



UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTE

GMGF - Génétique Médicale et Génomique Fonctionnelle UMR_S 910

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Pathologie Humaine

Spécialité : Génétique Humaine

Caroline LACOSTE DEIXONNE

**APPORT du SEQUENÇAGE HAUT DEBIT
dans l'AMELIORATION de la PRISE en CHARGE des
MALADIES MONOGENIQUES**

Soutenue le 12 décembre 2016 devant le jury :

Pr Damien SANLAVILLE- Rapporteur

Pr Loïc GARÇON - Rapporteur

Dr Martin KRAHN - Examineur

Pr Catherine BADENS - Examinatrice - Directrice de thèse

RESUME

La diffusion du séquençage haut débit (ou NGS pour *Next Generation Sequencing*) représente un tel changement d'échelle par rapport aux méthodes classiques de séquençage que les indications et l'organisation du diagnostic moléculaire s'en trouvent profondément modifiées. Le NGS permet à la fois de raccourcir le temps d'analyse et de rendu de résultat et d'élargir considérablement le nombre de gènes testés. Il promet donc d'augmenter la proportion de diagnostics posés et de faciliter l'identification de nouveaux variants et de nouveaux gènes impliqués en pathologie. Cependant dans tous les cas, il génère une quantité de données importante, données qui doivent être analysées et interprétées à l'aide d'outils bioinformatiques spécifiques.

Dans la première partie de ce travail, les stratégies existantes ainsi que les difficultés et les enjeux du séquençage haut débit pour le diagnostic moléculaire des maladies génétiques sont discutés. Dans la deuxième partie, la mise en place et la validation technique de cette approche diagnostique sont décrites au sein du laboratoire de Génétique Moléculaire de la Timone à Marseille et illustrées par trois exemples concrets de diagnostics moléculaires posés grâce à la technique de séquençage à haut débit.

Dans le domaine spécifique des maladies rares, ces nouvelles technologies sont porteuses d'un réel espoir pour les patients atteints de maladie génétique, permettant d'améliorer globalement leur prise en charge et d'accélérer les progrès dans le domaine de la recherche.

ABSTRACT

The diffusion of Next Generation Sequencing (NGS) technologies induces an important change that modifies molecular diagnostics indications and prompts laboratories to re-think their diagnostic strategies, up-to-now based on Sanger sequencing routine. Several high throughput approaches are available from the sequencing of a gene panel, to a whole exome, or even a whole genome. In all cases, a tremendous amount of data are generated, that have to be filtered, interpreted and analyzed by the use of powerful bioinformatics tools.

In part 1, existing strategies and the difficulties and challenges of high-throughput sequencing for molecular diagnosis in genetic diseases are discussed. In part 2, the set up and the technical validation of this diagnostic approach in the Molecular Genetics' Laboratory of the Timone Hospital in Marseille is presented and illustrated by 3 examples of complex diagnostics solved thanks to NGS.

NGS promises to shorten significantly the time of analysis and results reporting, and to expand the number of tested genes. It also promises to increase the proportion of positive diagnoses. Finally, the NGS can identify new variants and new genes involved in human pathology, thus will globally improve patient clinical care.

REMERCIEMENTS

Je remercie ma directrice de thèse, Catherine Badens. En 2004, tu as su me faire confiance pour intégrer le 8^{ème} étage du Département de Génétique Médicale et Biologie Cellulaire suite à mon passage furtif au siècle dernier au sein de l'ex U491. Comme quoi une proposition de poste tient parfois aux souvenirs de gâteaux maison laissés en salle de pause pour accompagner un café.

Tu as bien voulu que j'entame ce travail en 2012 avant même que la plateforme ne soit livrée et je suis heureuse d'avoir pu mener ce projet à bout. Merci pour ton soutien, tes encouragements, ta disponibilité.

Je remercie Martin Krahn de participer à ce jury. Tu m'as épaulée et soutenue durant toutes ces années et ce malgré ton emploi du temps de ministre et jusqu'au fin fond des plaines enneigées de St Aubin. Merci pour ton aide.

Je remercie aussi les professeurs Damien Sanlaville et Loïc Garçon d'avoir accepté d'être rapporteurs et de siéger dans ce jury.

Merci aussi à Nicolas Levy d'avoir accepté de me lancer dans une thèse de sciences pour redevenir une étudiante à plus de 30 ans.

Merci à Laurent Villard qui a su me donner une chance d'approcher la recherche au siècle dernier pour couler des gels d'acrylamide. Ces quelques semaines de juillet 1999 m'ont donné l'envie de continuer sur cette voie professionnelle. Depuis 2011, nous collaborons de nouveau et j'apprécie toujours de travailler avec toi. Merci pour ta confiance. Comme quoi un numéro de téléphone sur un tableau pour chercher un stage regarde où cela me mène 15 ans plus tard. Et merci aux membres de ton équipe, surtout Pierre, pour sa promptitude à m'aliquoter les amorces pour les confirmations en Sanger.

Merci aussi à mes compagnons de paille du laboratoire de Génétique Moléculaire du 8^{ème}. Depuis mon arrivée il y a 12 ans, nous suivons l'évolution des diagnostics en biologie moléculaire, aux rythmes des nouvelles technologies. Ces technologies qui grignotent notre espace à vivre sans que les murs ne s'écartent. Nous avons même perdu la vue sur la « Bonne Mère ». Il paraît que bientôt cela va s'arranger. Mais nous pouvons toujours compter sur notre bonne humeur autour d'un café et les petites douceurs dans la salle de pause. Merci à vous tous.

Une pensée particulière pour mes binômes successifs. Avec Katia, il a fallu mettre en place la plateforme Sanger avec l'ancêtre ABI 310. Deux petites nouvelles à l'époque, cela n'a pas été toujours facile mais nous avons pris nos marques. Il y a eu aussi Amandine, avec qui j'ai alterné mes grossesses et à qui j'ai dû laisser le Sanger (et ce ne fut pas facile de laisser ce bébé là). Avec Christophe maintenant, nous avons entamé l'histoire du NGS en diagnostic au laboratoire. Ce qui nous a permis de partager beaucoup de galère mais aussi de bons moments (et même des chambres d'hôtel). Merci pour ton aide, ton soutien et tes remarques durant tout ce travail qui est aussi le tien.

Toutes ces analyses ne pourraient se faire sans un CRB au top. Merci Karine et Cécile de répondre rapidement quand un ADN me manque.

Un merci appuyé à Patou, qui sait veiller sur nous comme une poule sur ses petits. Tu es essentielle au laboratoire et nous ne te le disons pas assez souvent. Alors un grand merci à toi. Une pensée aussi à Brigitte T qui est souvent une des premières personnes que je vois le matin. Vive nos bonjours matinaux.

Merci à Françoise et Sylvie, nos deux secrétaires du laboratoire pour leur aide et leur présence.

Un merci sincère à Patrice, pour ton aide précieuse lors de notre travail commun de rédaction *aux Archives de Pédiatrie*. Merci pour les corrections, les remarques, ta disponibilité et les relectures de ce manuscrit. Et surtout pour ton implication quand il faut se battre pour se faire entendre dans les hautes sphères de l'APHM.

Un merci particulier à Jean-Pierre pour sa patience sur des aspects informatiques du NGS que je maîtrise moins bien. C'est bien beau d'avoir des variants chez les patients étudiés mais il faut bien savoir les trier. Un merci sincère.

Merci à mes cadres de santé Nicole et Christine, que je n'arrive toujours pas à tutoyer. Un merci chaleureux à Christine, qui même en vacances, pense à moi pour m'améliorer à l'oral.

Merci aux biologistes, Rafaëlle, Annachiara, Perrine, Nathalie, Christophe et Mathieu pour leurs encouragements et leurs remarques constructives.

Merci aussi aux internes anciens et nouveaux, devenus assistants ou titulaires qui sont encore dans nos murs ou qui sont partis pour d'autres aventures, particulièrement Pauline avec qui j'ai partagé un peu de mon expérience de paillasse au début du NGS au laboratoire. Je te souhaite une belle réussite parisienne.

Merci aussi à toi Alexandre, de continuer à me solliciter depuis 2006. C'est un peu grâce à toi (ou à cause de toi) que je me suis lancée dans cette aventure une fois ta thèse soutenue. Pas mal d'hésitations mais merci de m'avoir encouragée jusqu'au bout. Tu sais qu'un bout de ma paillasse est réservé pour tes projets futurs.

Merci aux autres membres des secteurs du département, ceux du 7^{ème}, du 8^{ème} et 3^{ème} avec qui j'ai pu collaborer sur certains projets.

Merci aux membres de l'unité UMR_S910. Une pensée pour ceux que je côtoie au fil de mes allers-retours Hôpital-Faculté. Un merci sincère et chaleureux à Françoise pour ton aide quand il faut faire le lien entre les deux institutions. Et à Clotilde pour ton aide dans la partie administrative lors du dépôt de cette soutenance. Dur dur dur d'avoir 40 ans.

Une pensée particulière à ceux que je ne vois plus mais qui ont traversé ma vie professionnelle.

A Montpellier en 2000, à l'ex U_439, j'y ai fait mes premières armes à la paillasse après y avoir été stagiaire. Merci à Patrick Balaguer puis Eric Badia et leurs équipes pour m'avoir proposé mes premiers postes de vacataire. Ces deux ans sont gravés à jamais dans mon cœur et c'est avec beaucoup de nostalgie que je repense à tous les bons moments passés. Tu avais raison Anne-Marie B, je n'ai jamais retrouvé l'ambiance qui régnait à l'époque au labo. Un peu de toi est dans chacun de mes gestes à la paillasse.

A Marseille Nord, à l'UMR_641, je me suis initiée concrètement à la génétique. Merci Madeleine Garcia-Erard de m'avoir fait confiance pendant plus d'un an. Ton équipe de recherche (c'est à dire toi et moi) veillait sur les *C.elegans* qui de mémoire étaient épileptiques, en tout cas, ils convulsaient autant qu'un nématode peut convulser. Une belle expérience et de belles rencontres aussi dans cette unité.

Merci à vous tous.

Ce travail n'aurait pas pu se faire sans le soutien de mes proches, familles et amis.

A mon chéri de 20 ans, à mon mari de plus de 10 ans.

Au père de mes trois enfants : Boris je t'aime.

A Gabriel, Gaspard et Garance, mes 3G qui illuminent ma vie. Vous poussez trop vite.

A mes parents que j'aime. Je sais que ce n'est pas facile de répondre quand on vous pose la question « elle travaille dans quoi votre fille ? ». Je vous rassure, je n'arrive pas toujours moi-même à l'expliquer. Merci d'être toujours là comme parents et grands-parents actifs. Merci à mon frère, à ma belle maman et tous les membres de ma famille au sens large. Une pensée particulière pour mes grands-parents qui ont suivi toutes les étapes de ma vie. Une pensée aussi à tous ceux qui ne sont plus là aujourd'hui. Je pense à vous.

Merci à mes amis d'hier et d'aujourd'hui.

Il n'y a pas de réussite facile ni d'échec définitif.

Marcel Proust

TABLE DES MATIERES

RESUME	1
ABSTRACT	2
REMERCIEMENTS	3
TABLE DES FIGURES	10
TABLE DES TABLEAUX	11
AVANT PROPOS	12
1. PARTIE 1	13
1.1. INTRODUCTION	13
1.2. MALADIES RARES ET VARIATION GENETIQUE	15
1.3. CHANGEMENT D'ECHELLE POUR LE DIAGNOSTIC MOLECULAIRE	17
1.4. SEQUENÇAGE A HAUT DEBIT EN PRATIQUE	19
1.4.1. CONSTRUCTION DES LIBRAIRIES	21
1.4.2. SEQUENÇAGE NGS PROPREMENT DIT	23
1.4.3. PROFONDEUR ET COUVERTURE, LES DEUX PARAMETRES CLE DU NGS	27
1.4.4. TRAITEMENT INFORMATIQUE DES DONNEES BRUTES	29
1.4.5. INTERPRETATION BIOLOGIQUE DES DONNEES	29
1.4.6. VERS UNE NECESSAIRE HARMONISATION DES PRATIQUES	30
1.5. GENOME, EXOME, PANEL DE GENES OU GENE UNIQUE, QUELLE STRATEGIE CHOISIR ?	31
1.5.1. PANEL DE GENES D'INTERET	33
1.5.2. SEQUENÇAGE D'EXOME	33
1.5.3. SEQUENÇAGE DE LARGE PANEL (MUTOME) ET FILTRAGE BIO-INFORMATIQUE	34
1.5.4. SEQUENÇAGE DU GENOME ENTIER	35
1.5.5. QUEL CHOIX POUR QUELS RESULTATS ?	35
1.6. INTERPRETATION DES VARIANTS ET IMPORTANCE DE LA DESCRIPTION CLINIQUE	36
1.7. IMPLICATION ETHIQUE ET SOCIETALE DU NGS	37
1.8. PRESENTATION DU TRAVAIL DE THESE	39
1. PARTIE 2	40
1.1. INTRODUCTION	40
1.2. MATERIELS ET METHODES	43
1.2.1. PATIENTS ET ADN GENOMIQUE	43
1.2.2. KITS D'ENRICHISSEMENT – PREPARATION DES LIBRAIRIES	43
1.2.3. SEQUENÇAGE PLATEFORME ION PROTON	44
1.2.4. ANALYSE DES DONNEES BRUTES ISSUES DU NGS	46
1.2.5. ANALYSE ET FILTRE DES VARIANTS	46
1.2.6. CALCULS DES DONNEES DE COUVERTURE	47
1.2.7. CONFIRMATION DES VARIANTS SELECTIONNES	47
1.3. ANALYSE DES PERFORMANCES DE LA PLATEFORME ION PROTON™	48
1.3.1. VALIDATION DE LA PLATEFORME ION PROTON™	48
1.3.2. VALIDATION DU PROCESSUS D'ENRICHISSEMENT AMPLISEQ EXOME™	53
ARTICLE 1	55

1.3.3. DONNEES COMPLEMENTAIRES	56
1.4. PREMIERS RESULTATS DE LA MISE EN PLACE DU DIAGNOSTIC NGS AU LABORATOIRE	70
1.5. EXEMPLES DE DIAGNOSTICS POSES GRACE AU NGS	74
1.5.1. AMELIORATION DE LA DESCRIPTION DU PHENOTYPE D'ENCEPHALOPATHIE EPILEPTIQUE PRECOCE	74
ARTICLE 2	75
<i>IN UTERO</i> SEIZURES REVEALING A DENTATO-OLIVARY DYSPLASIA CAUSED BY <i>SCN2A</i> MUTATION	75
1.5.2. RESOLUTION D'UN CAS DE DEFICIENCE INTELLECTUELLE PRESENTANT UN PHENOTYPE CONFONDANT.	76
1.5.3. CARACTERISATION D'UN NOUVEAU GENE RESPONSABLE D'UNE FORME D'ANEMIE HEMOLYTIQUE HEREDITAIRE AVEC GLOBULES ROUGES DESHYDRATES.	78
ARTICLE 3	80
2. CONCLUSIONS GENERALES ET PERSPECTIVES	81
2.1. LA PLATEFORME ET LA STRATEGIE EN DIAGNOSTIC	81
2.2. LES SEQUENCEURS 3^{EME} GENERATION	82
2.3. LE DIAGNOSTIC DE DEMAIN : VERS UNE MEDECINE PERSONNALISEE PAR L'ANALYSE DU GENOME ?	83
2.4. LE POSITIONNEMENT DE LA FRANCE AU NIVEAU INTERNATIONAL	84
2.5. DIFFERENCIER LE DIAGNOSTIC ET LA RECHERCHE	85
BIBLIOGRAPHIE	87
RAPPEL DES LIENS	92
ABBREVIATIONS	93
COMMUNICATIONS ISSUES DE CE TRAVAIL DE THESE	95
AUTRES PUBLICATIONS	96

TABLE DES FIGURES

FIGURE 1 : LE GENOME HUMAIN	16
FIGURE 2 : LE SEQUENÇAGE « SANGER »	18
FIGURE 3 : LES DIFFERENTES ETAPES DU NGS	20
FIGURE 4 : REPRESENTATION SCHEMATIQUE DES DEUX GRANDS PRINCIPES D'ENRICHISSEMENT DE L'ADN GENOMIQUE DU PATIENT	22
FIGURE 5 : SCHEMA DU PRINCIPE DE SEQUENÇAGE SUR UNE PLATEFORME MISEQ (ILLUMINA)	24
FIGURE 6 : SCHEMA DU PRINCIPE DU SEQUENÇAGE SUR UNE PLATEFORME ION PROTON™ (THERMO FISCHER)	26
FIGURE 7 : NOTION DE COUVERTURE ET DE PROFONDEUR DE SEQUENÇAGE	28
FIGURE 8 : LES DIFFERENTS NIVEAUX D'APPROCHE DU SEQUENÇAGE	32
FIGURES 9 : SCHEMA DE LA STRATEGIE DE DIAGNOSTIC PAR NGS AU SEIN DU LABORATOIRE DE GENETIQUE MOLECULAIRE – TIMONE ENFANT	42
FIGURE 10 : REPRESENTATION GRAPHIQUE DU CHARGEMENT BRUT DES PUCES EN GB D'ADN EN FONCTION DE LA STRATEGIE UTILISEE	59
FIGURE 11: PROCESSUS ANALYTIQUE DEPUIS MAI 2016	69
FIGURE 12 : RESUME DES TROIS EXEMPLES DE DIAGNOSTICS POSES PAR NGS	73
FIGURE 13 : ARBRE GENEALOGIQUE ET DONNEES DU SEQUENÇAGE EN SANGER DE LA FAMILLE DU PATIENT ATTEINT DE DEFICIENCE INTELLECTUELLE	77

TABLE DES TABLEAUX

TABLEAU 1: RESUME DES DIFFERENTES STRATEGIES DEPUIS LA MISE EN PLACE DE LA PLATEFORME AU SEIN DU LABORATOIRE	45
TABLEAU 2: COMPARAISON DES INDICATEURS DE SEQUENCES SUR LES DONNEES D'EXOME	49
TABLEAU 3: RESUME DES DONNEES DES 10 PATIENTS UTILISEES POUR LA VALIDATION DE LA PLATEFORME	52
TABLEAU 4: COMPARAISON DES STRATEGIES 1 ET 2 SUR DES INDICATEURS DE DONNEES DE SEQUENCE	57
TABLEAU 5: COMPARAISON DES INDICATEURS DE SEQUENCE, DES VARIANTS ET DES DONNEES DE COUVERTURE ENTRE LES STRATEGIES 2 ET 2+	61
TABLEAU 6: RESUME DES DONNEES SUR 4 LISTES DE GENES CIBLEES	63
TABLEAU 7: COMPARAISON DES DONNEES DE COUVERTURE A UNE PROFONDEUR DE 20X MINIMUM SUR 4 LISTES DE GENES D'INTERÊT	65
TABLEAU 8: COMPARAISON DE LA SEMI AUTOMATISATION POUR LA SYNTHÈSE DES LIBRAIRIES AVEC LE ROBOT BRAVO™ SUR LES INDICATEURS DE SEQUENCE AVEC LA STRATEGIE 2	67
TABLEAU 9: BILAN D'ACTIVITE DU RENDU DE DIAGNOSTIC POSITIFS CONFIRMES SUR LA PLATEFORME ION PROTON™	71

AVANT PROPOS

La diffusion du séquençage haut débit ou NGS (pour *Next Generation Sequencing*) dans les laboratoires de génétique médicale représente un tel changement d'échelle par rapport aux méthodes classiques qu'il est important de repenser complètement les stratégies du diagnostic et de réorganiser le fonctionnement des laboratoires. Dès 2012, la Direction de l'Hospitalisation et l'Organisation des Soins (DHOS), a apporté un soutien financier à plusieurs laboratoires dont le nôtre, pour l'installation d'une plateforme NGS dédiée au diagnostic.

Le travail présenté ici retrace la mise en place de la plateforme NGS au laboratoire de Génétique Moléculaire à l'hôpital d'enfants de la Timone et l'évaluation de ses performances. Il rapporte également des exemples concrets de l'apport du NGS dans le domaine du diagnostic des maladies rares.

1. PARTIE 1

1.1. INTRODUCTION

La publication de la séquence du génome humain en 2001 a été célébrée comme l'une des plus grandes avancées de la science moderne. (Lander ES et al., 2001). En dépit du fait qu'elle était incomplète et comportait de nombreuses erreurs, cette séquence a marqué significativement l'avancée du Projet Génome Humain, projet considéré comme largement utopique au moment de son lancement en 1989. À cette époque, la technologie du séquençage d'ADN était encore lente et laborieuse et plusieurs jours étaient nécessaires pour produire les données relatives à quelques centaines de paires de bases seulement. Mais rapidement, les progrès technologiques ont été tels qu'ils ont permis de compléter la séquence du génome humain en moins de 15 ans.

Séquencer l'ADN est un procédé chimique qui permet d'établir l'ordre dans lequel les quatre nucléotides (A, G, C et T pour Adénosine, Guanosine, Cytidine et Thymidine) sont enchaînés dans un brin d'ADN, formant un message porteur d'information. Ce procédé est basé sur la synthèse d'un brin d'ADN complémentaire à celui dont on veut déterminer la séquence, en utilisant l'enzyme ADN polymérase, et grâce à une réserve de nucléotides libres dans le mélange réactionnel. C'est le biochimiste Frederick Sanger qui a mis cette méthode de séquençage au point en 1977 et qui depuis porte son nom, permettant le séquençage d'échantillons d'ADN d'abord de façon manuelle puis automatisée (Sanger F et al., 1977). Initialement, le débit des analyses manuelles était d'environ 5 jours pour 800 paires de bases (pb) et nécessitait l'utilisation de nucléotides radioactifs. Avec les séquenceurs automatiques capillaires, la cadence est passée à la fin du Projet Génome Humain à environ 500 kilobases (Kb) séquencées par jour, ce qui représentait déjà un progrès considérable. Puis, à partir de 2006, une convergence s'est opérée entre trois domaines jusqu'alors cloisonnés : les techniques microfluidiques (circulation de liquides dans des capillaires d'une taille de l'ordre du micromètre), les nanotechnologies (possibilité de fabriquer des objets de la taille d'un milliardième de mètre), et enfin les progrès de l'informatique, qui suivent la loi de Moore (doublement des capacités de traitement de l'information tous les 18 mois).

Cette convergence a permis l'émergence de nouveaux types de machines capables de réaliser plusieurs millions de réactions de séquence en parallèle, puis d'analyser les résultats et de traiter les informations obtenues. Ces automates sont appelés des séquenceurs à très haut débit ou des séquenceurs de nouvelle génération (Bai X et al., 2005). Avec l'émergence de ces nouvelles technologies de séquençage massif, souvent regroupées sous l'acronyme NGS pour *Next Generation Sequencing*, les capacités ont été exponentielles. Aujourd'hui, une seule machine peut séquencer 1 Térabases (Tb) en moins d'un jour, et les développements technologiques annoncés sont plus vertigineux encore (Kingsmore SF et al., 2011).

Même si nous estimons à plus de 7000 le nombre de maladies rares monogéniques, la majorité des patients avec une maladie génétique n'ont pas de diagnostic moléculaire posé. Pourtant, ils subissent une batterie d'examen souvent longs et coûteux: une étude des signes cliniques, des examens radiologiques, des biopsies, des analyses biologiques, des tests génomiques comme des caryotypes ou des Analyse Chromosomique sur Puce à ADN (ACPA), des analyses de gènes candidats en cascades, quelques fois sans résultats. Dans une majorité de cas, la démarche diagnostique n'aboutit pas et la prise en charge reste clinique et non spécifique. De même, un conseil génétique précis et éventuellement un diagnostic prénatal ne peuvent être proposés (Yang Y et al., 2013).

En revanche, l'identification de la mutation causale dans un gène est un point de départ important pour comprendre le mécanisme pathologique d'une maladie, pour proposer une thérapie spécifique et un conseil génétique éclairé (Gilissen C et al., 2012).

Une version condensée de la suite de la partie 1 a été soumise et acceptée sous le format d'une revue aux *Archives de Pédiatrie* sous le titre : « **Le séquençage haut débit en pratique diagnostique clinique** ».

1.2. MALADIES RARES ET VARIATION GENETIQUE

Environ 6% de la population mondiale est concernée par une maladie rare, maladie héréditaire la plupart du temps. En France, au moins 3 millions de français sont atteints par une de ces maladies qui seraient responsables d'un tiers de la mortalité infantile (source : Fondation Maladies Rares : <http://fondation-maladiesrares.org/>). Bien que rares au niveau individuel, les pathologies à transmission mendélienne sont impliquées dans au moins 20% de la mortalité infantile et dans environ 10% des hospitalisations pédiatriques (Bell CJ et al., 2011). Dans le cadre du diagnostic moléculaire d'une maladie d'origine génétique, l'identification du variant dans un gène est un point de départ nécessaire pour appréhender le mécanisme pathologique et espérer proposer une prise en charge adaptée (Gilissen C et al., 2012). Pour cela, il est nécessaire de déterminer la séquence du gène, c'est-à-dire l'ordre dans lequel les nucléotides qui le composent sont reliés sur un brin.

