley, E., ... Olivero, A. G. (2013). Discovery of 2-{3-[2-(1-Isopropyl-3-methyl-1H-1,2,4-triazol-5-yl)-5,6-dihydrobenzo[f]imidazo[1,2-d][1,4]oxazepin-9-yl]-1H-pyrazol-1-yl}-2-methylpropanamide (GDC-0032): A  $\beta$ -Sparing Phosphoinositide 3-Kinase Inhibitor with High Unbound Exposure and Robust in Vivo Antitumor Activity. *Journal of Medicinal Chemistry*, 56(11), 4597–4610. doi:10.1021/jm4003632

Maira, S.-M., Pecchi, S., Huang, A., Burger, M., Knapp, M., Sterker, D., ... Voliva, C. F. (2011). Identification and Characterization of NVP-BKM120, an Orally Available Pan-Class I PI3-Kinase Inhibitor. *Molecular Cancer Therapeutics*, 11(2), 317–328. doi:10.1158/1535-7163.mct-11-0474

Patnaik, A., Appleman, L. J., Tolcher, A. W., Papadopoulos, K. P., Beeram, M., Rasco, D. W., ... Ramanathan, R. K. (2016). First-in-human phase I study of copanlisib (BAY 80-6946), an intravenous pan-class I phosphatidylinositol 3-kinase inhibitor, in patients with advanced solid tumors and non-Hodgkin's lymphomas. *Annals of Oncology*, 27(10), 1928–1940. doi:10.1093/annonc/mdw282

Lopez, S., Schwab, C. L., Cocco, E., Bellone, S., Bonazzoli, E., English, D. P., ... Santin, A. D. (2014). Taselisib, a selective inhibitor of PIK3CA, is highly effective on PIK3CA-mutated and HER2/neu amplified uterine serous carcinoma in vitro and in vivo. *Gynecologic Oncology*, 135(2), 312–317. doi:10.1016/j.ygyno.2014.08.024

## Divers

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, et al. *Nature*. 2013 Aug 22;500(7463):415-21. doi:10.1038/nature12477. Epub 2013 Aug 14. Erratum in: *Nature*. 2013 Oct 10;502(7470):258. Imielinski, Marcin