Chez l'Homme, le génome nucléaire représente 3.2 Gigabases (Gb) et l'ADN mitochondrial représente un peu plus de 16 Kb d'ADN. Le génome est constitué aux deux tiers de séquences répétées dont la fonction est encore mal connue. Environ un quart concerne les gènes (exons, introns et séquences régulatrices), dont le nombre est estimé à environ 22 000 chez l'Homme (Figure 1) (Lander ES et al., 2001).

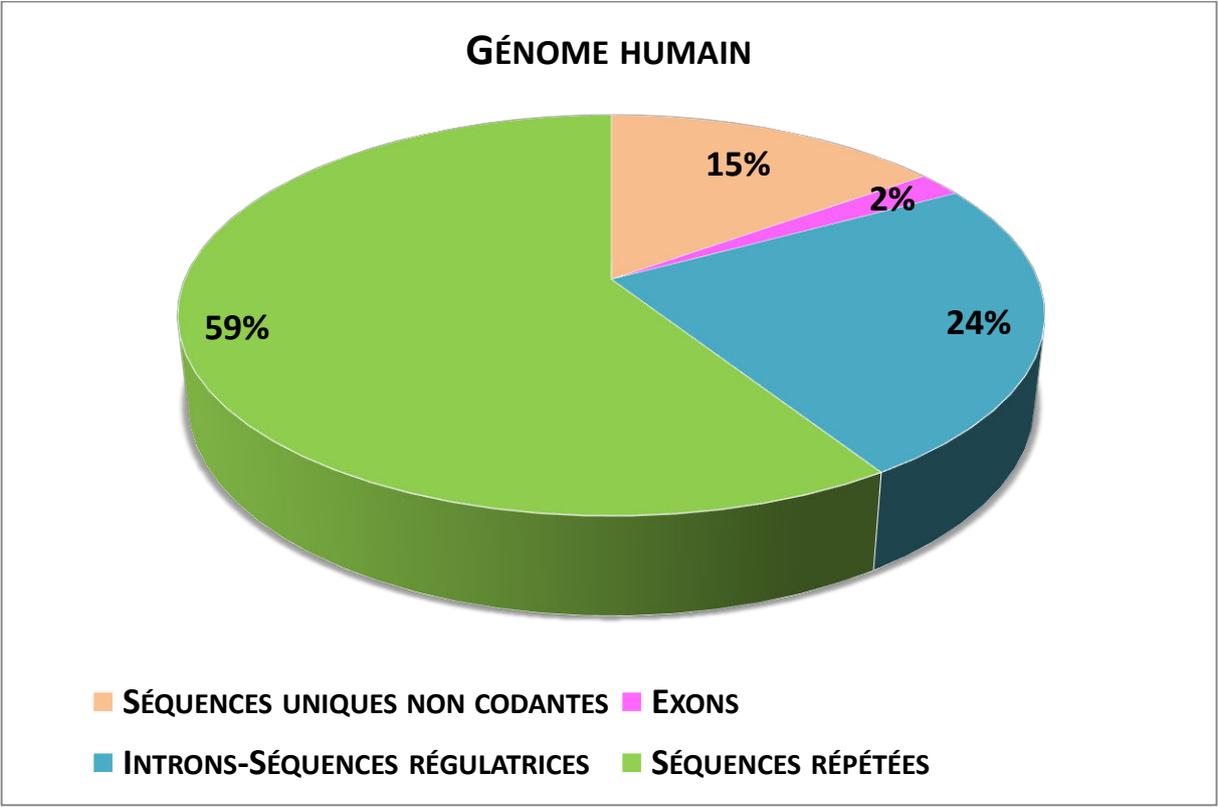


Figure 1 : Le génome humain
Répartition de l'information génétique nucléaire.

Cependant, seulement 3 000 d'entre eux environ ont été impliqués dans des pathologies mendéliennes à ce jour. Ce sous-groupe de gènes est parfois désigné par le terme de *mutome* ou de *mendélianome*. Dans les gènes eux-mêmes, seules les parties codantes ou exons contiennent un message permettant la synthèse protéique, si bien qu'au total, c'est moins de 2 % de l'ADN humain qui code pour des protéines.

Une mutation est définie, en pratique médicale, comme une variation pathogène de la séquence d'ADN d'une personne par rapport à une séquence dite « de référence », c'est-à-dire une séquence consensuelle du génome humain obtenue par compilation de la séquence d'individus réputés sains. La version actuelle du génome humain de référence est le GCRh38 (*Genome Reference Consortium, human version 38*) disponible en ligne au *National Center for Biology Information* (NCBI : <http://www.ncbi.nlm.nih.gov/>) ou hg38 (pour *human genome version 38*), disponible en ligne sur le site UCSC (*University of California Santa Cruz* : <https://genome.ucsc.edu/>).

1.3. CHANGEMENT D'ECHELLE POUR LE DIAGNOSTIC MOLECULAIRE

La méthode la plus utilisée jusqu'à présent dans les laboratoires de diagnostic moléculaire pour établir une séquence d'ADN et ainsi identifier un potentiel variant est le séquençage dit « Sanger », qui peut concerner un seul gène ou un groupe de gènes sélectionnés en fonction des signes cliniques recueillis et de l'hypothèse diagnostique faite par le prescripteur. On parle dans ce cas de stratégie « gène par gène ». Les régions exoniques de ces gènes que l'on peut nommer comme nos régions d'intérêt sont amplifiées par PCR (*Polymerase Chain Reaction*), indépendamment les unes des autres, à partir de l'ADN génomique. Puis, les fragments d'ADN obtenus sont séquencés sur un appareil d'électrophorèse capillaire. Une comparaison est alors effectuée entre les séquences générées à partir de l'ADN du patient et la séquence de référence pour mettre en évidence d'éventuels variants nucléotidiques (Figure 2). Cette approche est chronophage, couteuse et ne permet d'explorer qu'un nombre limité de gènes.

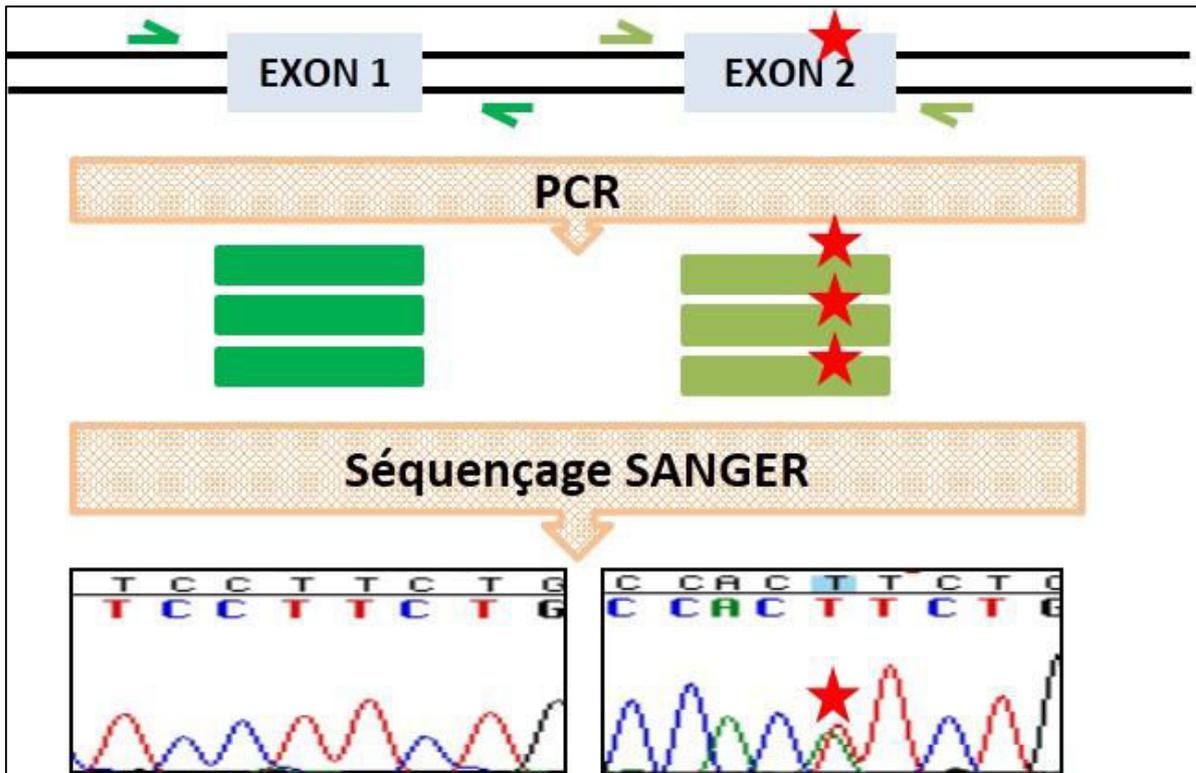


Figure 2 : Le séquençage « Sanger »

Représentation schématique de la méthode de diagnostic par PCR et séquençage capillaire dit « Sanger ». Chaque exon est amplifié par PCR (flèches vertes) puis séquencé séparément. Chaque nucléotide porte un fluorochrome de couleur différente, ce qui permet de les distinguer à la sortie du capillaire de séquençage et de visualiser leur enchainement. Les deux allèles sont séquencés simultanément, mais sur un seul brin d'ADN. Sur ce schéma, l'étoile indique un variant à l'état hétérozygote T>A dans l'exon 2, par rapport à la séquence de référence.

Avec le développement des séquenceurs dits de nouvelle génération (NGS), les capacités d'analyses des gènes ont été multipliées. En effet, ces machines sont capables de séquencer simultanément et indépendamment entre 80 Mégabases (Mb) et 100 Gigabases (Gb) d'ADN, ce qui représente 10 000 à 100 000 fois plus que les capacités des séquenceurs capillaires les plus performants, le génome humain étant de « seulement » 3.2 Gb.

Deux nouvelles stratégies ont ainsi vu le jour : le séquençage de génome qui consiste à lire l'ensemble de l'information génétique (pour l'homme, 3.2 Gb) et le séquençage d'exome qui se focalise sur les exons et concerne donc moins de 2% du génome, environ 62 Mb. En pratique, seul le séquençage d'exome est actuellement accessible en laboratoire de diagnostic, le séquençage de génome entier étant encore trop coûteux en termes d'équipement et de matériel informatique pour l'analyse et le stockage des données.

1.4. SEQUENÇAGE A HAUT DEBIT EN PRATIQUE

Concrètement, le séquençage haut débit ou séquençage NGS est basé sur la génération et la lecture en parallèle de millions de petites séquences d'ADN indépendantes. Ces données brutes sont ensuite réorganisées par un traitement informatique pour reconstituer le brin d'ADN cible, avant d'être analysées pour y déceler d'éventuelles variations par rapport à la séquence de référence. De nombreux articles détaillent les principes du séquençage haut débit (Mardis ER, 2008) ; (Teer JK et al., 2010) ; (Koboldt DC et al., 2013) ; (Altmüller J et al., 2014) ; nous en donnons ici une vision simplifiée et globale (Figure 3).

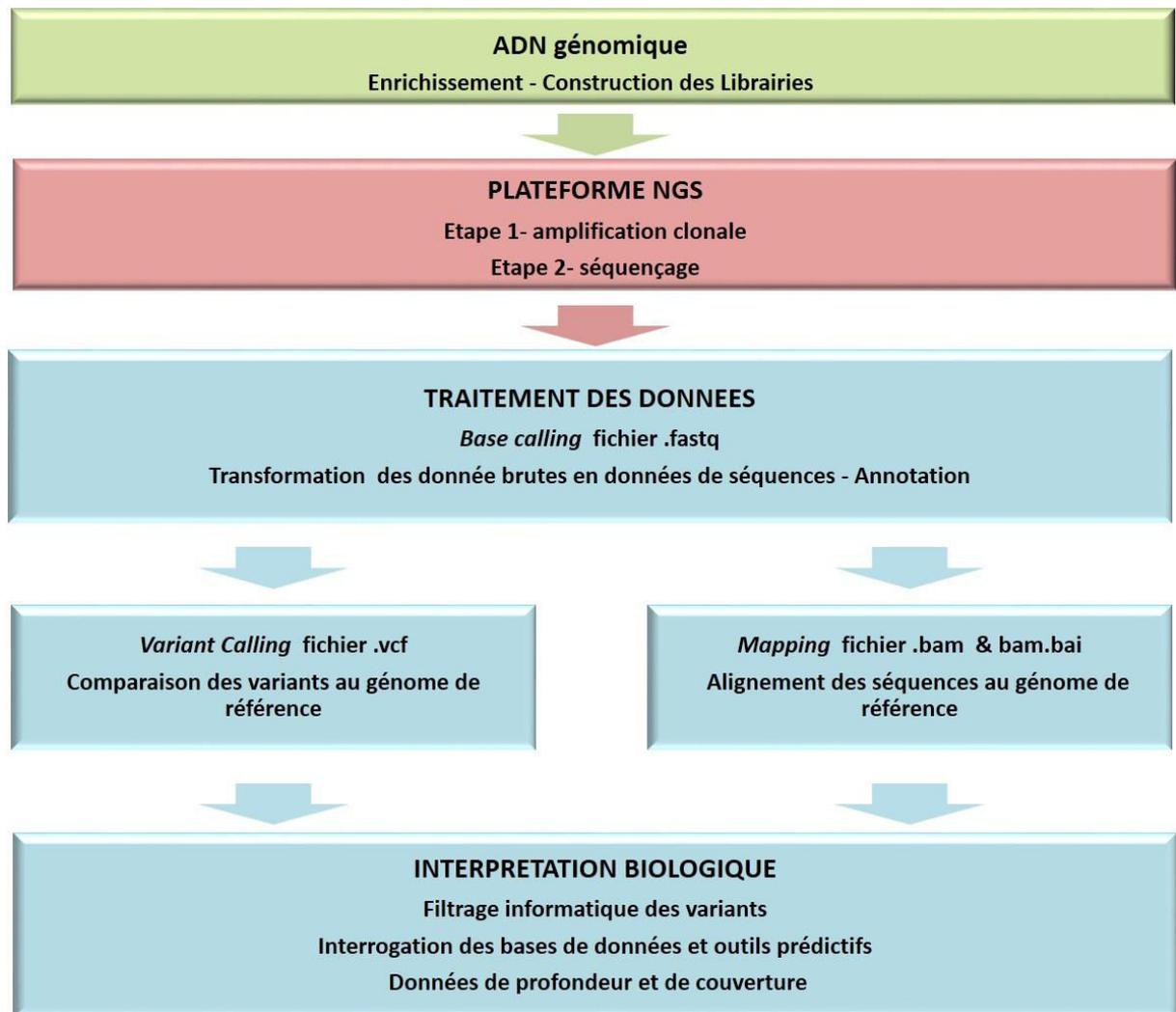


Figure 3 : Les différentes étapes du NGS

L'enrichissement des régions d'intérêt est réalisé en deux jours en moyenne. L'étape de séquençage proprement dit se déroule sur environ une journée. Les données brutes sont transformées en données de séquence pendant l'étape de *base calling* (quelques heures). Ces données de séquence sont ensuite alignées et annotées sur une séquence de référence du génome humain pour obtenir deux types de fichier. Avec les données de *mapping*, les fichiers .BAM permettent de calculer les données de profondeur et de couverture des régions d'intérêt. Lors du *variant calling*, le fichier .VCF met en évidence tous les variants du patient par rapport au génome de référence. L'étape la plus variable et délicate est l'interprétation biologique.

1.4.1. Construction des librairies

La première étape consiste à produire des millions de petits fragments d'ADN qui sont des copies des régions d'intérêt (régions du génome d'un patient que l'on souhaite analyser). Cette collection de fragments amplifiés est appelée « librairie » (anglicisme pour *library*, banque de fragments d'ADN ciblés). Plusieurs kits commerciaux existent, utilisant deux techniques différentes d'enrichissement de l'ADN génomique. La première est basée sur le principe de la **PCR-multiplexe** c'est-à-dire l'amplification de plusieurs fragments dans le même tube (Figure 4). L'autre technique d'enrichissement est basée sur l'hybridation, ou **capture d'ADN**. On génère alors des fragments de 200 à 250 paires de bases d'ADN génomique par digestion enzymatique ou par sonication (traitement aux ultrasons). Ces fragments sont hybridés avec des sondes spécifiques des régions d'intérêt: c'est la capture (Figure 4).

A la fin de cette étape, on dispose d'une librairie constituée des régions d'ADN enrichies que l'on souhaite étudier. Au cours de cette étape, chaque fragment de la librairie est associé à une petite séquence d'ADN synthétique, équivalent à un code barre, identifiant l'échantillon d'origine et permettant ainsi par la suite de séquencer plusieurs patients en même temps, on parle alors de multiplexage.

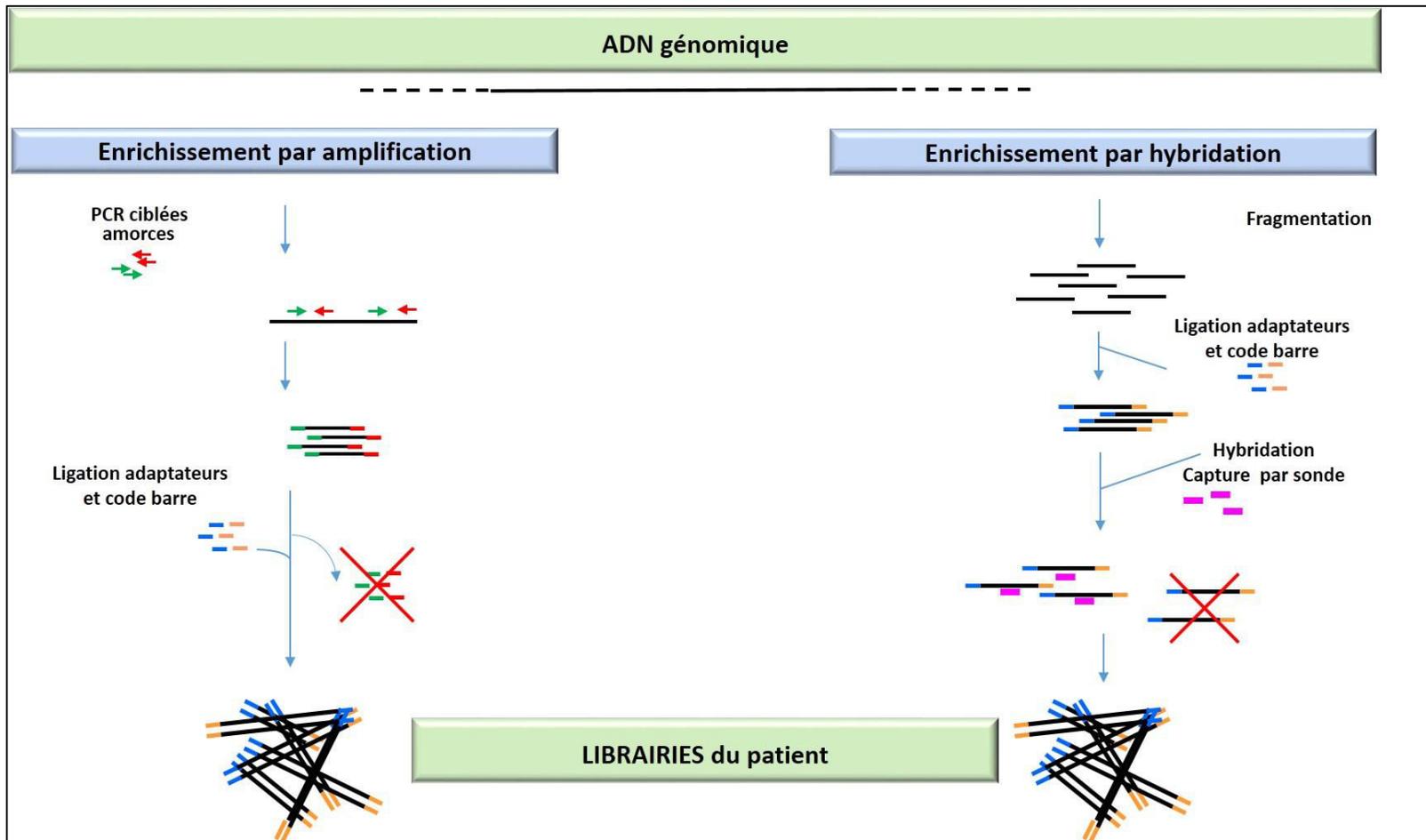


Figure 4 : Représentation schématique des deux grands principes d'enrichissement de l'ADN génomique du patient

Lors d'un enrichissement par amplification, étape d'environ une journée, on cible les régions d'intérêt par l'utilisation en multiplexage d'amorces spécifiques, représentées par les flèches rouges et vertes. On ajoute ensuite aux produits de PCR obtenus des adaptateurs spécifiques à la plateforme (en bleu et orange), ainsi qu'un code barre permettant d'identifier le patient. Avec la méthode par hybridation, l'ADN génomique est coupé en fragments de 200 bases puis associé aux adaptateurs compatibles avec la plateforme de NGS utilisée (en bleu et orange) et au code barre permettant d'indexer chaque échantillon lors du séquençage. Des sondes spécifiques des régions cibles (boîtes roses) sont ajoutées pour capturer les fragments ciblés, avant une courte étape finale d'amplification par PCR (7 à 10 cycles). Les régions non capturées sont éliminées. Suivant les fournisseurs, cette étape d'enrichissement dure entre 1 et 3 jours.

1.4.2. Séquençage NGS proprement dit

Pour séquencer ces bibliothèques sur la plateforme, il est ensuite nécessaire de réaliser une amplification clonale. Cette étape peut se faire par amplification sur lame de verre (*flow cells*) : à partir d'une molécule d'ADN de la bibliothèque fixée sur la *flow cell*. L'amplification clonale permet de reproduire des milliers de copies de ce fragment au voisinage immédiat de son point d'attachement (Figure 5). Puis, ces fragments clonés sont directement séquencés sur la *flow cell*. Un autre type d'amplification clonale consiste à attacher la molécule d'ADN à une bille microscopique (ISP pour *Ion Sphere Particle*) et de l'amplifier sur cette bille par PCR en émulsion (Figure 6). Ces ISP amplifiées sont ensuite chargées sur une puce ou *chip* pour que les fragments d'ADN clonés soient directement séquencés.

Même si les plateformes utilisent des techniques différentes, toutes sont basées sur la capacité de déchiffrer *in situ* la séquence en bases de chaque fragment pour en acquérir les données en temps réel, ceci grâce à l'activité d'une ADN polymérase incorporant des nucléotides complémentaires du brin d'ADN à séquencer (Mardis ER, 2008).

La plateforme **454** (Life Science –Roche) utilise une technique de pyroséquençage, où le phosphate libéré lors de l'incorporation du nucléotide est utilisé dans une série de réactions enzymatiques pour produire de la lumière qui est mesurée. Cette technique, assez robuste, est malheureusement trop onéreuse pour s'imposer comme un standard. La plateforme **SOLiD** (*Séquence par OligoLigation et Détection* - Applied Biosystems) utilise un procédé de séquençage par ligation complexe qui est très robuste, mais lent et cher en raison de nombreuses contraintes techniques. Ces deux plateformes sont déjà supplantées par deux autres plateformes qui s'imposent actuellement au niveau international.

Les plateformes proposées par la société **Solexa-Illumina**, utilisent une technologie basée sur la fluorescence. Pour chacun des 4 nucléotides, on associe un fluorochrome qui est excité lors de son incorporation par l'ADN polymérase. La mesure de la fluorescence est ensuite convertie en données de séquence. C'est un procédé également appelé Séquençage par Synthèse, *SBS* en anglais. Les *reads* obtenus sont très courts, mais la technique est robuste et relativement bon marché (Figure 5).

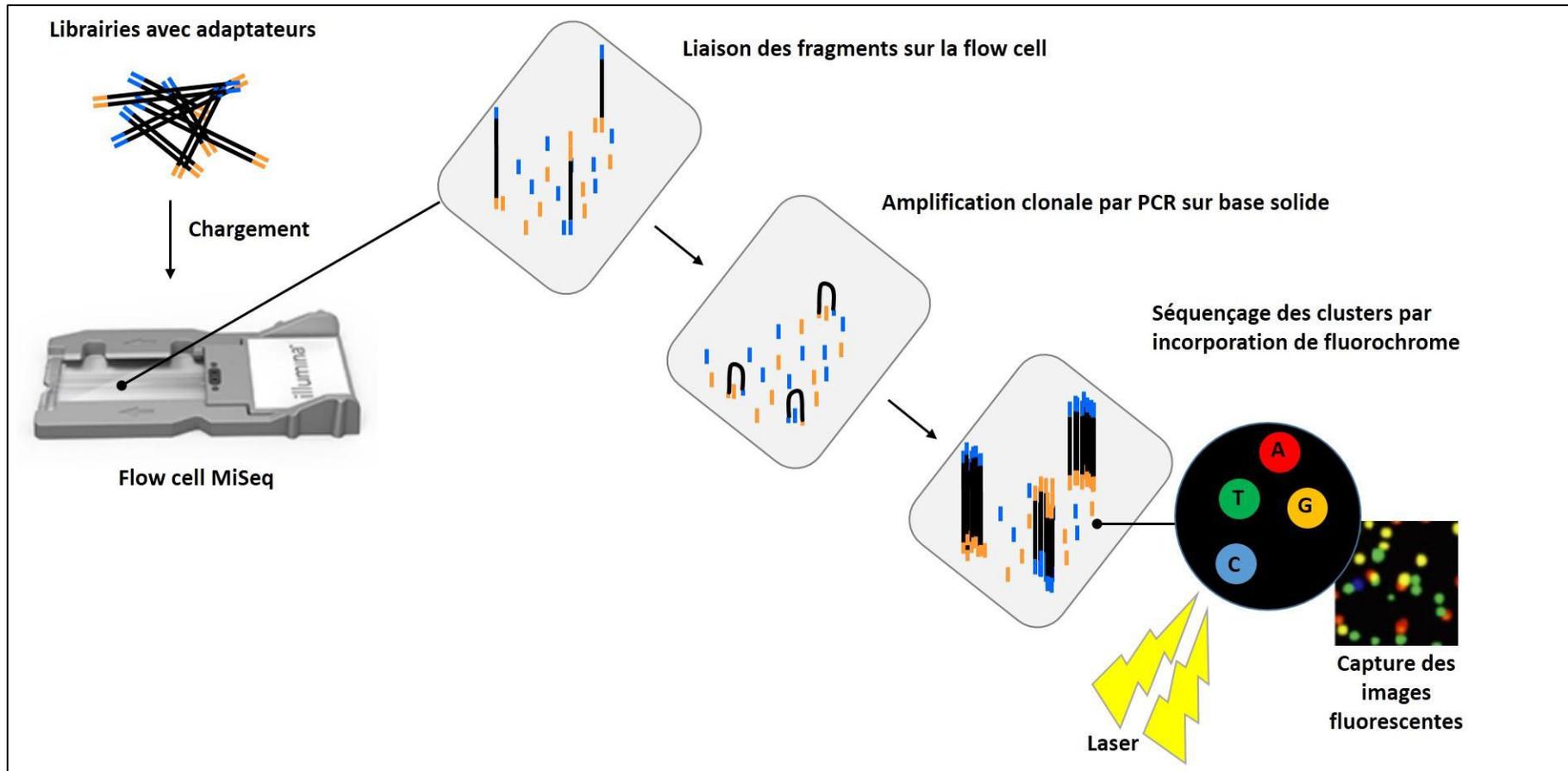


Figure 5 : Schéma du principe de séquençage sur une plateforme MiSeq (Illumina).

Les librairies produites sont chargées dans la station microfluidique (*flow cell*). Les différents fragments s'hybrident par leurs adaptateurs spécifiques sur le support pour ensuite être amplifiés et former des clusters (ou clones d'un même fragment). Le séquençage des clusters se fait par polymérisation et incorporation de dNTPs marqués par 4 fluorochromes permettant de différencier les 4 bases. À chaque flux de dNTPs, les données de fluorescence sont recueillies pour identifier les nucléotides incorporés sur chaque cluster. Cette étape du recueil des données est assez longue, car il est nécessaire de capturer la fluorescence de la surface totale de la flow cell en caméra à haute résolution. Les données brutes sont ensuite traduites en données de séquence.

La technologie **Ion-Torrent** (Thermo Fischer Scientific) utilise des semi-conducteurs et offre un autre type de plateforme. Les quatre nucléotides sont successivement apportés ; lors de son incorporation par l'ADN polymérase, le nucléotide va modifier le pH du milieu réactionnel en libérant un proton (H^+), modification qui est détectée et quantifiée. La différence de potentiel due au H^+ libéré est traduite en données de séquence. Cette approche est très bon marché et rapide, mais moins fiable en raison de difficultés techniques (Figure 6). Chaque plateforme a donc ses propres caractéristiques avec des avantages et des inconvénients (Boland JF et al., 2013).

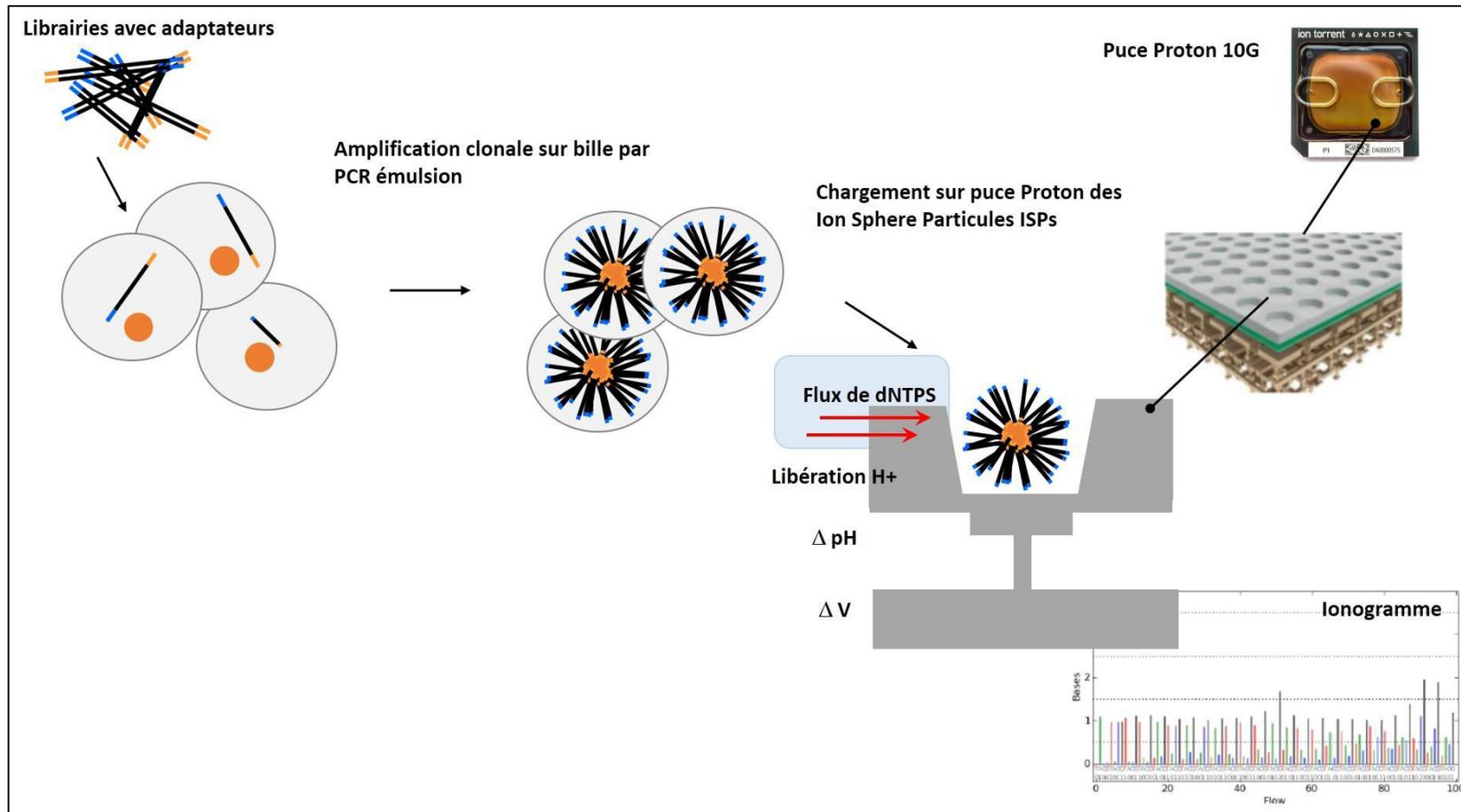


Figure 6 : Schéma du principe du séquençage sur une plateforme Ion Proton™ (Thermo Fischer).

L'amplification clonale des librairies se fait par PCR en émulsion sur des Ion Sphere Particles (ISPs) avec le système OT2/ES ou le Ion Chef™. Le ratio idéal de la PCR en émulsion est 1/1 (Fragment d'ADN de la librairie/ISP). Les ISPs sont chargées sur la puce Proton 10G, puis des flux successifs de dNTPs sont envoyés. Si le nucléotide est incorporé, il y a libération d'un H+. Cette libération entraîne une modification du pH, et donc du voltage, qui est enregistrée au niveau des senseurs de la puce et convertie en données brutes sous forme d'ionogramme, avant d'être transformée en données de séquences.

1.4.3. Profondeur et couverture, les deux paramètres clé du NGS

Deux critères fondamentaux permettent d'évaluer la qualité des données de séquençage et doivent être pris en compte pour appréhender tout résultat de séquençage à haut débit. Le premier est la **profondeur de séquençage**. C'est le nombre de lectures (ou *reads*) obtenues indépendamment pour chaque base ciblée : elle s'exprime en nombre de fois (X). Ainsi, une « profondeur de 100X » signifie que la base, ou la région considérée, a été séquencée 100 fois de manière indépendante. Très souvent, la profondeur de séquençage est une valeur moyenne sur l'ensemble des régions ciblées mais elle peut aussi être donnée en profondeur minimum. Pour pouvoir détecter des variants à l'état hétérozygote avec une qualité satisfaisante, il est recommandé de se placer à une profondeur minimum de 20X, suivant les recommandations de l'*European Society of Human Genetic* (ESHG), (Matthijs G et al., 2016).

Le second critère est la notion de **couverture**. Il correspond au pourcentage de bases effectivement séquencées par rapport au nombre total de bases ciblées au départ. La qualité globale de l'analyse sera donc appréciée par le pourcentage de couverture pour une profondeur minimale fixée ou pour une profondeur moyenne. Par exemple, si une région est couverte à 100% à 20X minimum, cela signifie que toutes les bases de la région ont été séquencées au moins 20 fois chacune de manière indépendante (Figure 7). Les performances de couverture et de profondeur sont très dépendantes des techniques d'enrichissement utilisées (Mertes F et al., 2011). Idéalement, dans un contexte diagnostique, il faudrait arriver à une profondeur minimum de 20X pour une couverture de 100%.

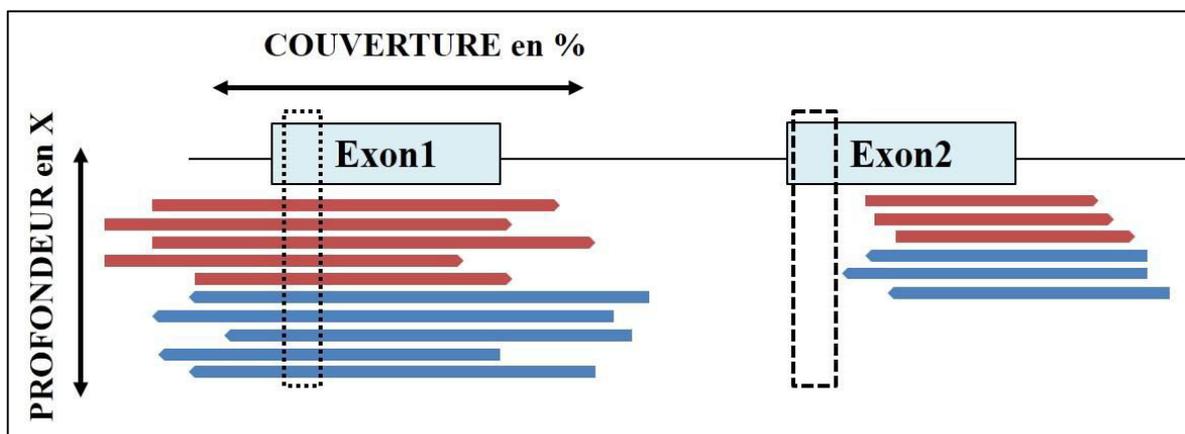


Figure 7 : Notion de couverture et de profondeur de séquençage

Chaque trait représente une lecture (ou *read* en anglais) indépendante d'une des régions ciblées. Les traits rouges sont les lectures du brin sens, les traits bleus sont des lectures du brin antisens, les deux allèles étant séquencés. La couverture (en %) représente le nombre total de bases séquencées par rapport au nombre total de bases ciblées. La profondeur de lecture (en X) est le nombre de séquences indépendantes obtenues pour une position donnée. Dans cette illustration, les régions ciblées ici sont les exons 1 et 2. L'exon 1 est couvert à 100% car la totalité de ses bases ont été lues au moins une fois. La partie encadrée en pointillés fins est couverte à 100% avec une profondeur de 10X c'est-à-dire 10 séquences indépendantes. À l'inverse, l'exon 2 n'est pas couvert à 100% puisque qu'une partie de l'exon 2 n'est pas séquencée (cadre en pointillés larges) : la fin de l'exon 2 est couverte seulement à une profondeur de 6X.

1.4.4. Traitement informatique des données brutes

Les données brutes subissent plusieurs pré-traitements suivant des algorithmes qui génèrent des données de séquence au format .FASTQ. Durant ces étapes de prétraitement, ou *base calling*, les données sont filtrées afin de ne garder que les meilleurs fragments séquencés (appelés *reads*) suivant un score de qualité (Figure 3). En cas de multiplexage, chaque *read* est réassigné suivant le code barre utilisé pour l'échantillon. Au final, des millions de *reads* pour chaque échantillon sont obtenus. Ceux-ci sont ensuite alignés informatiquement sur le génome de référence (voir plus haut), c'est ce qu'on appelle le *mapping*. Ces étapes sont en principe automatisées par le serveur de la plateforme de séquençage. Les données alignées exploitables sont au format .BAM (*Binary Alignment Map*). Ce sont des fichiers d'alignement compressés qui permettent d'accéder aux informations de couverture et de profondeur (Li H et al., 2009). Parallèlement, les coordonnées génomiques des différentes variations relevées entre les données expérimentales et la séquence de référence sont compilées et annotées dans un fichier au format .VCF (*Variant Call Format*) durant l'étape de *variant calling* (Danecek P et al., 2011).

1.4.5. Interprétation biologique des données

Dans la dernière étape, les variants sont classés en fonction de leur fréquence estimée dans la population générale et de leur pathogénicité théorique pour sélectionner celui ou ceux potentiellement responsables de la pathologie. L'interprétation nécessite une connaissance de la pathologie, du gène et des fonctions de ses produits, basée sur l'expertise humaine et sur l'utilisation d'outils de prédiction bio-informatiques et de bases de données, mais aussi des données de la littérature.

Plusieurs sites de prédictions du potentiel pathogène des mutations sont disponibles en ligne. Parmi eux, l'outil *UMD-Predictor* (<http://umd-predictor.eu/index.php>) offre une approche combinatoire aidant les biologistes à différencier les variations neutres des substitutions de nucléotides potentiellement pathogènes (Salgado D et al., 2016). Cependant, ce site n'analyse que l'effet des variations nucléotidiques entraînant une substitution d'acide aminé.

D'autres sites rapportent la fréquence des variants dans la population générale en se basant sur différentes sources: le projet *1000Genomes* rassemble les données sur le séquençage du génome de 1000 individus d'origines diverses, à priori exempts de pathologies mendélienne, formant ainsi une large base de polymorphismes (<http://www.1000genomes.org/>). Le site *ExAC* (pour *Exome Aggregation Consortium*, <http://exac.broadinstitute.org/>) rassemble des données de séquence sur plus de 60 000 individus, données issus de larges projets d'exomes indépendants et accessibles à la communauté scientifique (Lek et al., 2016). Ces individus séquencés ne présentent pas de pathologie sévère mais peuvent être atteints de maladies (asthme, autisme, schizophrénie...). Il est donc probable que cette base contient des variants pathogènes. Récemment en octobre 2016, une nouvelle base de données gnomAD (pour *Genome Aggregation Database*, <http://gnomad.broadinstitute.org>) complète le site ExAC par des données issues de 15 137 génome. Le site *Mutation Taster* (<http://www.mutationtaster.org/>) permet une évaluation rapide des variants en intégrant les données de fréquence et les données de pathogénicité (Schwarz JM et al., 2010).

Le logiciel VarAFT[®] (pour *Variant Annotation and Filtering Tool*, <http://varaft.eu/index.php>) permet d'intégrer une partie de ces analyses, notamment en interrogeant automatiquement les bases de données disponibles et les sites de prédiction. Il constitue une aide précieuse au filtrage et à l'interprétation des variants générés.

1.4.6. Vers une nécessaire harmonisation des pratiques

Actuellement, l'utilisation du NGS en diagnostic est encore en progression et les réseaux nationaux de laboratoires de génétique se structurent pour essayer d'établir une pratique harmonisée et cohérente. De même, au niveau européen, des guides sont publiés pour accompagner les laboratoires dans l'évaluation et la validation du NGS pour le diagnostic des maladies génétiques, afin d'arriver à des résultats comparables quelle que soit la méthode de séquençage, les logiciels d'analyse et le paramétrage choisis (O'Rawe J et al., 2013) ; (Matthijs G et al., 2016).

Les approches par NGS présentent un avantage majeur, qui justifie leur diffusion dans les laboratoires. Par diagnostic avec la méthode de séquençage « Sanger », l'approche gène par gène est à ressource limitée, on génère une information de séquence en fonction de l'avancée du diagnostic. Au contraire, par NGS, on génère une réserve d'information qu'il est possible d'exploiter en fonction des besoins et de l'avancée du diagnostic : on n'est limité que par la capacité (ou les choix) d'analyse. Il est également possible de revenir sur les données brutes du patient, s'il donne son accord et si les données ont été conservées, en cas d'échec du diagnostic initial, pour ré-analyser un gène qui viendrait d'être impliqué dans une pathologie similaire, sans nécessiter un nouveau prélèvement.

1.5. GENOME, EXOME, PANEL DE GENES OU GENE UNIQUE, QUELLE STRATEGIE CHOISIR ?

Suivant la stratégie développée, mais aussi selon les plateformes à disposition, il existe actuellement plusieurs types d'approches haut-débit, correspondant à des échelles d'analyse différentes et offrant des rapports coût/bénéfices très variés. Dans tous les cas, une vérification de la mutation identifiée par NGS est pratiquée par séquençage Sanger, qui reste la méthode de référence dans les laboratoires de diagnostic hospitalier. L'ensemble des approches de séquençage disponibles est représenté sur la Figure 8.

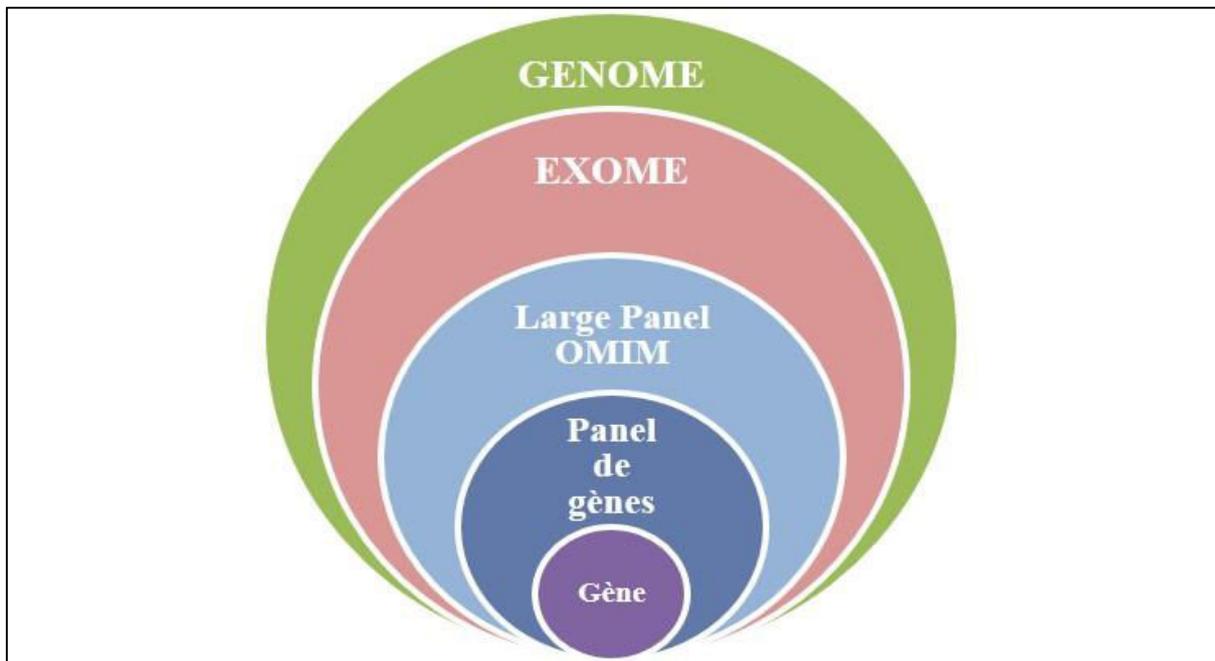


Figure 8 : Les différents niveaux d'approche du séquençage.

Un gène, cible de la méthode de diagnostic moléculaire classique en Sanger, représente en moyenne 3 kb de séquence. Le panel de gènes cible tous les gènes décrits pour une pathologie donnée. En NGS, il permet d'atteindre une grande profondeur de lecture et une bonne couverture car le nombre de régions à séquencer est faible, de l'ordre de quelques dizaines ou centaines de kb. Le large panel, commercialisé par plusieurs fournisseurs, cible tous les gènes déjà décrits comme responsables de pathologies. Ces gènes, au nombre de 3000 environ, couvrent environ 10 Mb de séquence. Le séquençage d'exome permet de cibler la totalité des séquences codantes des gènes (1 à 2% du génome, environ 62 Mb d'ADN), dans lesquelles sont décrites la majeure partie des mutations. Les régions d'intérêt étant plus grandes, les données de couverture et de profondeur sont moindres. Enfin, le séquençage du génome analyse l'ensemble de l'ADN d'un patient sans enrichissement (environ 3,2 Gb). Il reste très coûteux et génère un important volume de données. Il n'est pour le moment pas adapté en pratique diagnostique.

1.5.1. Panel de gènes d'intérêt

Dans un contexte clinique précis, les régions d'intérêt enrichies peuvent concerner un groupe de gènes particuliers responsables d'une pathologie donnée, comme par exemple l'ensemble des gènes impliqués dans les myopathies (Gorokhova S et al., 2015). On parle alors de NGS en **panel de gènes**. Le plus souvent, le panel sera constitué des gènes déjà décrits dans la littérature comme étant la cause de la pathologie à diagnostiquer, mais il peut s'agir également de gènes candidats. Dans ce cas cependant, la frontière entre le diagnostic et la recherche n'est plus très bien délimitée et les recommandations éthiques ne vont pas dans ce sens. Cette approche permet d'obtenir de très bons résultats en termes de couverture et de profondeur, car les régions ciblées sont petites. Les fournisseurs proposent soit des kits prêts à l'emploi (par exemple pour le gène *CFTR* impliqué dans la mucoviscidose ou les gènes de prédispositions au cancer du sein *BRCA1* et *BRCA2*), soit des kits à façon pour cribler certains gènes, dont la liste est alors établie par le laboratoire lui-même (Bell CJ et al., 2011). Ces kits à façon présentent l'avantage d'être très adaptés à une pathologie donnée, mais la liste des gènes couverts étant fixée au moment de la conception du kit, l'ajout d'un gène nouvellement décrit ne peut se faire qu'à la condition d'une nouvelle mise au point technique et d'une nouvelle validation, selon les recommandations de l'ESHG (Matthijs G et al., 2016). Le coût de ces kits est assez faible, mais la nécessité de les réactualiser augmente leur coût à long terme.

1.5.2. Séquençage d'exome

Si les régions enrichies regroupent toutes les parties codantes du génome, ou exons, on parle alors **d'exome** (ou de *WES*, pour *Whole Exome Sequencing*). Ces régions codantes ne représentent que 1 à 2% du génome total, soit environ 62 Mb d'information génétique, mais contiennent environ 58% des mutations mises en cause dans les maladies génétiques (Choi M et al., 2009). Le volume de données générées dans ce cas est important et plus difficile à analyser. En effet, le séquençage d'un exome révèle, selon la technique d'enrichissement utilisée, entre 20 000 et 50 000 variants. Après filtrage informatique strict, il subsiste en moyenne entre 150 et 500 variants privés non synonymes ou ayant un effet potentiel sur l'épissage qui peuvent être impliqués dans la pathologie (Xue Y et al., 2015). En conséquence, l'interprétation clinico-biologique de ces données représente encore un travail colossal qui visera à confirmer l'implication de l'un ou l'autre de ces variants. Les premières études publiées

sur des cohortes de nouveaux nés chez qui aucune exploration génétique n'avait été réalisée au préalable, montrent un taux de diagnostic d'au moins 25% dans le champ des déficiences intellectuelles, ce qui semble prometteur par rapport aux approches traditionnelles de diagnostic moléculaire, dans lesquelles une majorité des patients restent sans diagnostic (Gahl WA et al., 2012) ; (Yang Y et al., 2014). Actuellement, un exome revient à environ 600 à 1200 € selon les techniques utilisées.

1.5.3. Séquençage de large panel (Mutome) et filtrage bio-informatique

Une stratégie intermédiaire entre ces deux approches de panel et d'exome est possible, c'est le séquençage d'un **large panel OMIM** (*Online Mendelian Inheritance in Man*). Cela consiste à séquencer les gènes décrits en pathologie dans la base de données OMIM qu'on désigne aussi comme **mutome** ou **mendélianome**, version électronique de l'encyclopédie de référence des maladies mendéliennes. Ce séquençage peut être réalisé à partir de kits commerciaux mis au point par les fournisseurs, ce qui dédouane le laboratoire du travail de mise à jour technique. Dans cette stratégie, on peut choisir d'interpréter les données sur tous les gènes ciblés ou bien uniquement celles obtenues sur les gènes impliqués dans la pathologie explorée, grâce à un filtre informatique, ce qui revient à cibler un **panel bio-informatique**. Cette approche présente plusieurs avantages (Majewski J et al., 2011). D'une part, elle est robuste car elle est basée sur des sondes commerciales qui ont déjà passé une étape de validation par le fournisseur. Elle est adaptable à tous les types de pathologies et présente une bonne flexibilité, car si un nouveau gène doit être ajouté à l'analyse, c'est le pipeline d'analyse informatique et non pas la technique d'enrichissement qui est modifié. L'inconvénient principal est que les séquences ciblées étant plus nombreuses que lors d'un panel, la profondeur est moins importante en moyenne et peut être insuffisante dans certaines zones des régions cibles par rapport au panel de gènes. De même, les gènes récemment décrits et non intégrés dans la base OMIM ne sont pas encore dans ces kits commerciaux. Le coût de cette stratégie est un peu plus avantageux que les deux autres car le kit est actualisé et validé par le fournisseur à prix fixe, et l'on peut multiplexer les échantillons lors d'une même analyse.

1.5.4. Séquençage du génome entier

La dernière stratégie possible, mais nécessitant une plateforme NGS de grande capacité et une plateforme d'analyse de données brutes très performante, est le **séquençage du génome complet** ou *WGS* pour *Whole Genome Sequencing*.

Cette stratégie ne nécessite pas de technique d'enrichissement puisqu'elle consiste à séquencer l'intégralité du génome. Toutefois, elle est encore loin d'être utilisée en routine diagnostique et reste pour l'instant limitée à la recherche. L'analyse informatique des variants, quant à elle, reste focalisée sur les séquences des gènes. En effet, lorsqu'on séquence deux génomes d'individus pris au hasard (quelle que soit leur origine ethnique), on trouve en moyenne 99.9% d'identité de séquence (Jorde LB et al., 2004). Cela peut paraître élevé, mais cela implique qu'il subsiste environ 3 millions de différences ponctuelles ou SNP (*Single nucleotide polymorphism*) dont il faudrait pouvoir analyser l'effet pour poser un diagnostic. Par ailleurs, le coût de cette analyse est encore trop élevé pour être utilisé en routine diagnostique.

1.5.5. Quel choix pour quels résultats ?

Lorsque la pathologie est bien étiquetée, que le diagnostic clinique est précis et soutenu par des critères biochimiques ou anatomo-pathologiques certains, le NGS paraît peu indiqué, car les stratégies actuelles permettent généralement d'identifier la cause moléculaire de la pathologie. C'est le cas par exemple pour les hémophilies, la myopathie de Duchenne ou la mucoviscidose. Si le diagnostic peut être affiné par des examens paracliniques, comme dans les neuropathies périphériques héréditaires sensitives et motrices, mais que la pathologie est connue pour être hétérogène sur le plan génétique, un panel sera la solution la plus rapide, la plus efficace et la moins onéreuse. En revanche, lorsque le spectre clinique est peu spécifique, qu'il n'y a pas d'examen paraclinique et que la pathologie est multifactorielle, comme dans le cas des déficiences intellectuelles non syndromiques, alors le séquençage d'exome peut être indiqué pour poser un diagnostic.

Dans certains pays, aux États-Unis notamment, il est envisagé d'utiliser ce type d'approche dès la naissance dans les services de néonatalogie, chez les patients présentant des tableaux sévères, pour établir un diagnostic très précoce et permettre une prise en charge adaptée (Petrikin JE et al., 2015) ; (Smith LD et al., 2015).

1.6. INTERPRETATION DES VARIANTS ET IMPORTANCE DE LA DESCRIPTION CLINIQUE

Parmi les variants retenus après analyse informatique, plusieurs types de situations peuvent se présenter, pour lesquelles les difficultés d'interprétation diffèrent grandement. Le cas le plus simple est la découverte d'un variant connu dans un gène connu, les données bibliographiques et scientifiques levant toute ambiguïté sur la pathogénicité. L'interprétation est plus difficile si le variant est inconnu, c'est-à-dire non rapporté auparavant dans la littérature ou les bases de données. Elle reposera alors sur l'estimation de son caractère pathogène par différents moyens. Le cas d'un variant non connu dans un gène non connu dans la pathologie explorée nécessite des étapes d'exploration complémentaires qui relèvent plus du champ de la recherche que de celui du diagnostic.

Les petites insertions ou délétions de quelques bases ou les mutations non-sens ne posent généralement pas de problème d'interprétation et sont classées comme mutations pathogènes. Les mutations d'épissage sont également non ambiguës puisqu'elles perturbent *a priori* la production du transcrit. A l'opposé, les mutations qui n'ont pas d'impact sur la séquence protéique (mutations iso-sémantiques) sont généralement considérées comme non pathogènes.

La principale difficulté d'interprétation réside dans l'analyse des mutations faux-sens: elles correspondent à un changement d'acide aminé et peuvent avoir un effet très variable allant de l'absence de conséquence phénotypique jusqu'à un effet pathogène incompatible avec la vie. Plusieurs critères *a priori* sont utilisés pour tenter de prédire l'effet de ces substitutions, en s'appuyant sur des logiciels de prédiction de pathogénicité dont nous avons déjà parlé plus haut (Schwarz JM et al., 2010) ; (Salgado D et al., 2016).

L'évaluation du caractère pathogène d'un variant s'appuiera enfin sur une orientation clinique précise et sur l'étude familiale de ségrégation. Ainsi, pour valider l'identification d'une nouvelle mutation, même dans un gène connu, le mode de transmission et le lien entre le phénotype (les signes cliniques) et le génotype (la présence ou non de la mutation) chez plusieurs membres de la famille seront contrôlés.

Avec l'analyse en haut débit, le problème d'interprétation est forcément très axé sur les données cliniques et familiales.

La réinterprétation de variants rares rapportés dans la littérature peut amener à réévaluer leur pathogénicité. Ainsi, une récente étude sur les déficiences intellectuelles liées au chromosome X a remis en cause 10% des gènes précédemment impliqués dans la pathologie, car les variations de séquences rapportées se révèlent finalement fréquentes dans la population générale (Piton A et al., 2013).

1.7. IMPLICATION ETHIQUE ET SOCIETALE DU NGS

Le récent avis du Comité Consultatif National d'Éthique (CCNE) pour les sciences de la vie et de la santé sur la question du séquençage à haut débit suggère qu'il est urgent pour la communauté médicale de prendre la mesure de la révolution annoncée par ces technologies globales de séquençage (CCNE- avis N°124). Dans un avenir proche, il est possible que tout individu ait son génome séquencé dès la naissance, et que ces données l'accompagnent tout au long de sa vie, comme une carte d'identité. En effet, l'information génétique constitutionnelle est par nature propre à un individu et fixée pour la vie, contrairement aux informations épigénétiques ou aux mutations acquises. Ces considérations pourraient nous amener à vouloir définir une norme comme référence absolue, qui figerait la « séquence normale de référence de l'espèce ». Or, comme le démontre l'avis du CCNE, « la complexité biologique ne peut [...] s'accommoder de l'idée qu'un génome standard serait définissable. Imaginer qu'il existerait un « génome normal », figé en quelque sorte lors de l'évolution [...] et qu'il conviendrait de préserver, est une illusion. Parler de gènes normaux et de gènes mutés est une aberration, puisque tous nos gènes, produits de cette évolution, sont, par définition, des gènes mutés. Par ailleurs, il est absolument nécessaire que la pratique du NGS soit très clairement encadrée. D'après le CCNE toujours, « l'analyse globale du génome demande une justification en termes de nécessité (besoin de résoudre un problème clinique) et de proportionnalité (équilibre des avantages et inconvénients pour le patient). Quand un séquençage global d'ADN est décidé, un protocole rigoureux doit guider la conduite à tenir pour les découvertes non sollicitées ». En effet, lors du séquençage à haut débit, on se trouve confronté à des variants décrits comme étant pathogènes, mais dans des gènes que l'on ne souhaitait pas forcément explorer. On comprend intuitivement que, plus on analyse de larges portions d'ADN, plus on identifie de telles variations nucléotidiques, que l'on appelle « découvertes fortuites » (*incidental findings*).

Les résultats récents d'exploitation des données du projet 1000Genomes par exemple, montrent qu'un individu considéré comme exempt de toute pathologie porte entre 40 et 110 variants classés comme potentiellement pathogènes à l'issue de l'analyse, sans que cela soit toujours explicable dans l'état actuel des connaissances (Xue Y et al., 2012) ; (Rabbani B et al., 2014). Que faire dans ce cas ?

Un argument fréquemment avancé est que la connaissance du statut de porteur d'une « anomalie » génétique pourrait inciter l'individu à modifier ou à adapter son comportement, à l'instar des campagnes de prévention contre le tabac par exemple. Or, le même CCNE conteste cette vision utopique (naïve ?) des choses « dans un pays comme la France, dans lequel plus de 30% des citoyens fument, où un fumeur sur 2 mourra d'une maladie liée au tabac, et où nombre de patients ne suivent pas le traitement médical prescrit ». En opposition à la notion de risque subi, lié à l'hérédité, la notion de risque choisi pourrait devenir prépondérante, ce risque étant fonction du mode de vie adopté par la personne. Ainsi la maladie ne serait plus perçue comme une simple fatalité, mais un état contre lequel il serait possible de se prémunir en adaptant ses comportements. Qu'en sera-t-il alors de ceux qui n'adopteront pas le bon mode de vie ?

Enfin, le NGS fait bouger la frontière entre la pratique diagnostique, visant à identifier la cause moléculaire d'une pathologie chez un patient, et la recherche, qui tend à décrire de nouvelles causes de la pathologie pour en comprendre les mécanismes. Cette évolution est directement liée « à l'avènement d'une prétendue société de la connaissance, [comme le clament] les slogans tels que le rapprochement de la paillasse du lit du malade (*from bench to bedside*), qui se décline dans la promotion d'une recherche dite translationnelle » (CCNE- avis N°124). Jusqu'ici, le système de santé était organisé autour de cette stricte séparation, que les approches globales ont tendance à remettre en cause.

Les conditions de stockage et le respect de la confidentialité des données représentent d'autres problèmes qui ne sont pas encore parfaitement encadrés. Récemment, la société américaine *23&Me*, spécialisée dans le diagnostic moléculaire de susceptibilité génétique sans interface médicale (officiellement validée par l'autorité de santé américaine en octobre 2015) a vendu à l'entreprise de biotechnologie *Genetech*, pour quelques dizaines de millions de dollars, les informations personnelles et les prélèvements à des fins de re-séquençage, d'une cohorte de 14 000 personnes réalisée dans le cadre d'une étude sur la maladie de Parkinson. Qui est propriétaire des données générées au cours des analyses de NGS ?

Combien de temps conserver les données de séquence ? Une durée fixe ? Toute la vie de l'individu ? Comment assurer la sécurité des fichiers et des serveurs ? Où trouver le financement nécessaire à la constitution et à l'entretien de telles bases de données ?

Autant de questions qui restent actuellement sans réponse. Et qui nécessitent une réflexion de fond, pluridisciplinaire et mesurée, indépendamment des contraintes temporelles liées à la poursuite de nos pratiques.

1.8. PRESENTATION DU TRAVAIL DE THESE

Depuis quelques années déjà, les laboratoires de recherche utilisent les données obtenues par les techniques de NGS en faisant appel à de grandes plateformes comme le Centre National de Génotypage (CNG) ou des sociétés privées. Le grand succès de ces technologies dans le domaine de la recherche annonce une nouvelle ère dans le domaine du diagnostic génétique. Aussi, il est important que les laboratoires de diagnostic soient capables d'utiliser ces technologies pour répondre aussi bien aux demandes des prescripteurs qu'aux attentes des patients et de leur famille. Grâce au soutien financier de la DHOS, le laboratoire de Génétique Moléculaire de la Timone s'est équipé fin 2012 d'une plateforme NGS Ion Proton™.

Dans la deuxième partie de ce manuscrit, nous présentons l'ensemble du travail préliminaire nécessaire pour installer, développer et mettre en route une plateforme de NGS à des fins de diagnostic génétique dans un grand centre hospitalier, tout en s'inscrivant dans une démarche européenne d'harmonisation des pratiques (Matthijs G et al., 2016). Ce travail a notamment permis d'étudier les données de profondeur et de couverture sur une trentaine de données d'exome générés dans le laboratoire de Génétique Moléculaire de la Timone, paramètres importants pour rendre un diagnostic au patient. Notre pipeline d'analyse s'est restreint à l'étude de panels bioinformatiques de gènes ciblés dans les différents groupes de pathologies étudiées, au sein du laboratoire. Nous présentons également des exemples de cas pour lesquels cette méthode d'analyse a permis d'apporter des réponses rapides aux prescripteurs, chez des patients aux phénotypes pourtant complexes.

1. PARTIE 2

1.1. INTRODUCTION

L'efficacité du séquençage fondée sur l'approche des données issues d'exome a été prouvée depuis quelques années maintenant et a permis d'élucider des causes de maladies génétiques hétérogènes (Choi M et al., 2009) ; (Lupski JR, 2012) ; (Xue Y et al., 2015).

Le succès de ces technologies dans le domaine de la recherche a conduit au transfert de ces approches en diagnostic et ouvre une nouvelle ère dans le domaine du diagnostic des maladies génétiques hétérogènes (Neveling K et al., 2013) ; (Biesecker LG et al., 2014). Aussi, il est important que les laboratoires de diagnostic soient en mesure d'utiliser et d'évaluer ces technologies en lien étroit avec les prescripteurs pour répondre au mieux aux besoins des patients et de leur famille. Les fournisseurs développent de nouvelles plateformes dites de « pailleasse » plus accessibles tout en offrant plusieurs stratégies adaptées.

Au sein de du laboratoire de Génétique Moléculaire de la Timone, une cinquantaine de gènes impliqués dans des pathologies hétérogènes comme les myopathies, la déficience intellectuelle, les encéphalopathies épileptiques précoces, les neuropathies périphériques héréditaires sensibles et motrices ou les surdités sont séquencés en routine par la méthode de Sanger ce qui représente environ 8 Mb de données brutes de séquence.

Le laboratoire s'est équipé fin 2012 d'une plateforme NGS Ion Proton™ (Rothberg JM et al., 2011) (Thermo Fisher Scientific). Cette plateforme permet une mise en œuvre simplifiée du séquençage de l'exome dans les laboratoires de taille moyenne, produisant des données d'exome de haute qualité en moins de 8 heures (Boland JF et al., 2013). En effet, elle est capable de générer plus de 10 Gb de données brutes de séquence. Toutefois, pour utiliser le NGS en routine et en diagnostic, il faut au préalable valider la plateforme mais aussi la méthode d'enrichissement que nous utilisons pour le diagnostic chez les patients qui ont été adressés au laboratoire. Ceci est indispensable pour voir si nos besoins sont couverts par cette nouvelle technologie.

La question de la stratégie lors de la mise en place de la plateforme NGS s'est posée. En raison de la diversité des pathologies génétiques traitées, il fallait mettre en place un processus analytique commun, quelle que soit la pathologie pour laquelle les patients avaient été adressés (Figure 9). Une première stratégie basée sur la génération de données de séquençage d'exome a été testée et analysée en filtrant informatiquement uniquement sur des listes de gènes sélectionnées en fonction des groupes de pathologies. Ceci a permis de cribler les gènes correspondants à l'indication spécifiquement et d'adapter le diagnostic à la demande du clinicien. Ces listes de gènes ciblés pathologie dépendante ont été établies avec les cliniciens, les biologistes mais aussi les chercheurs dont le projet de recherche est centré sur la pathologie. De même pour des raisons de coût et de capacité de séquençage de la plateforme, l'analyse d'exome en trio (Cas index- père et mère *par exemple*) n'a pas été retenue.

Il est aussi important de réaliser une analyse détaillée de la couverture de séquence sur les listes de gènes d'intérêt mise en place au laboratoire. Ceci est indispensable pour voir si nos besoins sont couverts par cette nouvelle technologie et pour évaluer ce critère essentiel pour une utilisation en diagnostic selon les préconisations de l'ESHG (Matthijs G et al., 2016).

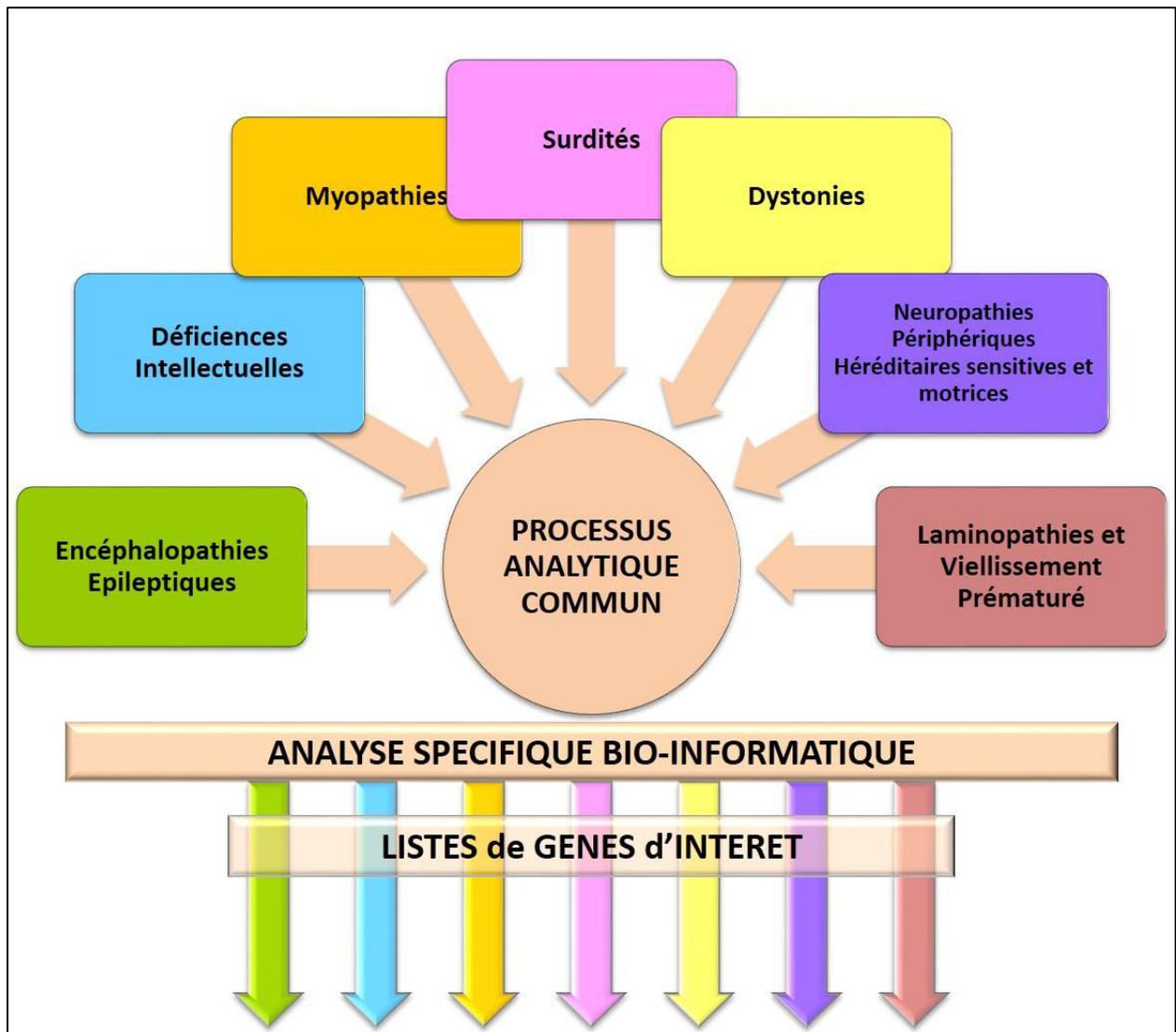


Figure 9 : Schéma de la stratégie de diagnostic par NGS au sein du laboratoire de Génétique Moléculaire – Timone Enfant.

Les patients qui sont adressés au laboratoire présentent des pathologies variées. La stratégie a été d'appliquer un processus analytique commun. L'ADN génomique est traité de façon identique quelle que soit la pathologie du patient. C'est au niveau informatique que chaque patient bénéficiera d'une analyse ciblée sur les gènes impliqués dans sa pathologie. Cette stratégie offre à la fois une bonne robustesse et une grande souplesse, permettant par exemple de pouvoir ouvrir la plateforme à de nouvelles demandes des cliniciens pour d'autres pathologies.

1.2. MATERIELS ET METHODES

1.2.1. Patients et ADN génomique

Lors de la consultation de génétique, le patient signe un formulaire de consentement éclairé conformément à la réglementation institutionnelle.

Le Centre de Ressources Biologique (CRB) du Département de Génétique Médicale et de Biologie Cellulaire de l'Hôpital La Timone procède à l'extraction de l'ADN et assure le stockage à long terme des échantillons. L'extraction d'ADN génomique à partir de 5 ml de sang total est automatisée à l'aide du robot Quiagen Autopure LS[®], en utilisant le kit Puregene[®] et le protocole Compromised Whole Blood (Qiagen).

La qualité et la quantité d'ADN extrait sont dosées au spectrophotomètre NanoDrop[®] ND-1000. L'ADN est considéré de qualité adéquate pour le séquençage si la concentration est supérieure à 100ng/μl, si le rapport d'absorbance A260/A280 est autour de 1.80 (+/- 0.1) et celui A260/A230 compris entre 2.0 et 2.20.

La quantité d'ADN est aussi mesurée en parallèle avec le fluoromètre Qubit[®] 2.0 en utilisant le kit The Qubit[™] dsDNA HS Assay, suivant les recommandations du fournisseur (Thermo Fischer). En fonction de la technique d'enrichissement utilisée, entre 50 ng et 1 μg d'ADN génomique sont nécessaires.

1.2.2. Kits d'enrichissement – Préparation des bibliothèques

Lors de la mise en route de la plateforme en 2013, plusieurs kits d'enrichissement ont été testés. Pour les études d'exome, les premiers essais ont été faits avec le kit TargetSeq[™] Exome (Cat n° 4477742, 4477743 ou 4477744 - Life Technologies[™] - Thermo Fischer). En parallèle, nous avons testé un kit SureSelect XT Human All Exon V5 (Cat n° 5190-3208- Agilent Technologies). Ces deux kits d'enrichissement sont basés sur le principe de capture (Figure 4 PARTIE 1) mais ils n'ont pas été retenus pour continuer la mise en place de la plateforme au sein du laboratoire. Puis en 2014, nous avons testé le kit Ion AmpliSeq[™] Exome Library (Cat n° 4487084 - Life Technologies[™] - Thermo Fischer) qui utilise le principe de PCR pour réaliser les bibliothèques des patients (Figure 4 PARTIE 1), et l'avons utilisé ensuite pour débiter l'activité diagnostic (Tableau 1). Pour chaque kit, nous avons suivi les recommandations du fournisseur.

Depuis 2015, nous utilisons le kit large panel SureSelect XT Inherited Disease (Cat n°5190-7520 et Cat n° G9605B - Agilent Technologies). Nous avons d'abord suivi le protocole et les recommandations du fournisseur pour un processus analytique manuel, puis nous l'avons adapté pour générer les bibliothèques par semi automatisation à l'aide du robot Bravo™ (Agilent Technologies).

1.2.3. Séquençage Plateforme Ion Proton

Les bibliothèques issues des différents kits d'enrichissement ont été séquencées sur la plateforme Ion Proton™ couplé au Torrent Server et Torrent Suite™ Software avec le système Ion OneTouch™2 /Enrichissement System (OT2/ES System) (Life Technologies™ - Thermo Fischer). Lors de la mise en route de la plateforme, la chimie de séquençage proposait les kits Ion PI™Template OT2-200 V2 (cat n° 4485146 - Life Technologies™ - Thermo Fischer) associés aux kits Ion PI™Sequencing200 V2 couplés avec les puces Ion PI™Chip V2 (Cat n°4485149 et Cat n° 4482321- Life Technologies™ Thermo Fischer). Plusieurs évolutions de chimie de séquençage ont été apportées depuis (Tableau 1). Nous avons utilisé par la suite la chimie V3 pour les kits Ion PI™Template OT2-200 V3 et kits Ion PI™Sequencing200 V3 (Cat n° 4488991 et 4488315- Life Technologies™ Thermo Fischer). Actuellement, avec l'automate Ion Chef™ mis en route pour remplacer le système OT2/ES (Life Technologies™ Thermo Fischer), nous utilisons la chimie Hi-Q™ (pour High Quality) avec le kit Ion PI™ Hi-Q™ IC (Cat n° A27198- Life Technologies™ Thermo Fischer) associé au kit Ion PI™ Hi-Q™ (Cat n° A26771- Life Technologies™ Thermo Fischer). Pour chaque kit, nous avons suivi les recommandations du fournisseur.

Tableau 1 : Résumé des différentes stratégies depuis la mise en place de la plateforme au sein du laboratoire

Le processus analytique change au fur et à mesure de l'évolution des kits de matériel disponibles au niveau de la construction des librairies mais aussi au niveau de la plateforme du séquenceur (les changements apparaissent **en gras**).

		STRATEGIE 1	STRATEGIE 2	STRATEGIE 2+
LIBRAIRIES	Kit d'enrichissement	AmpliSeq Exome	SureSelect XT Inherited Disease	SureSelect XT Inherited Disease
PROCESSUS TECHNIQUES- PLATEFORME SEQUENCAGE	Chimie de séquençage	Chimie V2-V3	Chimie V3	Chimie HiQ
	Amplification clonale	Système OT2-ES	Système OT2-ES	Système Ion Chef
	Chargement puce PI	Manuel	Manuel	Système Ion Chef
	Multiplexage	Non - passage en simplex	Oui - passage en triplex	Oui - passage en triplex

1.2.4. Analyse des données brutes issues du NGS

Les données brutes sont alignées et annotées sur hg19 une première fois par les logiciels d'alignement inclus dans le serveur Ion Torrent qui génère ainsi un fichier .bam (Figure 3). Les variants nucléotidiques sont déterminés et annotés en utilisant le Torrent Browser Variant caller (version 4.0.2) pour générer un fichier .vcf.

1.2.5. Analyse et Filtre des variants

À partir de ces deux fichiers, nous utilisons le logiciel VarAFT[®] (pour *Variant Annotation and Filtering Tool*). C'est un logiciel développé en interne qui annote et filtre les fichiers de variant pour les prioriser. Il est disponible sur le lien suivant : <http://varaft.eu/index.php>. Il permet non seulement de visualiser et de filtrer les variants candidats à partir des formats .vcf, mais aussi de générer et de visualiser les données de couverture à partir des formats .bam. Il interroge automatiquement les bases de données disponibles et les sites de prédiction de pathogénicité. Parmi ces sites, nous utilisons également l'outil *UMD-Predictor* (<http://umd-predictor.eu/index.php>) qui utilise une approche combinatoire en attribuant des scores de pathogénicité ou de polymorphisme sur les substitutions nucléotidiques (Salgado D et al., 2016). Il constitue une aide précieuse à l'interprétation des variants. Le site *Mutation Taster* (<http://www.mutationtaster.org/>) permet une évaluation rapide des variants en intégrant les données de fréquence et les données de pathogénicité (Schwarz JM et al., 2010). Ce site renseigne également sur la fréquence du variant issue des bases ExAC (*pour Exome Aggregation Consortium*, <http://exac.broadinstitute.org/>) et 1000 Genomes.

Nous utilisons également, pour des études *in silico* des effets sur la protéine mutée, les logiciels PolyPhen (<http://genetics.bwh.harvard.edu/pph2/>) et EIPD (<http://sift.jcvi.org/>) qui prédisent respectivement par un score, l'effet pathogène de notre changement d'acide aminé induit par la variation nucléotidique trouvée.

1.2.6. Calculs des données de couverture

Le logiciel VarAFT® permet de générer les données de couverture sur la liste de gènes candidats à la pathologie définie au préalable. Pour les analyses de couverture, nous fixons un seuil de profondeur de lecture de 20X.

Le pourcentage de couverture des gènes d'intérêt à 20X permet de définir 3 catégories :

1. Gènes couverts à 100% à 20X;
2. Gènes couverts entre 90 et 100% à 20X;
3. Gènes couverts à moins de 90% à 20X.

Nous considérons que les gènes d'intérêt sont suffisamment couverts quand la couverture moyenne est supérieure ou égale à 90% à 20X, ce qui correspond aux deux premières catégories. Il faut toutefois noter que même si un gène est couvert à plus de 90% à une profondeur de 20X, il est possible qu'un exon de ce gène soit mal couvert voire pas couvert du tout. Le logiciel permet d'extraire ces données et de connaître les exons mal ou bien couverts spécifiquement pour chaque gène de la liste ciblée de gènes.

1.2.7. Confirmation des variants sélectionnés

Les variants moléculaires mis en évidence par l'analyse des données NGS sont confirmés par un séquençage en Sanger qui reste la méthode de référence. Les références génomiques sont issues de la base de données UCSC sur la version hg19 du génome (<https://genome.ucsc.edu/>) pour nous permettre de dessiner des amorces de PCR criblant le variant à vérifier à l'aide du logiciel Primer 3 (v0.4.0 en libre accès sur <http://bioinfo.ut.ee/primer3-0.4.0/>). La PCR est réalisée avec une Taq Polymerase FIREPol® (Cat n°01-01-1000 Solis Biodyne) en suivant le protocole standard du fournisseur dans des conditions standard de PCR. Les produits de PCR sont ensuite purifiés avec le kit AmPure® (Cat n°A63881 - Beckman Coulter) en suivant le protocole du fournisseur avant d'être séquencés par le séquenceur multicapillaire 3500XL (Life Technologie – Thermo Fischer) avec le kit BigDye® Terminator v1.1 Cycle Sequencing (Cat n°: 4336774- Life Technologie – Thermo Fischer), toujours en suivant le protocole standard du fournisseur.

1.3. ANALYSE DES PERFORMANCES DE LA PLATEFORME ION PROTON™

1.3.1. Validation de la plateforme Ion Proton™

La stratégie adoptée initialement par le laboratoire a été de réaliser le séquençage d'exome, après utilisation d'un kit commercial pour tous les types d'échantillon, puis d'appliquer un filtre bio-informatique sur une liste de gènes candidats correspondants à la pathologie du patient étudié. Cette procédure technique a permis de générer des données de séquences rapidement et de manière robuste et reproductible, en lien avec les contraintes qualité des laboratoires hospitaliers (Dias C et al., 2012) ; (Bartoli M et al., 2014).

Un travail préliminaire a été cependant nécessaire pour maîtriser et évaluer la technologie NGS afin de valider les performances de la plateforme.

1.3.1.1. Comparaison de trois kits d'enrichissement d'exome

Lors de la mise en route de la plateforme, trois types de kits d'enrichissement ont été testés permettant de séquencer des exomes : le kit *Sure Select Exome* (Agilent Technologies) et le *TargetSeq Exome* (Thermo Fisher Scientific) qui fonctionnent par capture des régions cibles, et le kit *AmpliSeq Exome* (Thermo Fisher Scientific) avec lequel les régions exoniques sont obtenues par PCR (Figure 4 PARTIE 1).

Quatre patients atteints de myopathie et dont le diagnostic moléculaire avait été clairement établi auparavant par séquençage Sanger, ont été séquencés par la plateforme après traitement avec ces 3 kits d'enrichissement en parallèle. La technologie NGS est basée des indicateurs précis qui ne sont pas forcément rapportés dans le rapport de rendu de résultat, mais qui doivent être surveillés en continu. Parmi ces indicateurs, le nombre de bases totales chargées sur la puce, le nombre de *reads* pour chaque patient si nous multiplexons, la longueur moyenne de lecture, le nombre total de variant détectés et le paramètre AQ20 (c'est un score d'exactitude : lors de l'alignement sur le génome *hg19*, la probabilité que la base lue soit incorrect est de 1 pour 100) sont surveillés. Ces différents indicateurs évalués sont résumés dans le Tableau 2.

Tableau 2 : Comparaison des indicateurs de séquences sur les données d'exome

Les différents indicateurs de séquences ont été analysés à partir des données brutes issues du séquençage de 4 patients au statut moléculaire connu séquencés avec les 3 kits d'enrichissements sans multiplexage (simplex). L'indicateur Bases totales représente la somme des bases de tous les reads obtenus sur une puce (Proton Piv2) sans appliquer les filtres qualité. L'indicateur AQ20 est le nombre de bases réellement alignables sur le génome *hg19* où 99 % de chance que la base lue soit juste. Le nombre de *reads* est le nombre de lectures assignées à chaque patient via son code barre. La longueur moyenne indique la taille de ces *reads* en bases. Le nombre de variants est donné sans filtre informatique, ces données au format .vcf sont issues de l'alignement au format .bam des *reads* de chaque patient alignés sur le génome *hg19*. C'est à partir des fichiers .vcf et .bam que nous recherchons les variants pathogènes et vérifions les données de couverture à une profondeur de 20X minimum. Certaines données n'ont pu être déterminées (ND).

	Librairie Target Seq Exome (Thermo Fischer)					Sure Select Exome (Agilent)					AmpliSeq Exome (Thermo Fischer)				
	Bases Totales (Gb)	AQ20 (Gb)	Nombre de lectures (reads)	Longueur moyenne de lecture (pb)	Nombre total de variants	Bases Totales (Gb)	AQ20 (Gb)	Nombre de lectures (reads)	Longueur moyenne de lecture (pb)	Nombre total de variants	Bases Totales (Gb)	AQ20 (Gb)	Nombre de lectures (reads)	Longueur moyenne de lecture (pb)	Nombre total de variants
Patient 1	8,4	6,5	52 712 933	117	41 523	11,2	7,4	ND	132	46 102	8,3	5,8	38 600 181	151	49 286
Patient 2	9,8	7,7	78 095 565	125	34 369	10,7	8,3	ND	124	48 762	8,3	5,8	76 413 424	149	51 143
Patient 3	12,0	8,7	ND	112	29 352	10,2	8,2	ND	104	42 951	12,1	8,6	70 611 398	163	56 367
Patient 4	9,2	6,6	ND	127	36 805	9,8	7,7	ND	111	42 924	10,0	6,6	62 346 848	160	51 543

Le chargement de la puce se fait de manière manuelle. Le nombre total de bases chargées dépasse les capacités de la puce (habituellement 10Gb), il faut toutefois se référer au paramètre AQ20 qui permet d'évaluer la qualité des données recueillies. La longueur moyenne des lectures (ou *reads*) est plus importante en moyenne de 20 bases avec les librairies obtenues par PCR (*AmpliSeq Exome*), que par capture. Ces indicateurs de séquence renseignent sur la qualité des données brutes mais une analyse plus poussée avec le logiciel d'analyse peut renseigner sur le choix du meilleur kit d'enrichissement. Néanmoins avec une longueur moyenne de lecture plus longue avec le kit *AmpliSeq Exome*, de meilleures couvertures des régions cibles sont obtenues ainsi qu'un plus grand nombre de variants. Avec les trois kits d'enrichissement d'exome testés, les variants pathogènes ont été correctement et efficacement retrouvés par le pipeline d'analyse bio-informatique pour les quatre patients séquencés (patient 1 à 4 Tableau 3, hétérozygote composite et homozygote,).

Suite à ces premiers essais, la stratégie choisie a été de travailler avec le kit *AmpliSeq Exome* pour des raisons de facilité de mise en œuvre et de meilleures couvertures de régions cibles. En effet, l'obtention des librairies est plus rapide : une journée et demie, contre trois à cinq jours pour les deux autres kits par capture. Par ailleurs, il ne nécessite qu'une petite quantité d'ADN génomique (100ng). Ce kit était donc le kit d'enrichissement pour des données d'exome le mieux adapté à nos souhaits même en terme de prix de revient. Dans la suite du texte, cette stratégie sera désignée par « stratégie 1 » (Tableau 1).

1.3.1.2. Validation du kit *AmpliSeq Exome* –Stratégie 1

Pour valider la plateforme et la stratégie 1 choisie, nous avons complété les essais de validation en séquençant six autres patients (Patient 5 à 10 Tableau 3). Cette petite cohorte de dix patients au statut moléculaire connu comporte des variations pathogènes dans des gènes impliqués dans les dystrophies musculaires des ceintures (6 patients), dans les gènes impliqués dans des déficiences intellectuelles (2 patients) et dans des gènes d'encéphalopathies épileptiques précoces (2 patients).

Les différents indicateurs qui nous renseignent sur la qualité de l'échantillon séquencé sans multiplexage sont résumés dans le Tableau 3. Ces indicateurs sont systématiquement pris en compte au cours des différentes analyses réalisées par la suite.

Le nombre total de bases chargées en moyenne est de 11,3 Gb sur cette cohorte de 10 patients (entre 8,1 Gb et 14,1 Gb). Ces données sont ensuite filtrées et environ 30 % sont éliminées par les filtres qualité du serveur (données AQ20, soit 99 % de chances que la base soit juste). En moyenne le paramètre AQ20 est autour de 7,8 Gb. Le nombre de *reads* en moyenne sur ces 10 patients est de 68 000 pour une longueur moyenne de lecture de 162 bases et pour un nombre de variants détectés en moyenne par patient de 54 000 (Tableau 3).

Grâce à ces échantillons, le pipeline informatique des traitements des données de séquences a été validé, puisque les variants ponctuels déjà caractérisés en Sanger ont été détectés quel que soit le statut moléculaire du patient (Tableau 3). Toutes les analyses bio-informatiques des variants et des données de couverture ont été faites initialement au sein de l'équipe de recherche « Génétique et Bio-informatique » de l'UMR 910 à l'aide du logiciel VarAFT®. Le kit d'enrichissement *AmpliSeq Exome –Stratégie 1* a été donc validé ainsi que la plateforme. Les 10 patients (Tableau 3) au statut moléculaire connu ont été séquencés avec succès par la stratégie 1 puisque toutes les variations nucléotidiques responsables de leur pathologie ont été découvertes. Par cette stratégie, le diagnostic de patients en prospectif pour tenter de poser un diagnostic moléculaire peut donc être lancé.

Tableau 3 : Résumé des données des 10 patients utilisées pour la validation de la plateforme

Le statut moléculaire des gènes impliqués et les indicateurs de qualité des données brutes pour la cohorte de 10 patients choisis pour valider la plateforme Ion Proton avec la stratégie 1 sont résumés dans ce tableau. Les librairies ont été réalisées par AmpliSeq Exome™ et séquencées sans multiplexage (en simplex). Les indicateurs analysés sont les mêmes que ceux rapportés dans le Tableau 2.

	Statut	Bases totales (Gb)	AQ20 (Gb)	Nombre de lectures (<i>reads</i>)	Longueur moyenne de lecture (pb)	Nombre total de variants
Patient 1	CAPN3 (NM_000070) Hétérozygote composite	14,1	9,6	81 531 520	171	ND
Patient 2	DYSF(NM_003494) Homozygote	9,9	6,3	63 349 088	155	55 408
Patient 3	DYSF (NM_003494) Homozygote	12,1	8,6	70 611 398	169	56 367
Patient 4	CAPN3 (NM_000070) Hétérozygote composite	10,0	6,6	62 346 848	160	51 543
Patient 5	ANO5 (NM_213599) Hétérozygotes composite	13,2	9,1	77 868 586	168	56 742
Patient 6	CAPN3 (NM_000070) Hétérozygote composite	11,5	9,6	68 762 924	167	56 423
Patient 7	KCNQ2 (NM_172107) Hétérozygote	8,1	5,9	51 145 133	158	54 157
Patient 8	ATRX (NM_000489) Hémizygote	11,9	8,7	68 115 092	169	51 452
Patient 9	ATRX (NM_000489) Hémizygote	9,3	6,0	61 986 936	148	61 520
Patient 10	KCNQ2 (NM_172107) Hétérozygote	12,6	8,4	78 034 307	160	49 920

1.3.2. Validation du processus d'enrichissement AmpliSeq Exome™

Une fois la validation de la plateforme réalisée sur la cohorte de 10 patients au statut moléculaire connu, l'analyse de 31 échantillons adressés au laboratoire pour différentes pathologies a été effectuée (Figure 9). Parallèlement à la mise en place du diagnostic sur les premiers patients, les données de séquençage pour évaluer dans son ensemble la stratégie 1 déployée au sein du laboratoire ont été générées et étudiées sur des listes de gènes présentant un intérêt pour le laboratoire.

Ces résultats ont fait l'objet d'une publication soumise en 2015 et parue en 2016 dans *Journal of Genetics* « **Coverage Analysis of Lists of Genes involved in Heterogeneous Genetic Diseases following Benchtop Exome Sequencing using the Ion Proton™** » dont le texte est présenté ci-après.

En résumé, les données de couverture à une profondeur de 20X pour les 31 exomes ont été générées. Ceci pour déterminer spécifiquement les couvertures sur les différentes listes de gènes établies pour 6 groupes hétérogènes de maladies génétiques : les myopathies, les neuropathies périphériques héréditaires sensitives et motrices, les encéphalopathies épileptiques précoces, les dystonies isolées et combinées, les surdités non syndromiques-pertes auditives héréditaires et la déficience intellectuelle. Les données sur une liste de 57 gènes « actionnable » pour lesquels *l'American College of Medical Genetics (ACMG)* recommande de signaler les découvertes fortuites (Green RC et al., 2013) ont été étudiées. Les données de couverture des listes de gènes sont des données essentielles pour valider la qualité de l'échantillon séquençé et le rendu du diagnostic.

À l'exception des dystonies isolées et combinées, les 2/3 des gènes d'intérêt analysés pour les 6 grands groupes hétérogènes de maladies génétiques sont couverts à plus de 90% pour une profondeur de 20X. On note cependant qu'il existe une variabilité de couverture sur certains gènes d'intérêt d'un patient à un autre. Si la stratégie séquençage d'exome est largement utilisée en recherche, sa mise en place dans les laboratoires de diagnostic ne doit pas en être limitée par des aspects techniques.

Le processus analytique de la stratégie 1 développé avec la plateforme Ion Proton™ semble adapté si on ne multiplexe pas les patients sur la puce. Mais il est nécessaire d'améliorer ces

données de couverture pour cribler au mieux les gènes ciblés des listes des pathologies d'intérêt du laboratoire.

ARTICLE 1

Coverage Analysis of Lists of Genes involved in Heterogeneous Genetic Diseases following Benchtop Exome Sequencing using the Ion Proton™

RESEARCH NOTE

Coverage analysis of lists of genes involved in heterogeneous genetic diseases following benchtop exome sequencing using the ion proton

CAROLINE LACOSTE^{1,2}, JEAN-PIERRE DESVIGNES¹, DAVID SALGADO¹, CHRISTOPHE PECHEUX², LAURENT VILLARD^{1,2}, MARC BARTOLI^{1,2}, CHRISTOPHE BEROUUD^{1,2}, NICOLAS LEVY^{1,2}, CATHERINE BADENS^{1,2} and MARTIN KRAHN^{1,2*}

¹Aix Marseille Université, INSERM UMR_S 910, GMGF, 13385, Marseille, France

²APHM, Département de Génétique Médicale, Hôpital Timone Enfants, 13385, Marseille, France

[Lacoste C., Desvignes J.-P., Salgado D., Pecheux C., Villard L., Bartoli M., Beroud C., Levy N., Badens C. and Krahn M. 2016 Coverage analysis of lists of genes involved in heterogeneous genetic diseases following benchtop exome sequencing using the ion proton. *J. Genet.* **95**, 203–208]

Introduction

Exome sequencing (ES) has been proven efficient for clinical applications in heterogeneous genetic diseases (Neveling *et al.* 2013; Biesecker and Green 2014), including ES associated with data filtering for selected genes (Dias *et al.* 2012; Bartoli *et al.* 2014). However, previous reports have been based mainly on the use of large-scale next-generation sequencing (NGS) platforms, difficult to implement in small to medium scale genetic diagnosis laboratories. Recent technological developments include the Ion ProtonTM (Rothberg *et al.* 2011) (Life Technologies, Carlsbad, USA), allowing ES through an easy setup in 6 h runs on a benchtop apparatus. The Ion ProtonTM has been the first commercialized platform specifically adapted to the environment and diagnostic setting of small to medium scale laboratories, in particular through the simplified workflow and considerably shortened sequencing run times, while the production of high quality exome data has been demonstrated, including comparable average depth and read length, and accuracy for SNP variant calls between the Ion ProtonTM semiconductor technology and the Illumina technology (Boland *et al.* 2013; Motoike *et al.* 2014).

The rationale of our study was to specifically evaluate sequence coverage using Ion AmpliSeqTM exome enrichment and Ion ProtonTM sequencing, for genes implicated in diverse genetically heterogeneous diseases, to determine which genes may be retained using this procedure as first-tier mutation screening.

We selected the Ion ProtonTM platform to implement ES associated with data filtering for selected genes, to allow for standardization of mutation screening procedures for

the analysis of diverse groups of heterogeneous diseases, through a common workflow, not feasible through the use of disease-group-specific gene-panel sequencing. While available sequence coverage data for different platforms mainly focus on average exome values, the selection of genes which can be analysed for diagnostic purposes through first-tier ES warrants targeted coverage analysis of genes of interest. Therefore, as a requirement for diagnostic mutation screening using Ion ProtonTM ES, we evaluated sequence coverage data for six groups of genetically heterogeneous diseases, of interest for genetic diagnostic laboratories (table 1 in electronic supplementary material at <http://www.ias.ac.in/jgenet/>): myopathies (82 genes), hereditary motor and sensory neuropathies (55 genes), early onset epileptic encephalopathies (30 genes), isolated and combined dystonia (12 genes), non-syndromic deafness and hereditary hearing loss (60 genes), and intellectual disability (107 genes for X-linked transmission; 39 genes for autosomal recessive transmission, 37 genes for autosomal dominant transmission). In addition, we evaluated sequence coverage for a list of 57 genes for which the American College of Medical Genetics and Genomics (ACMG) recommends to report incidental findings (Green *et al.* 2013).

Methods

We generated and analysed sequence coverage data of 45 exome runs (14 duplex runs following barcoding using Ion XpressTM Barcode Adapters (Life Technologies) and 31 simplex runs) from DNA samples from our Biological Resources Center (Timone Hospital, Marseille). This study was conducted according to legal institutional and national regulations, and according to the Declaration of Helsinki

*For correspondence. E-mail: martin.krahn@univ-amu.fr.
Catherine Badens and Martin Krahn contributed equally to this work.

Keywords. exome; coverage; diagnosis; sequencing; mutation.

on ethical principles for medical research involving human subjects. All patients signed a consent form before the procedure. ES was performed using Ion AmpliSeq™ exome enrichment and library preparation (Life Technologies), template preparation using the Ion PI™ Template OT2-200-Kitv2 on the Ion OneTouch™ 2 System (Life Technologies), and sequencing using the Ion PI™ Chip-Kitv2 and Ion PI™ Sequencing200-Kitv2 on the Ion Proton™ Sequencer (Life Technologies) with sequencing data processing using the Torrent Suite™ Software (ver. 4.0.2) on a Torrent Server (Life Technologies) (figure 1). The Torrent Suite™ has been used to handle the bioinformatics data analysis from the acquisition through the generation of bam files. The suite chains several applications; we used the default parameters for all of them. The base calling step has been accomplished using the Ion Torrent's base-caller algorithm. Quality

trimming has been performed using the per-base quality scores. The window size is set to 30 bases, and the threshold below which the trimming occurs is a quality score of 15. The alignment was performed using the TMAP aligner with map4 parameter, which is the default option to balance rapidity and maintain a high degree of sensitivity and specificity, on human reference genome 19.

The Ion AmpliSeq™ exome enrichment kit targets ~33 Mb of coding exons (97% of coding regions as described by consensus coding sequences (CCDS) annotation), corresponding to a total design coverage including padding and flanking regions (mean number of five flanking base pairs 5' and 3' of coding exons) of ~58 Mb. To comprehensively determine the coverage for the gene lists of interest, we selected for each gene in all lists the longest CCDS transcript isoforms (as recommended by the Human Genome Variation

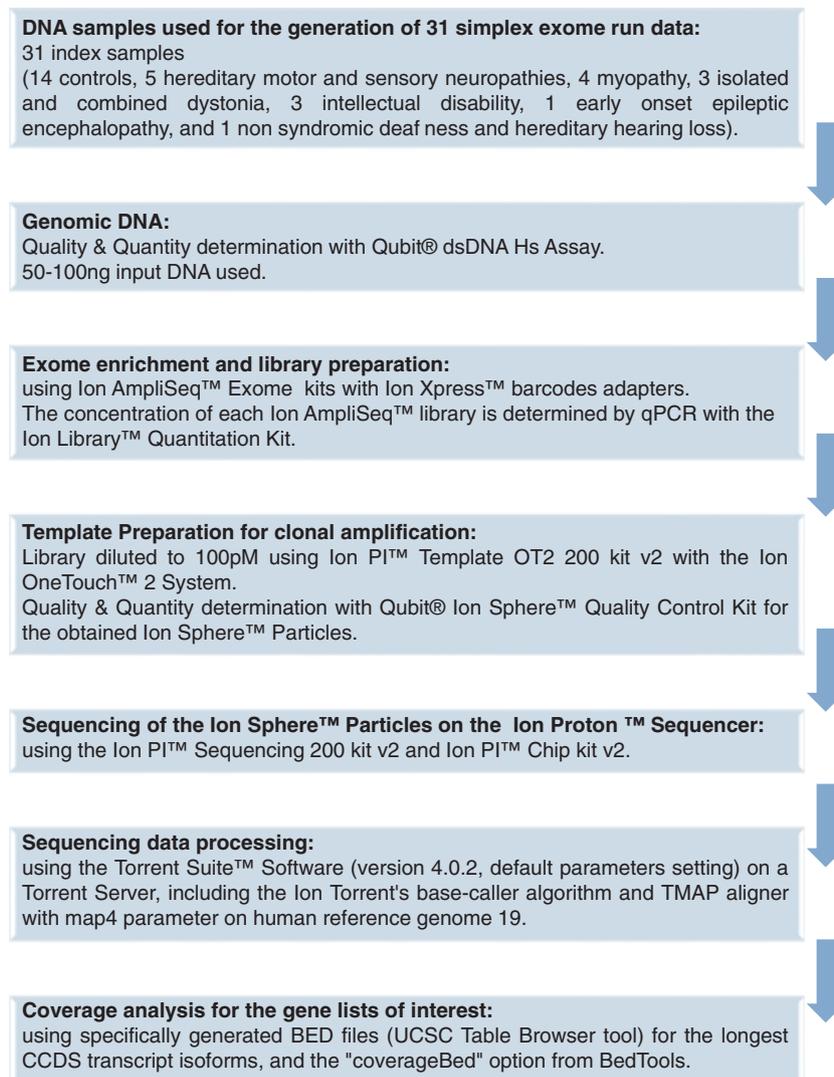


Figure 1. Workflow used to determine the mean sequence coverage at $\geq 20\times$ for simplex exome runs analysed in the present study. All reagents are from Life Technologies and were used according to the manufacturer's recommendations.

Society guidelines (<http://www.hgvs.org/mutnomen>) for mutational analyses). We then used the UCSC Table Browser tool (<http://genome.ucsc.edu/cgi-bin/hgTables>) to retrieve a single BED file containing all coding exons. Effective coverage calculation based on obtained sequencing data from 31 simplex runs was based on the generated BED files specifically for the respective transcripts, using the 'coverageBed' option from BedTools (Quinlan and Hall 2010), to report the depth of each position for each sample processed. Finally, we implemented a perl script to group the coverage values at exon and transcript levels. This perl script summarized the per cent of base pairs covered at different depth levels (1×, 5×, 10×, 20× and 30×). In our study, genes were considered as sufficiently covered for diagnostic mutation screening at a ≥90% mean sequence coverage at ≥20×, with mean exome sequencing depths ≥100×, based on the guidelines from the American College of Medical Genetics and Genomics (Rehm *et al.* 2013).

Results

The sequence data generated for 45 exome runs yielded an average of 11.2 Gb per run (ranging from 6.6 to 15.8 Gb; standard deviation (SD) 2.38), including an average of 7.68 Gb of AQC20 (ranging from 4.3 to 12.5 Gb; SD: 2.08), corresponding to a mean of 70765587 reads (ranging from 40981018 to 90139803; SD: 12264334) with a mean length of 157.62 bases (range from 129 to 178; SD: 12.48). The exome sequence depth obtained for simplex runs ranged from 103× to 223×, with a mean value of 162× (SD: 36) for the 31 analysed simplex exome runs. This corresponds to mean sequence coverage of 96.26% at 1× (SD: 0.33), 92.64% at 10× (SD: 1.02) and 89.85% at 20× (SD: 1.98).

Using duplex runs, the exome sequence quality decreased to a mean value of 93× (ranging from 42× to 143×, SD: 25), corresponding to a mean sequence coverage of 95% at 1× (SD: 2.86), 88.82% at 10× (SD: 5.35) and 82.48% at 20× (SD: 7.08). Therefore, further detailed coverage analysis focussing on the different gene lists was performed for the 31 simplex runs. To identify genes suitable or not for diagnostic mutation screening, we determined mean sequence coverage at a depth of at least 20× (recommended for the detection of heterozygous constitutive germline mutations) as detailed in tables 1 and 2 in electronic supplementary material.

Among the different gene lists of interest, genes considered as sufficiently covered for diagnostic mutation screening (≥90% mean sequence coverage at ≥20×) distributed as: 58 (71.9%) among the 82 genes involved in myopathies; 45 (81.8%) among the 55 genes involved in hereditary motor and sensory neuropathies; 18 (60%) among the 30 genes involved in early onset epileptic encephalopathy; six (50%) among the 12 genes involved in isolated and combined dystonia; 45 (75%) among the 60 genes involved in

nonsyndromic deafness and hereditary hearing; 77 (71.9%) among the 107 genes involved in X-linked intellectual disability; 29 (74.3%) among the 39 genes involved in autosomal recessive intellectual disability; 24 (64.8%) among the 37 genes involved in autosomal dominant intellectual disability; and 38 (66.6%) among the 57 genes for which the ACMG recommends to report incidental findings to patients. Among the different experiments, mean sequence coverage values are consistent for the respective lists of interest (figure 2).

Discussion

First-tier mutation screening using ES associated with data filtering offers promising perspectives for molecular diagnosis of genetically heterogeneous diseases. While most initial diagnostic applications of NGS relied on the use of disease-specific gene panel enrichment procedures (Rehm 2013), the feasibility of molecular diagnosis using ES has been demonstrated in the past years through the pioneering work of different laboratories having access to large-scale NGS platforms (Choi *et al.* 2009; Ng *et al.* 2010; Neveling *et al.* 2013; Yang *et al.* 2013; Biesecker and Green 2014). As compared with disease-specific gene panels, ES associated with data filtering for selected genes has the advantage to allow for regular updating of gene lists of interest, with regard to the always growing number of genes in several diseases groups, such as intellectual disability and neuromuscular disorders. Effectively, updating of disease-specific gene panel enrichment procedures through adding novel targets readily leads to disequilibrium in amplification and/or sequence capture performances, and therefore warrants customized revalidation of the respective gene panels at each update. On the contrary, for ES associated with data filtering for selected genes, updating of gene lists simply requires modifying the bioinformatics filtering settings. Moreover, ES offers the second-tier possibility of analysing the whole exome, including genes not previously implicated in the respective disease-groups of interest. In particular, in small to medium scale diagnostic laboratories implicated in the diagnosis of genetically heterogeneous diseases, but with limited numbers of proceeded samples for the respective diseases, the aforementioned flexibility constitutes in our opinion an important element of choice for using ES associated with data filtering for selected genes as an alternative to disease-specific gene panel enrichment procedures.

Importantly, as ES is being more widely used in genetic diagnosis, its implementation in small to medium scale diagnostic laboratories should not be limited by technical hurdles relating to large-scale platforms, to allow for large access and availability in prospective diagnosis. The simple workflow of the Ion Proton™ platform (Rothberg *et al.* 2011) is adapted to this purpose, with an easy setup and technical handling for ES in a benchtop format while producing high quality exome data (Boland *et al.* 2013).

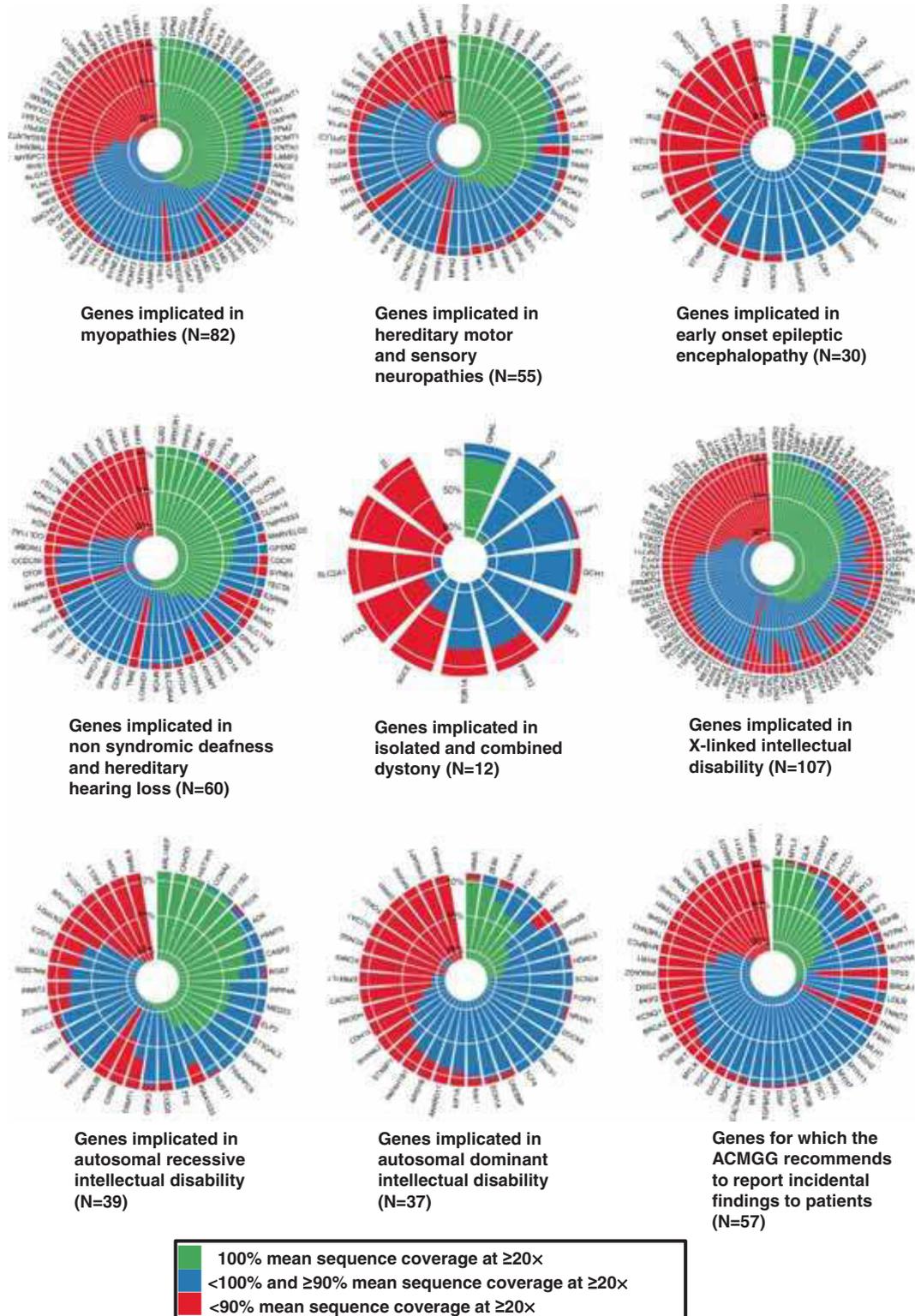


Figure 2. Sequence coverage at a depth of $\geq 20\times$ among 31 simplex exome runs for the analysed gene lists. The polar histogram shows the distribution of sequence coverage at a depth of $\geq 20\times$ among the 31 simplex exome runs analysed (representing 100% on the radial axis), for each gene of the respective gene lists of interest. Polar histograms were plotted using R software scripts and a specific library developed by C. Ladroue (<http://chrisladroue.com/2012/02/polar-histogram-pretty-and-useful/>).

Our results determine for the first time the specific genes implicated in diverse genetically heterogeneous diseases, for which the Ion Proton™ platform in its current configuration is suitable for first-tier mutation screening using ES associated with data filtering.

While mean sequence coverage values are consistent among the different experiments, variability was observed for individual genes of the respective lists of interest (figure 2), and has to be taken into account. In fact, it constitutes a criteria of optimization (including through automation), that may be used for the definition of thresholds such as total number of generated AQ20 reads before variant calling steps. Importantly, exception made for the list of 12 genes involved in isolated and combined dystonia, approximately two-third of all genes among the different groups are efficiently sequenced at >90% mean sequence coverage at $\geq 20\times$.

The sequence coverage achieved in the current configuration is adapted for efficient first-tier mutation screening using ES associated with data filtering for the majority of genes, and anticipated increase in diagnostic yield with regard to the large genetic heterogeneity of the diseases addressed here. More specifically, the information on the mean sequence coverage values determined in our study gene by gene among the analysed lists (table 2 in electronic supplementary material), allows for retaining or excluding selected genes for diagnostic applications using the first-tier Ion AmpliSeq™ exome enrichment and Ion Proton™ sequencing workflow, while mutational analysis of genes currently insufficiently covered using this procedure can rely on supplementation with second-tier Sanger sequencing for full coverage, as previously demonstrated (Dias *et al.* 2012). Genomic regions to be targeted with second-tier Sanger sequencing can be identified precisely for any sample of interest using the same coverage analysis workflow as used in our study for the determination of mean sequence coverage values. Importantly, to avoid biases relating to variability among exome sequencing runs, this should optimally rely on individual coverage determination on a sample by sample basis, to ensure full coverage, as exemplified for one selected sample in tables 3 and 4 in electronic supplementary material, respectively for the 82 genes involved in myopathies, and the 57 genes for which the ACMG recommends to report incidental findings to patients. In particular, supplementation with second-tier Sanger sequencing has to be considered for cases in which clinical data strongly orientate towards the implication of specific genes among the respective lists (i.e. suspected specific single gene disorders), or if comprehensive gene list screening is required.

For laboratories implementing the Ion Proton™ benchtop platform for first-tier mutation screening using ES associated with data filtering, the choice of providing either a mutation screening approach (with precisely determined associated limitations especially regarding coverage of the coding regions of interest), or a comprehensive gene list analysis, will further determine the diagnostic strategy and inclusion

of second-tier sequencing supplementation, while the future increase in sequencing capacities of the platform should allow for correlated increase in sequence coverage.

Acknowledgements

We sincerely thank the patients for their contribution to this study. We thank Nicole Philip, Emmanuelle Salort-Campana, Jean Pouget, Shahram Attarian, Véronique Blanck-Labelle, Annachiara DeSandre-Giovannoli, Rafaëlle Bernard, Nathalie Bonello-Palot, Perrine Malzac, Patrice Bourgeois, Mathieu Milh, Gwenaëlle Collod-Bérout, Karine Bertaux, Cécile Mouradian and Valérie Delague for their contribution to this work. We also thank the Association Française contre les Myopathies, the Jain Foundation, Assistance Publique Hôpitaux de Marseille, Inserm, Aix-Marseille Université, ARS-PACA and Agence de la Biomédecine for supporting this work. The researchers implicated in this project also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 2012-305121 'Integrated European -omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS).

References

- Bartoli M., Desvignes J. P., Levy N. and Krahn M. 2014 Exome sequencing as a second-tier diagnostic approach for clinically suspected dysferlinopathy patients. *Muscle Nerve* **50**, 1007–1010.
- Biesecker L. G. and Green R. C. 2014 Diagnostic clinical genome and exome sequencing. *N. Engl. J. Med.* **371**, 1169–1170.
- Boland J. F., Chung C. C., Roberson D., Mitchell J., Zhang X., Im K. M. *et al.* 2013 The new sequencer on the block: comparison of life technology's proton sequencer to an illumina hiseq for whole-exome sequencing. *Hum. Genet.* **132**, 1153–1163.
- Choi M., Scholl U. I., Ji W., Liu T., Tikhonova I. R., Zumbo P. *et al.* 2009 Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101.
- Dias C., Sincan M., Cherukuri P. F., Rupps R., Huang Y., Briemberg H. *et al.* 2012 An analysis of exome sequencing for diagnostic testing of the genes associated with muscle disease and spastic paraplegia. *Hum. Mutat.* **33**, 614–626.
- Green R. C., Berg J. S., Grody W. W., Kalia S. S., Korf B. R., Martin C. L. *et al.* 2013 ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574.
- Motoike I. N., Matsumoto M., Danjoh I., Katsuoka F., Kojima K., Nariai N. *et al.* 2014 Validation of multiple single nucleotide variation calls by additional exome analysis with a semiconductor sequencer to supplement data of whole-genome sequencing of a human population. *BMC Genomics* **15**, 673–686.
- Neveling K., Feenstra I., Gilissen C., Hoefsloot L. H., Kamsteeg E. J., Mensenkamp A. R. *et al.* 2013 A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum. Mutat.* **34**, 1721–1726.
- Ng S. B., Buckingham K. J., Lee C., Bigham A. W., Tabor H. K., Dent K. M. *et al.* 2010 Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35.
- Quinlan A. R. and Hall I. M. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.

- Rehm H. L. 2013 Disease-targeted sequencing: a cornerstone in the clinic. *Nat. Rev. Genet.* **14**, 295–300.
- Rehm H. L., Bale S. J., Bayrak-Toydemir P., Berg J. S., Brown K. K., Deignan J. L. et al. 2013 ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **15**, 733–747.
- Rothberg J. M., Hinz W., Rearick T. M., Schultz J., Mileski W., Davey M. et al. 2011 An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352.
- Yang Y., Muzny D. M., Reid J. G., Bainbridge M. N., Willis A., Ward P. A. et al. 2013 Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511.

Received 15 April 2015, in revised form 2 July 2015; accepted 11 August 2015

Unedited version published online: 14 August 2015

Final version published online: 24 February 2016

1.3.3. Données complémentaires

Au cours de l'année 2014, la stratégie a été changée en modifiant le kit d'enrichissement suite à plusieurs constatations. Tout d'abord, les données de couverture publiées des 31 premiers patients montrent que seulement deux tiers des gènes d'intérêt analysés pour les six groupes de maladies génétiques hétérogènes sont couverts à plus de 90% pour une profondeur de 20X et ceci en raison de la taille importante des régions ciblées (Lacoste C et al., 2016) Les possibilités pour améliorer cette couverture étaient soit de diminuer le nombre de patients séquencés simultanément sur une puce et fixé initialement à 2, soit de diminuer la taille des régions cibles tout en gardant un processus commun quelle que soit la pathologie. L'inconvénient de diminuer le multiplexage des patients est l'augmentation importante de coût du séquençage et par là, du coût de l'analyse. Pour ne pas diminuer le multiplexage, le nombre de séquences cibles a été diminué et un kit d'enrichissement commercial correspondant à un large panel de gènes décrits dans la base de données OMIM (voir PARTIE 1 section 1-5-3) a été testé. La majorité des gènes ciblés dans les listes d'intérêts ont été décrits depuis plusieurs années et sont bien caractérisés. Ils figurent donc majoritairement dans les kits larges panels proposés dans le commerce. Cette nouvelle stratégie est appelée Stratégie 2 (Tableau 1).

1.3.3.1. Validation du kit Sure Select XT Inherited Diseases

Le kit *Sure Select XT Inherited Diseases* (Agilent Technologies) a été testé. Ce kit d'enrichissement utilise la méthode de capture pour séquencer les régions ciblées (Figure 4 PARTIE 1). Avec cette stratégie 2 appliquée sur deux patients déjà criblés avec la stratégie 1, les données de couverture à une profondeur de 20X ont été évaluées. Les indicateurs de séquence des données obtenues de deux mêmes patients avec deux stratégies d'enrichissement différentes (Stratégie 1- exome vs Stratégie 2-large panel) (Patient 11 et 12- Tableau 4) ont été comparés.

Tableau 4 : Comparaison des stratégies 1 et 2 sur des indicateurs de données de séquence

Le patient 11 et le patient 12 ont été séquencés suivant deux techniques d'enrichissement différentes. Pour la stratégie 1, les données recueillies sont issues d'un séquençage d'exome et pour la stratégie 2 pour des données issues d'un large panel de gènes. Le patient 11 a été séquencé en multiplexant deux patients par puce (duplex) avec la stratégie 1, le nombre de reads est donc plus petit que pour un séquençage en simplex.

	STRATEGIE 1 – Régions cibles 62 Mb					STRATEGIE 2 - Région cibles 10 Mb				
	Bases Totales (Gb)	AQ20 (Gb)	Nombre de lectures (reads)	Longueur moyenne de lecture (pb)	Nombre total de variants	Bases Totales (Gb)	AQ20 (Gb)	Nombre de lectures (reads)	Longueur moyenne de lecture (pb)	Nombre total de variants
Patient 11	12,3	8,0	40 571 192	158	50 293	9,8	7,0	78 313 510	123	9 289
Patient 12	12,5	8,4	79 625 637	155	51 161	8,6	6,9	66 982 300	110	9 006

Les régions ciblées par le kit d'enrichissement de la stratégie 2 sont 6 fois moins nombreuses, puisque le kit cible environ 3 000 gènes soit 10 Mb de séquence, alors que le kit d'enrichissement de la stratégie 1 couvre l'exome, c'est à dire 1 à 2% du génome soit 62 Mb (Tableau 4). Cette réduction de la taille des régions cibles laissait espérer de meilleures données de couverture que pour la stratégie 1 (comparaison des données de couvertures détaillées plus bas). C'est pour la même raison que le nombre de variants obtenus après filtrage des données brutes est 6 fois plus faible (en moyenne 51 000 variants en données d'exome contre 8 500 pour le kit *Sure Select XT Inherited Diseases Panel*).

La longueur moyenne des lectures (ou *reads*) est plus petite de 20 bases pour cette stratégie 2, mais les données de couvertures des régions cibles sont meilleures puisque plus petites. Le nombre de *reads* par patient étant plus grand, la profondeur moyenne est améliorée. Les variants pathogènes attendus chez les patients ont été mis en évidence quelle que soit la stratégie utilisée, validant également le logiciel d'analyse et notre pipeline informatique.

Cette stratégie 2 a permis d'augmenter la profondeur et la couverture des régions cibles puisque celles-ci sont plus petites, mais surtout de séquencer plusieurs patients sur une même puce en faisant des multiplexages. Ainsi, non seulement la capacité de séquençage est augmentée mais le coût de l'analyse en routine est diminué.

1.3.3.2. Evolution des données totales de chargement sur la puce Ion™ Proton

Parallèlement à cette évolution du kit d'enrichissement qui a conduit à mettre en place la stratégie 2, le fournisseur de la plateforme a apporté des améliorations dans la chimie de séquençage (chimie version V2, puis V3 et enfin *HiQ*) ou dans les étapes d'amplification clonale (Figure 6 Partie 1). Les différents processus analytiques ou stratégies employées au sein du laboratoire sur la plateforme NGS sont résumés dans le Tableau 1.

De même, le système *Ion Chef™* remplaçant le système OT2/ES a été mis en place. Ce système permet d'automatiser plusieurs étapes, dont l'amplification clonale et celle du chargement des puces de séquençage, améliorant ainsi l'efficacité et la reproductibilité. On parlera dans la suite du texte de la stratégie 2+ (Tableau 1).

Si on analyse les données brutes de chargement au cours du temps, on remarque une nette amélioration entre la version de séquençage de la chimie V3 et la version HiQ (Figure 10).

Une puce Proton doit pouvoir fournir au moins 10G de données brutes. Sur les 25 premières puces chargées manuellement, les données brutes réelles sont légèrement inférieures à 10G avec la chimie V3 pour une moyenne de 8,6 Gb (mini 6,9 Gb- maxi 11,3 Gb). Avec la nouvelle chimie HiQ, les 10G de données brutes sont obtenues de manière plus régulière avec une moyenne de 10,7 Gb (mini 8,3Gb - maxi 13,1 Gb). L'effet de l'utilisation du *Ion Chef™* sur ce paramètre est moins important, mais il change radicalement la gestion de la plateforme en termes de temps d'analyse et de rendement (Figure 10). Lors d'un chargement manuel, la plateforme ne peut séquencer qu'une seule puce par jour avec le matériel disponible. Avec la mise en place du *Ion Chef™*, deux puces de 10G peuvent être chargées simultanément, multipliant ainsi par deux le nombre patients séquencés par jour.

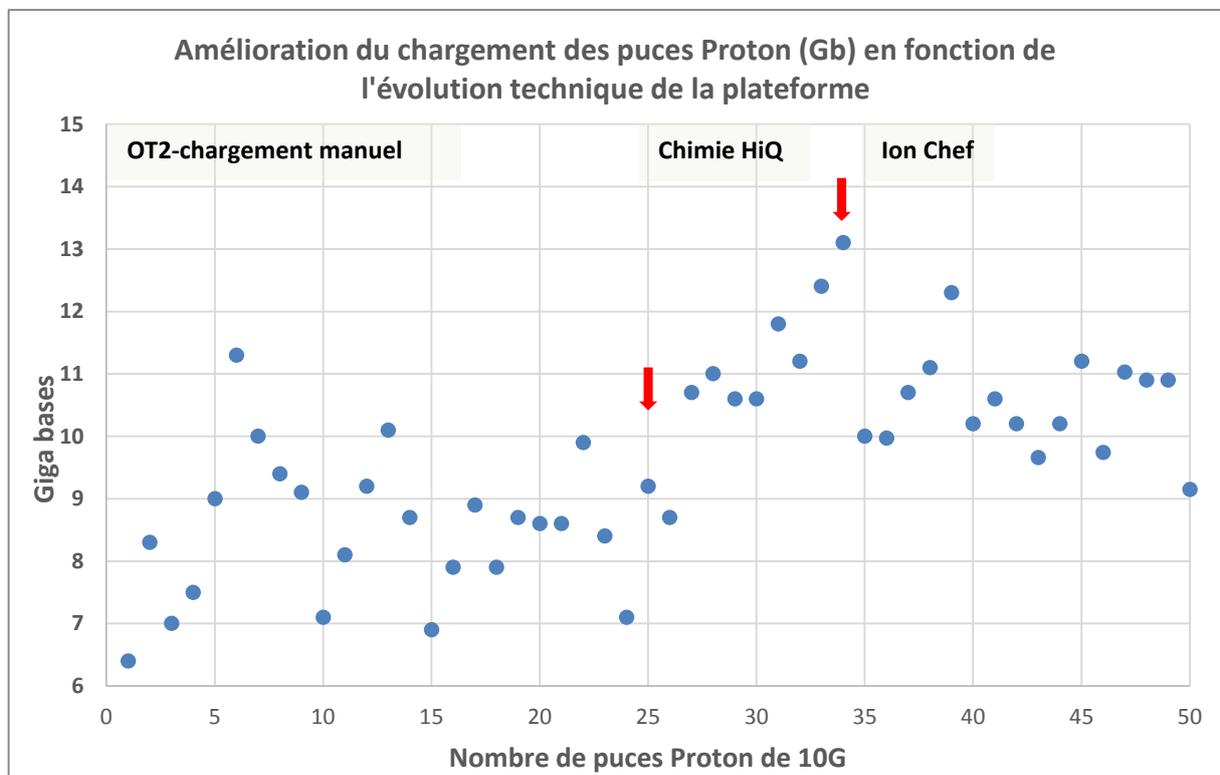


Figure 10 : Représentation graphique du chargement brut des puces en Gb d'ADN en fonction de la stratégie utilisée

Chaque point bleu représente le chargement total en Gb d'ADN pour chaque puce Proton PI. Les 25 premières puces ont été chargées manuellement après une PCR en émulsion par le système OT2/ES. La chimie de séquence *HiQ* a été introduite lors de la 26^{ème} puce, puis le système *Ion Chef™* après 8 puces supplémentaires.

1.3.3.3. Validation de l'évolution de la stratégie 2 vers la stratégie 2+

La stratégie 2 et la stratégie 2+ ont été comparées en évaluant les indicateurs de séquence mais aussi l'efficacité du rendement de diagnostic chez un patient sur une demande de diagnostic sur la liste ciblée sur les gènes impliqués dans la déficience intellectuelle (Tableau 5). Un premier séquençage par NGS avec la stratégie 2 a permis d'obtenir des indicateurs de qualité de séquence dans la moyenne avec un multiplexage en triplex (plus de 23 000 000 reads et une longueur moyenne de 113 pb). La couverture moyenne des gènes ciblés de la déficience intellectuelle (DI) à une profondeur de 20X est de 87,4%, avec environ 19% des exons des gènes d'intérêt couverts à moins de 90% (15%) ou pas couvert du tout (4%).

La même librairie de ce patient a été séquencée en l'utilisant comme témoin lors de la mise en route de la stratégie 2+ (Tableau 2). Il n'y a pas eu de multiplexage lors de cet essai, d'où le nombre plus important de lectures (plus de 79 000 000 reads) entraînant un biais dans l'amélioration des données de couverture. On remarquera cependant que la chimie *Hi-Q* couplée au système *Ion Chef*TM augmente la longueur moyenne des reads (ici de 10 bases), améliorant ainsi la couverture des gènes ciblés. En effet, plus de 97% des exons sont couverts à 100%, contre 73% avec la stratégie 2 (Tableau 5). De même, plus de 600 variants supplémentaires ont été détectés, soit 11 variants de plus sur la liste ciblée de gènes déficience intellectuelle. Mais aucun variant potentiellement pathogène sur cette liste n'a pu être mis en évidence ni avec l'une, ni avec l'autre de ces stratégies, laissant le patient sans diagnostic moléculaire sur la liste de gènes ciblés.

Tableau 5 : Comparaison des Indicateurs de séquence, des variants et des données de couverture entre les stratégies 2 et 2+ pour un même patient

Le patient a été adressé pour une étude ciblée des gènes impliqués dans la déficience intellectuelle. Le kit d'enrichissement large panel *Sure Select XT Inherited Disease* a été utilisé pour construire la librairie. Elle a été séquencée deux fois sur la même plateforme, mais en suivant des stratégies différentes. Malgré cette variabilité technique, nous pouvons comparer les indicateurs de séquences brutes tout comme les données de couvertures. L'analyse des variants est réalisée sur la liste ciblée de 92 gènes « diagnostic de la déficience intellectuelle » établie au sein du laboratoire. Les données de couverture sont à une profondeur de 20X minimum. Même si le multiplexage biaise les données brutes, l'amélioration des données de couverture n'a pas permis de mettre en évidence un variant pathogène avec la nouvelle chimie de séquence *Hi-Q* et un chargement de puce Proton 10G par le *Ion Chef™*.

	STRATEGIE 2 Chargement manuel - Chimie v3	STRATEGIE 2+ Chargement Ion Chef - Chimie <i>HiQ</i>
	Indicateurs de qualités du séquençage	
Bases totales (Gb)	7,0	10,0
AQ20 (Gb)	4,4	8,0
Nombre de lectures (<i>reads</i>)	23 269 313	79 259 867
Longueur moyenne de lecture (pb)	113	125
Multiplexage	OUI Triplex	NON Simplex
Nombre total de variants	8 098	8 740
	Analyse des variants	
Nombre de variants après filtre sur liste de gènes d'intérêt	70	81
Variant candidat	0	0
	Analyse de couverture à 20X	
Moyenne de couverture	87,4%	98,2%
% exons couvert à 100%	73%	97%
% exons couvert à > 90%	7%	1%
% exons couvert à ≤ 90%	15%	1%
% exons non couverts	4%	1%

1.3.3.4. Comparaison des données de couverture entre la stratégie 1 et 2+

Actuellement, c'est la stratégie 2+ qui est déployée au sein du laboratoire avec les dernières évolutions de la plateforme Ion Proton™. La stratégie 1 et la stratégie 2+ ont été comparées sur 4 listes de gènes étudiées dans l'article Lacoste et al. 2016. Les données de couverture à une profondeur de 20X sur la liste de gènes myopathies (82 gènes décrits), des neuropathies périphériques héréditaires (55 gènes décrits), des encéphalopathies épileptiques précoces (30 gènes décrits) et la liste des gènes « actionnables » (57 gènes) ont été générées. Il faut néanmoins noter que 13 gènes ne sont pas ciblés par le kit *Sure Select XT Inherited Diseases Panel* (Tableau 6).

Tableau 6: Résumé des données sur 4 listes de gènes ciblées par le Kit SureSelect XT Inherited Disease

On précise que 13 gènes au total sur ces 4 listes ne sont pas ciblés par le kit *Sure Select Inherited Disease*. Pour estimer au mieux les données de couverture, la taille des régions ciblées en pb a été calculée pour les 4 listes de gènes d'intérêt.

Groupe de pathologies	Myopathies	Neuropathies périphériques héréditaires sensibles et motrices	Encéphalopathies Épileptiques Précoces	Gènes actionnables
Nombre de gènes	82	55	30	57
Nombres d'exons	2125	918	513	1177
Taille des exons	383 631 pb	136 608 pb	81 274 pb	198 959 pb
Nombre de gènes non capturés par le kit SS-id	5	5	3	0

Les données générées ont été extraites à partir des 21 premiers patients séquencés avec la stratégie 2+ (multiplexage en triplex) et ont été comparées aux données publiées (Lacoste et al., 2016). Elles sont résumées dans le Tableau 7. Les gènes sont classés en 3 catégories (1- 100% -2 \geq à 90% et 3- < à 90%) comme décrit dans le matériel et méthode (PARTIE 2 Section 2-2-6). De plus, pour évaluer au mieux ces données de couverture à 20X, la taille de toutes les régions cibles des listes (correspondant à la taille de chaque exon) a été calculée pour évaluer plus précisément le pourcentage des exons couverts à 20X.

Les données de couverture à 20X sur les gènes sont bien meilleures avec le kit *Sure Select XT Inherited Diseases Panel* puisque en moyenne, il ne reste plus que 4 à 5 gènes dans le groupe des gènes couverts à moins de 90%, contre 24 avec la stratégie 1 sur la liste des myopathies par exemple. Même si le nombre de gènes couverts à 100% n'est pas fortement amélioré (aucun gène couvert à 100% sur les 21 premiers patients sur la liste des 30 gènes d'encéphalopathies épileptiques précoces par exemple), on améliore fortement la proportion d'exons couverts (et non plus de gènes) à plus de 99% à une profondeur de 20X (Tableau 7). La qualité de couverture est au moins doublée dans tous les cas : de 13% avec la stratégie 1 à 27% avec la stratégie 2+ sur cette même liste de gènes d'encéphalopathies épileptiques précoces. D'autre part avec cette nouvelle stratégie 2+, il est possible de séquencer trois patients simultanément (multiplexage en triplex) sur une puce Proton 10G, permettant ainsi une diminution des coûts du séquençage et une augmentation du rendement de la plateforme.

Ces résultats soulignent la qualité de séquençage obtenue pour la majorité des gènes d'intérêt avec la stratégie 2+, mais également l'importance de l'évaluation systématique et précise des données de couverture pour les gènes inclus dans des listes d'intérêt de pathologies génétiques hétérogènes. L'information de la couverture de séquence des gènes d'intérêt est transmise de manière synthétique aux prescripteurs dans le rapport d'analyse afin de leur communiquer l'information concernant les gènes insuffisamment couverts, pour lesquels ils pourront éventuellement demander une analyse complémentaire (séquençage Sanger complémentaire ou autre) en fonction du contexte clinique.

Tableau 7: Comparaison des données de couverture à une profondeur de 20X minimum sur 4 listes de gènes d'intérêt

Les données ont été calculées à partir de 31 patients séquencés avec la stratégie 1 et des 21 premiers patients séquencés avec la stratégie 2+. Les données d'exome sont issues d'échantillons séquencés sans multiplexage (simplex), contrairement aux échantillons séquencés avec la chimie HiQ qui ont été multiplexés par 2 (duplex), 3 (triplex) et même 4 (quadruplex). Ces essais de multiplexage ne diminuent en rien les données de couverture : si on analyse plus précisément le pourcentage d'exons couverts à au moins 99% (bandeau vert), on notera une amélioration nette des données de couverture à 20X avec la dernière stratégie 2+.

	Myopathies		Neuropathies périphériques héréditaires sensibles et motrices		Encéphalopathies Épileptiques Précoces		Gènes actionnables	
	STRATEGIE 1	STRATEGIE 2+	STRATEGIE 1	STRATEGIE 2+	STRATEGIE 1	STRATEGIE 2+	STRATEGIE 1	STRATEGIE 2+
	Nombre de GENES couverts à 20 X							
Gènes < à 90%	24	4	10	4	12	5	19	5
Gènes ≥ à 90%	55	59	41	41	17	22	37	48
Gènes à 100%	3	5	4	5	1	0	1	4
	Pourcentage de couverture à 20 X des EXONS							
Exons couverts < à 90%	29%	22%	18%	16%	40%	27%	33%	9%
Exons couverts ≥ à 90% et <100%	67%	72%	75%	75%	57%	73%	65%	84%
Exons couverts à 100%	4%	6%	7%	9%	3%	0%	2%	7%
Exons couverts ≥ 99%	21%	43%	27%	49%	13%	27%	7%	47%

1.3.3.5. Validation de la semi automatisation par le Robot Bravo™

La dernière évolution sur la plateforme est la semi automatisation par le Robot Bravo™ (*Agilent Technologies*) de la synthèse des librairies Sure Select XT™. Pour évaluer l'impact de cette semi-automatisation, nous avons séquencé 4 patients précédemment séquencés avec la stratégie 2 et dont le statut moléculaire a été clairement établi (Tableau 8). Les différences entre le nombre de *reads* sont dues aux multiplexages des patients par puce. Les autres indicateurs restent similaires par rapport à une synthèse manuelle des librairies. Depuis début 2016, l'automate Bravo™ permet de préparer au moins 24 librairies simultanément à partir de 24 ADN génomiques en 3 jours, sans modifier la qualité des résultats de séquençage.

Les premiers essais lors de la formation sur l'automate Bravo ont été initiés en avril 2015 avec la stratégie 2 sur seulement 8 patients. Pour des raisons administratives et logistiques indépendantes de notre volonté, la stratégie 2+ ainsi que l'utilisation du robot n'ont débuté qu'en mai 2016.

Tableau 8: Comparaison de la semi automatisation pour la synthèse des librairies avec le robot Bravo™ sur les indicateurs de séquence avec la stratégie 2

Quatre échantillons (11 à 14) ont été séquencés suivant la stratégie 2. Pour valider la mise en route de la semi automatisation du robot Bravo, ces quatre échantillons ont été de nouveau séquencés après avoir obtenus les librairies par le Bravo™. La différence des nombres de lecture est due au multiplexage lors du séquençage (le type de multiplexage par puce est indiqué sous le nombre de lectures). Néanmoins les analyses des variants et des données de couverture continuent à donner les résultats attendus. Le nombre de variants total reste dans la gamme de valeurs attendues (entre 8500 et 9000 variants avec la Stratégie 2).

	STRATEGIE 2					STRATEGIE 2 - semi automatisation avec Bravo™				
	Bases totales (Gb)	AQ20 (Gb)	Nombre de lectures (<i>reads</i>) multiplexage	Longueur moyenne de lecture (pb)	Nombre total de variants	Bases totales (Gb)	AQ20 (Gb)	Nombre de lectures (<i>reads</i>) multiplexage	Longueur moyenne de lecture (pb)	Nombre total de variants
Patient 11	9,8	7,0	78 313 510 Simplex	123	9 289	8,3	5,8	22 105 083 Triplex	120	8 846
Patient 12	8,6	6,9	66 982 300 Simplex	110	9 006	6,4	4,2	18 008 009 Triplex	116	8 482
Patient 13	12,0	8,6	29 640 792 Triplex	137	9 085	10,7	6,8	19 742 140 Quadruplex	104	8 528
Patient 14	9,8	6,6	31 993 767 Triplex	128	9 180	10,7	6,8	19 573 395 Quadruplex	122	8 728

Lors de la mise en route fin 2012 de la plateforme Ion Proton™, une évolution de la puce Proton PI était annoncée par le fournisseur Life Technologies – Thermo Fischer avec une augmentation annoncée de la capacité de chargement total de la puce de 100G. Cette nouvelle version de la Puce PII n'a jamais été commercialisée et a même été abandonnée par le fournisseur. Néanmoins, des évolutions ont été apportées et depuis 2016, la plateforme fonctionne donc avec le processus analytique suivant (Figure 11):

- Une semi automatisation de la synthèse des librairies par l'automate Bravo™ (Agilent) couplé au kit *Sure Select XT™ Inherited Disease* permettant sur 3 jours d'obtenir au moins 24 librairies de patients prêts au séquençage.
- Une chimie de séquence version *Hi-Q* favorisant la longueur des *reads* pour améliorer les données de couverture à une profondeur de 20X. Cette chimie permet aussi de multiplexer 3 patients par puce afin de diminuer les coûts.
- L'automatisation de la PCR en émulsion et du chargement de deux puces Proton 10G par jour par le système *Ion Chef™* permet de doubler le nombre de patients à séquencer dans un temps donné.



Figure 11: Processus analytiques depuis mai 2016

Représentation schématique du processus d'analyse mis en place sur la plateforme depuis mai 2016 depuis la réception du prélèvement jusqu'à la vérification par la méthode de Sanger des variants détectés par la plateforme Ion™ Proton du laboratoire.

1.4. PREMIERS RESULTATS DE LA MISE EN PLACE DU DIAGNOSTIC NGS AU LABORATOIRE

Avec la stratégie 2, environ 130 analyses sur des patients adressés pour un diagnostic au laboratoire ont été lancées. Depuis mars 2016, plus de 60 patients avec une analyse par la stratégie 2+ (Tableau 9) ont été faites.

Un résultat positif a été rendu pour 43 patients (22%) toutes pathologies confondues. Ce pourcentage varie de 2 à 44% selon les groupes de pathologie. (Tableau 9). La mise en évidence du variant a toujours été confirmée par séquençage Sanger et par une étude familiale de la ségrégation quand cela était possible. Quand le variant n'est pas connu, son caractère pathogène est prédit après interrogation des bases de données et de logiciels spécialisés (voir PARTIE 1 Section 1-5 et PARTIE 2 Section 2-2-5). Les variants considérés comme positifs et communiqués aux cliniciens font partie des catégories « pathogènes » (classe 5) et « probablement pathogènes » (classe 4) de la classification ESHG (Matthijs G et al., 2016). Les variants non classés (nommés VUS - *Variant of Unknown Signification*, classe 3) peuvent éventuellement être rapportés, mais cette dernière catégorie doit être clairement mentionnée.

Dans les cas où aucun variant non ambigu n'a été détecté, le biologiste propose au clinicien de faire une réévaluation clinique, en particulier pour estimer l'éventuelle implication de gènes couverts à moins de 90% et la nécessité de faire réaliser des analyses complémentaires. Les échanges entre biologistes et cliniciens sont ici essentiels.

Parallèlement à la mise en place technique de la plateforme, les biologistes du laboratoire ont élaboré les procédures pour la validation biologique et un rendu de résultat concis, homogène et complet aux cliniciens, en accord avec les recommandations de l'ANPGM (Association des Praticiens de Génétique Moléculaire : <http://www.anpgm.fr/index.php/recommandations-professionnelles/29-recommandations-professionnelles-anpgm>) et de l'ESHG (Matthijs G et al., 2016).

Tableau 9 : Bilan d'activité du rendu de diagnostic positifs confirmés sur la plateforme Ion Proton™

Le processus analytique est commun pour les 7 grands groupes de pathologies étudiées au laboratoire. L'analyse post séquençage étant longue, certains dossiers sont encore en cours d'analyse d'où la grande variation du nombre de positifs confirmés entre les indications. En complément des analyses de variants, les données de couvertures à une profondeur de 20X sont transmises aux cliniciens en classant les gènes en 3 catégories : gènes couverts à 100%, couverts à plus de 90% et couverts à moins de 90%. Lorsque le diagnostic met en évidence un variant candidat, le résultat est rendu aux cliniciens avec une demande de prélèvement des membres de la famille pour réaliser une étude de la ségrégation familiale du variant.

Nombre de patients	STRATEGIE 2 / STRATEGIE 2+	
	Séquencés	Positifs confirmés (% en fonction de l'indication)
Neuropathies Périphériques héréditaires sensitives et motrices	41	18 (41,8%)
Myopathies	45	1 (2,2%)
Dystonies	17	0
Épilepsies	51	19 (37,2%)
Déficiência Intellectuelle	18	3 (16,6%)
Surdités héréditaires	5	0
Laminopathies - Vieillessement Prématgré	19	2 (10,5%)
TOTAL	189	43 (22,7%)

Dans la suite de ce travail, trois cas dans lesquels l'analyse en haut débit a permis d'améliorer les connaissances soit par l'amélioration de la description clinique d'un syndrome, soit par la résolution d'un phénotype confondant, soit encore par la description d'un nouveau gène en pathologie sont présentés (Figure 12).

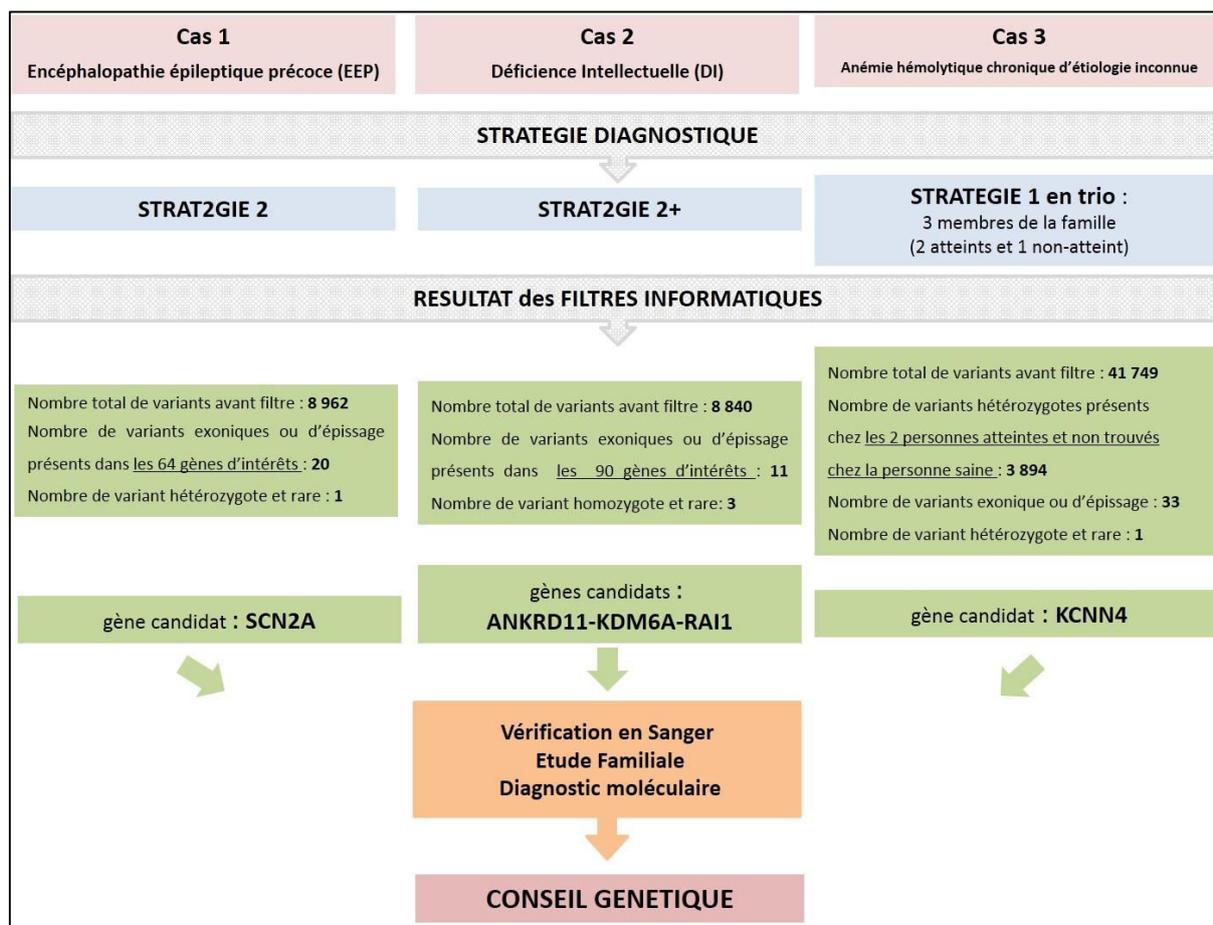


Figure 12 : Résumé des trois exemples de diagnostics posés par NGS.

La stratégie diagnostique est adaptée au cas par cas et dépend de la pathologie. Le Cas 1 concerne un fœtus interrompu en raison de convulsions *in utero*. Le Cas 2 décrit un enfant présentant un phénotype de déficience intellectuelle atypique. Le Cas 3 décrit un nouveau-né atteint d'une anémie d'étiologie inconnue. Pour ces 3 cas, la corrélation des données cliniques et phénotypiques, ainsi que l'étude de la ségrégation familiale ont permis de valider le variant candidat.

1.5. EXEMPLES DE DIAGNOSTICS POSES GRACE AU NGS

1.5.1. Amélioration de la description du phénotype d'encéphalopathie épileptique précoce

Une demande d'analyse d'un échantillon d'ADN fœtal a été adressée au laboratoire pour une étude par NGS des listes de gènes ciblés dans les encéphalopathies épileptiques précoces. La grossesse a été interrompue suite à des convulsions du fœtus survenues *in utero* à 30 semaines d'aménorrhée.

Les données cliniques mentionnaient, outre les convulsions, une ventriculomégalie et des membres mal positionnés. L'examen de fœto-pathologie rapportait une dysplasie dentato-olivaire, une malformation rare qui n'est pas détectable par les techniques d'imagerie.

Suite à l'analyse par NGS avec la stratégie 2, une mutation faux-sens hétérozygote dans le gène *SCN2A* a été mise en évidence. L'analyse familiale montrait que le variant était survenu *de novo*. Le gène *SCN2A* est connu pour être impliqué dans divers syndromes épileptiques, y compris des EEP, tels que le syndrome Ohtahara. Dans la littérature, on retrouvait un cas de mutation de *SCN2A* associée à une dysplasie dentato-olivaire diagnostiquée post-mortem chez un enfant présentant une encéphalopathie épileptique précoce (Howell KB et al., 2015). La fréquence de cette anomalie morphologique est probablement sous-estimée, car le diagnostic de dysplasie dentato-olivaire ne peut se faire que par un examen post-mortem. Le lien fonctionnel entre la survenue de cette anomalie et les mutations de *SCN2A* reste à déterminer.

Ce cas a fait l'objet d'un article soumis pour publication et en cours de révision. Le manuscrit soumis est présenté ci-après.

ARTICLE 2

***In utero* seizures revealing a dentato-olivary dysplasia
caused by *SCN2A* mutation**

***In utero* seizures revealing dentato-olivary dysplasia caused by *SCN2A*
mutation**

Running title: Antenatal seizures and *SCN2A* mutation

Fanny Sauvestre^{1,*}, Sébastien Moutton^{2,*}, Catherine Badens³, Bernard Broussin⁴, Dominique Carles¹, Nada Houcinat², Caroline Lacoste³, Florent Marguet⁵, Christophe Pecheux³, Laurent Villard³, Fanny Pelluard¹, Annie Laquerrière⁵, Gwenaëlle André¹

¹ Department of Pathology, Bordeaux University hospital, Bordeaux, France

² Department of Medical Genetic, Bordeaux University Hospital, Bordeaux, France

³ Department of Medical Genetic, APHM, Timone Hospital, Aix Marseille University, GMGF, Inserm UMR 910, Marseille

⁴ Radiology center, 120 bis rue Georges Bonnac, Bordeaux

⁵ Pathology Laboratory, Rouen University Hospital and Normandie University, UNIROUEN, NéoVasc, Rouen, France

* Both authors contributed equally to the work.

Corresponding author:

Fanny Sauvestre

Unité de Pathologie Fœtoplacentaire, CHU de Bordeaux, 33076 Bordeaux Cedex, France

Tel: +33 5 56 79 56 02, Fax: +33 5 56 79 55 93

fanny.sauvestre@u-bordeaux.fr

Key words: Dentato-olivary dysplasia, early-onset epileptic encephalopathy, *in utero* seizures, brain pathology, fetal epilepsy

- Number of text pages: 7

- Word count:

- Abstract: 148

- Short summary: 93

- Manuscript: 1654

- References: 15

- Figure: 1

- Supporting information: 1 table and 1 video

Abstract:

We report on a fetal case presenting with *in utero* seizures associated with ultrasonographically discovered ventriculomegaly, malposition of the upper and lower limbs detected at 30 gestational weeks, which led to the termination of the pregnancy two weeks later. Autopsy confirmed fetal akinesia sequence and revealed dentato-olivary dysplasia (DOD), a rare malformation that has never been antenatally reported. Next-generation sequencing (NGS) of a gene panel dedicated to epileptic syndromes allowed to identify a novel *de novo* *SCN2A* heterozygous missense mutation. Mutations in this gene have previously been implicated in diverse epileptic syndromes including early-onset epileptic encephalopathies (EOEE) such as Ohtahara syndrome, and have been reported in a context of DOD in only one post-natal case. DOD diagnosis is made on post-mortem neuropathological examination; therefore its incidence is probably underestimated. To improve our understanding of epileptogenesis and DOD physiopathology, neuropathological examination could be offered in EOEE with fatal outcome.

Key words:

Dentato-olivary dysplasia, early-onset epileptic encephalopathy, *in utero* seizures, brain pathology, fetal epilepsy

INTRODUCTION

Most early-onset epileptic encephalopathies (EOEE) are caused by genetic defects. In the past years, mutations in many genes in particular in genes encoding sodium channels have been identified by means of linkage studies, array-CGH and more recently next-generation sequencing^{1,2}. Mutations in *SCN2A* gene have been identified in a wide variety of early-onset epileptic syndromes, including benign familial neonatal infantile seizures (BFNIS)³, and more severe forms leading to encephalopathy such as Ohtahara and West syndromes⁴ as well as in autism spectrum disorders⁵.

Dentato-olivary dysplasia (DOD) was reported for the first time in 1991 by Harding and Boyd from five unrelated infants with intractable epilepsy⁶ but to date, the link between DOD and severe EOEE, especially Ohtahara syndrome, remains unclear. It is tempting to speculate that DOD is related to EOEE, but it is difficult to definitely establish the link between them⁵. DOD is a rare disease, even if its frequency is underestimated since this structural brain anomaly is usually not detected on brain MRI. This report aims at describing the first DOD fetal case revealed by *in utero* seizures and for whom sequencing of a gene panel dedicated to epileptic syndromes made it possible to identify a novel *de novo* heterozygous missense mutation in the *SCN2A* gene.

PATIENT AND METHODS

Case report

A gravida 1, para 1, 26-year-old woman underwent routine ultrasonography (US) at 20 weeks of gestation (WG) which was considered as normal. The pregnancy was uneventful until the third trimester, when routine US performed at 30 WG revealed severe symmetric bilateral ventriculomegaly, distal arthrogyriposis, hydramnios and abnormal movements. During subsequent follow-up, repetitive tonico-clonic movements were visualized, highly suggestive

of *in utero* seizures (Video S1). There was no personal or family medical history except for the maternal uncle who suffered from 4H syndrome (hypomyelinating leukodystrophy with hypo/oligodontia and hypogonadotropic hypogonadism), an autosomal recessive disorder whose clinical signs do not classically include epilepsy. Retrospectively, the mother reported occasional shaking movements considered as hiccups. Fetal karyotype analysis was normal (46, XX). A medical termination of the pregnancy was achieved at 32 WG in accordance with French law with the written consent of both parents.

Autopsy procedures

A complete autopsy was performed with the informed consent of both parents in accordance with the French law and following standardized protocols including X-rays, photographs, macroscopical and microscopical examination of all viscera. The brain was immersed in a 10% formalin-zinc buffer solution for one month. Seven-micrometer paraffin-embedded sections were processed from multiple brain areas, and stained with hematoxylin-eosin.

Molecular analyses

Molecular analyses were carried out in agreement with our institutional recommendations. Fetal DNA was extracted from thymus and parental DNA from blood samples according to standardized protocols. Fetal DNA was first analysed. Targeted regions were enriched using SureSelect XT Inherited Diseases Panel (Agilent Technologies) and sequencing was performed by means of an Ion Proton platform (Thermo Fisher – Life Technologies). Sequences were aligned to the hg19 genome reference and variant calling was performed targeting 64 genes involved in early-onset epilepsy, with the Torrent Suite using low stringency filters for germline disorders (`germline_low_stringency_targetseq`) (Thermo Fisher – Life Technologies). Candidate variants were confirmed by Sanger sequencing and parental DNA analysis was performed to study the mode of inheritance.

RESULTS

Autopsy findings

External examination confirmed fetal akinesia deformation sequence with non-specific dysmorphism and limb deformities. Brain weight was normal for the term, there were no gyral anomalies and the brainstem and cerebellum appeared macroscopically normal in size and shape. On coronal sections, the corpus callosum was thin in its posterior part, with mild ventricular dilatation. On infratentorial sections, the dentate and olivary nuclei were poorly convoluted. Histologically, the cytoarchitecture of the cerebellar cortex was normal, but the dentate nuclei were dysmorphic, forming a single C-shaped mass of neurons (Figure 1A) compared with an age matched control (Figure 1B). In the medulla, olivary nuclei were coarse and almost unconvoluted (Figure 1C and D for comparison with a control case at the same developmental age), but no fragmentation or associated ectopias were present. Other infratentorial anatomic structures were free of lesions. In the cerebral hemispheres, nodular or linear heterotopic calcified clusters of neurons were observed in the intermediate zone along with dispersed calcified neurons within the striatum.

Gene panel sequencing

A missense mutation chr2(GRCh37):g.166237627A>G /NM_001040142(SCN2A):c.4471A>G was identified in exon 25 of *SCN2A* gene, resulting in the substitution p.Thr1491Ala, located in the cytoplasmic domain connecting S6 of domain III and S1 of domain IV. Even though mild physico-chemical differences exist between these two amino acids, the substitution occurs in a highly conserved region among species, and is predicted to be deleterious according to SIFT (Deleterious; score: 0) and MutationTaster (disease causing; p-value: 1), but benign to Polyphen2 (0.011). Splicing scores were predicted to be unaffected. No other anomaly could be identified based on standard variant

prioritization procedures (Table S1). This most likely pathogenic variant was confirmed by Sanger sequencing and was absent in the parents, indicating its *de novo* occurrence.

DISCUSSION

Apart from X-LAG syndrome and pyridoxine dependent epilepsy, together with seizures and combining various structural malformations, antenatal seizures have very unfrequently been reported in other pathological conditions and only in case of cerebral infarcts, haemorrhages or trauma, congenital infections, amino and organic acid disorders, and maternal drug withdrawal. Regardless of the etiology, prognosis of fetal seizures is consistently very poor^{7, 8}. Although pregnant women sometimes report abnormal fetal movements, fetal seizures most often remain misdiagnosed^{7, 8}, though they may represent the most severe end of the spectrum in several epileptic syndromes. In the present observation, antenatal onset epilepsy might have been related to the severe form of Ohtahara syndrome (or other unspecified EOEE). Ohtahara syndrome is one of the earliest epileptic syndromes leading to severe developmental delay, usually manifesting in the first days of life with most often intractable early-onset tonic seizures and suppression-burst patterns, enabling its distinction from other early infantile encephalopathies⁹.

DOD is a rare structural brain malformation, recognized at neuropathological examination, and linked to EOEE^{5, 6, 10}, though not systematically observed (one patient (patient three) from the series of Howell *et al*)⁸. Only few postnatal DOD cases have been described in the literature. On imaging, abnormal signals have been described and interpreted as resulting from cytotoxic oedema. Corpus callosum hypoplasia, ventricular enlargement, white matter atrophy have also been reported¹¹.

On post-mortem brain examination, dentate and olivary nucleus dysplasia, two functionally interconnected structures, affect one or both structures and may occur in isolation or with

other brain abnormalities such as cerebellar hypoplasia, microcephaly or corpus callosum agenesis. Besides, DOD may be observed in several other conditions such as Zellweger syndrome or thanatophoric dysplasia, lissencephalies, Joubert and related disorders and ponto-cerebellar hypoplasia¹². These entities are generally ruled out based on other associated specific brain and sometimes visceral abnormalities¹³. Microscopically, DOD, considered to be due to a maturational arrest¹², is characterized by unfolded dentate and olivary nuclei, forming an oval or a C-shaped mass made of interconnected islands of neurons interspersed with myelin^{6, 10}. Hyperconvoluted dentate and olivary nuclei have also been described. All these features allow to differentiate DOD from ponto-cerebellar hypoplasia^{5, 12}.

Mutations in the *SCN2A* gene have been identified in different forms of infantile epilepsy including benign familial neonatal infantile seizures (BFNIS)³, other early-onset epileptic syndromes associated with severe developmental delay², and in EOEE such as Ohtahara or West syndromes⁴. In a recent report by Howell et al.⁸, *SCN2A* has been shown to represent the second most common cause of epilepsy of infancy with migrating focal seizures.

The *SCN2A* gene encodes the major α subunit of voltage-gated sodium channels Na_v1.2 in excitatory neurons¹⁴. The vast majority of *SNC2A* alterations are missense mutations, located in diverse domains of the protein. Although there are no obvious genotype-phenotype correlations, mutations responsible for severe forms frequently arise *de novo*, conversely to benign phenotypes usually inherited in an autosomal dominant trait. In our patient, the mutation was located near a previously reported substitution affecting the last amino acid of S6 of domain III (NM_001040142.1 (*SCN2A*):c.4419A>G ;p.Ile1473Met) which has been shown to alter the voltage-dependence of Na_v1.2 channel conductance¹⁴. *De novo* occurrence and *in silico* predictions led us to propose that the variation is pathogenic. From the literature, the molecular diagnosis of DOD was made in a single patient using whole genome sequencing. The patient carried a *SCN2A* missense mutation affecting another domain

(p.Ala263Val)⁵. Pangenomic analyses including array-CGH and whole exome/genome sequencing make it possible to identify numerous disease-causing genes. Meanwhile, molecular diagnostic laboratories have taken advantage of next-generation sequencing to design gene panels in order to improve mutational detection rates and diagnosis delay for various heterogeneous genetic diseases, particularly epileptic syndromes¹⁵.

In conclusion, we report for the first time the earliest and most severe form of EOEE due to *SNC2A* mutation with antenatal *in utero* presentation, highlighting the wide clinical spectrum of the disease. Our case is the second to be reported in relation to DOD, whose link with epileptogenesis still remains unclear. Due to the severity of postnatal outcome in the majority of cases, the diagnosis must be established before birth whenever possible. Post-mortem examination is an essential step to guide molecular investigations for confirmation of the disease. Nevertheless, additional patients should be investigated to know if mutations in other genes could also be associated with DOD. Therefore, in EOEE patients, the combination of array-CGH, gene panel or whole-exome/genome sequencing¹ and neuropathological examination in case of fatal outcome¹² will undoubtedly provide interesting data for improving genotype-phenotype correlations.

Acknowledgements

The authors thank the parents for their participation to this study.

Disclosure of Conflicts of Interest

None of the authors has any conflict of interest to disclose.

Ethical Publication Statement

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Figure Legends

Figure 1: Neuropathology of **dentato-olivary dysplasia**

A- Dysmorphic dentate nuclei with loss of their normal undulated shape (arrow) [H&E,

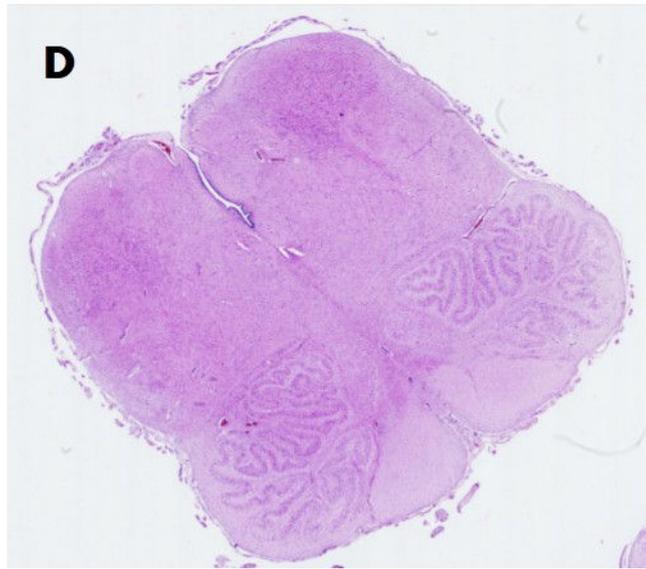
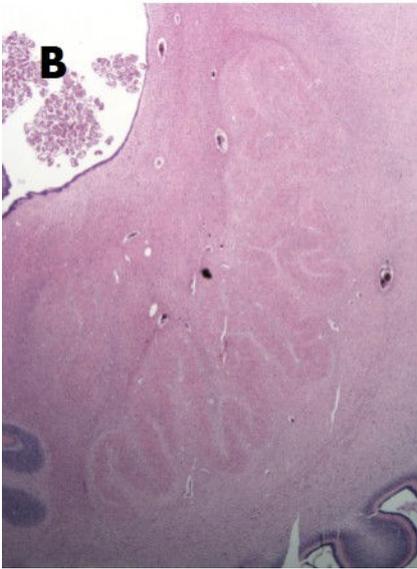
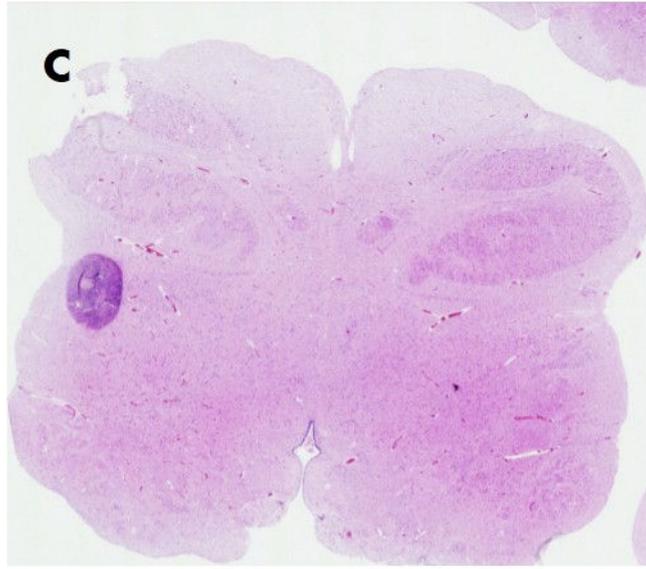
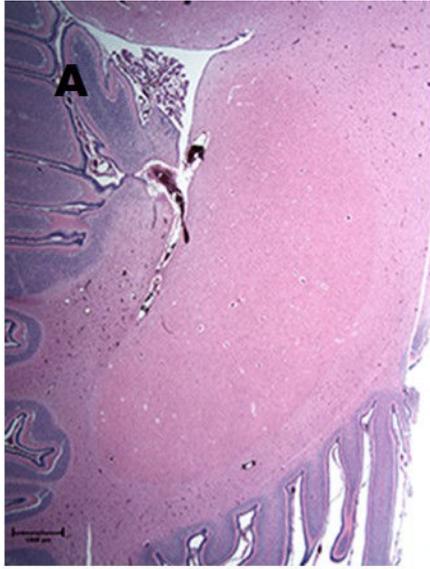
OM x 40]

B- Compared with normal dentate nucleus at 32 WG

C- Poorly convoluted and coarse olivary nuclei (arrow) [H&E, OM x 25]

D- Normal inferior olivary nucleus at the same term (scale bar = 1mm).

(H&E: Haematoxylin eosin staining, OM: original magnification)



REFERENCES

1. Ream MA, Patel AD. Obtaining genetic testing in pediatric epilepsy. *Epilepsia*. 2015;56:1505-14.
2. McTague A, Howell KB, Cross JH, et al. The genetic landscape of the epileptic encephalopathies of infancy and childhood. *The Lancet Neurology*. 2016;15:304-16.
3. Heron SE, Crossland KM, Andermann E, et al. Sodium-channel defects in benign familial neonatal-infantile seizures. *Lancet*. 2002;360:851-2.
4. Nakamura K, Kato M, Osaka H, et al. Clinical spectrum of SCN2A mutations expanding to Ohtahara syndrome. *Neurology*. 2013;81:992-8.
5. Touma M, Joshi M, Connolly MC, et al. Whole genome sequencing identifies SCN2A mutation in monozygotic twins with Ohtahara syndrome and unique neuropathologic findings. *Epilepsia*. 2013;54:e81-5.
6. Harding BN, Boyd SG. Intractable seizures from infancy can be associated with dentato-olivary dysplasia. *Journal of the neurological sciences*. 1991;104:157-65.
7. Usta IM, Adra AM, Nassar AH. Ultrasonographic diagnosis of fetal seizures: a case report and review of the literature. *BJOG: an international journal of obstetrics and gynaecology*. 2007;114:1031-3.
8. Howell KB, McMahon JM, Carvill GL, et al. SCN2A encephalopathy: A major cause of epilepsy of infancy with migrating focal seizures. *Neurology*. 2015;85:958-66.
9. Ohtahara S, Ohtsuka Y, Yamatogi Y, et al. The early-infantile epileptic encephalopathy with suppression-burst: developmental aspects. *Brain & development*. 1987;9:371-6.
10. Trinkaus E, Rauscher C, Nagler M, et al. A case of Ohtahara syndrome with olivary-dentate dysplasia and agenesis of mamillary bodies. *Epilepsia*. 2001;42:950-3.

11. Raspall M, Ortega-Aznar A, del Toro M, et al. Neonatal rigid-akinetic syndrome and dentato-olivary dysplasia. *Pediatric neurology*. 2006;34:132-4.
12. Barth PG. Cerebellar dentate dysplasia. *Brain & development*. 2011;33:621-6.
13. Golden JA HB. *Developmental Neuropathology*. First ed: Wiley-Blackwell; 2006. 388p.
14. Oliva M, Berkovic SF, Petrou S. Sodium channels and the neurobiology of epilepsy. *Epilepsia*. 2012;53:1849-59.
15. Kodera H, Kato M, Nord AS, et al. Targeted capture and sequencing for detection of mutations causing early onset epileptic encephalopathy. *Epilepsia*. 2013;54:1262-9.

1.5.2. Résolution d'un cas de déficience intellectuelle présentant un phénotype confondant.

Un patient hypotonique avec un retard psychomoteur important associé à des troubles du comportement et une dysmorphie atypique a été adressé pour étudier par NGS les gènes impliqués dans la déficience intellectuelle.

L'ADN du patient a été traité avec la stratégie 2+. Les données de séquences ont été filtrées sur une liste de 70 gènes ciblés (90 transcrits) pour la déficience intellectuelle. Après filtrage informatique sur le logiciel d'analyse VarAFT[®], trois variants dans les gènes *RAI1*, *ANKRD11* et *KDM6A* ont été retenus comme potentiellement pathogènes. Ils ont tous été confirmés en Sanger. Après concertation avec le clinicien sur la concordance avec le phénotype, 2 des gènes paraissaient pertinents : *ANKRD11* et *KDM6A*. En effet, le père et le frère aîné du cas index présentaient un phénotype discret de déficience intellectuelle compatible avec un tableau de syndrome KBG, affection rare à transmission dominante due à des mutations du gène *ANKRD11* (Goldenberg A et al. 2016). Par séquençage Sanger, il était montré que tous 2 étaient porteurs de la variation nucléotidique dans le gène *ANKRD11*. Par ailleurs, la mère du cas index était porteuse hétérozygote de la variation dans le gène *KDM6A* (situé sur le chromosome X) et présentait un biais total d'inactivation du chromosome X. Elle ne présentait aucun trait phénotypique. Le gène *KDM6A* est un des gènes responsables du syndrome de Kabuki. Le cas index est donc porteur de 2 variations nucléotidiques dans 2 gènes différents tous deux impliqués dans 2 syndromes différents (Figure 13). Son phénotype atypique résulte donc vraisemblablement de la superposition de ces 2 défauts génétiques. Le variant dans *RAI1* a été exclu suite à l'analyse de ségrégation familiale et à une absence de concordance avec les signes cliniques.

Un oncle maternel non testé mais ayant eu il y a quelques années une consultation en génétique pour un tableau de déficience intellectuelle compatible avec un syndrome de Kabuki est probablement porteur hémizygotique de la variation dans le gène *KDM6A*. Par la technique de NGS, nous avons donc pu mettre en évidence un génotype pathologique impliquant deux gènes donnant un phénotype confondant. Un cas a été déjà décrit dans la littérature pour un phénotype similaire impliquant le gène *ANKRD11* et le gène *KDM1A*, autre gène connu pour être responsable dans le syndrome de Kabuki (Tunovic S et al. 2014). Dans ce type de cas, les échanges avec le clinicien sont essentiels pour statuer sur les variants candidats mis en évidence par le NGS.

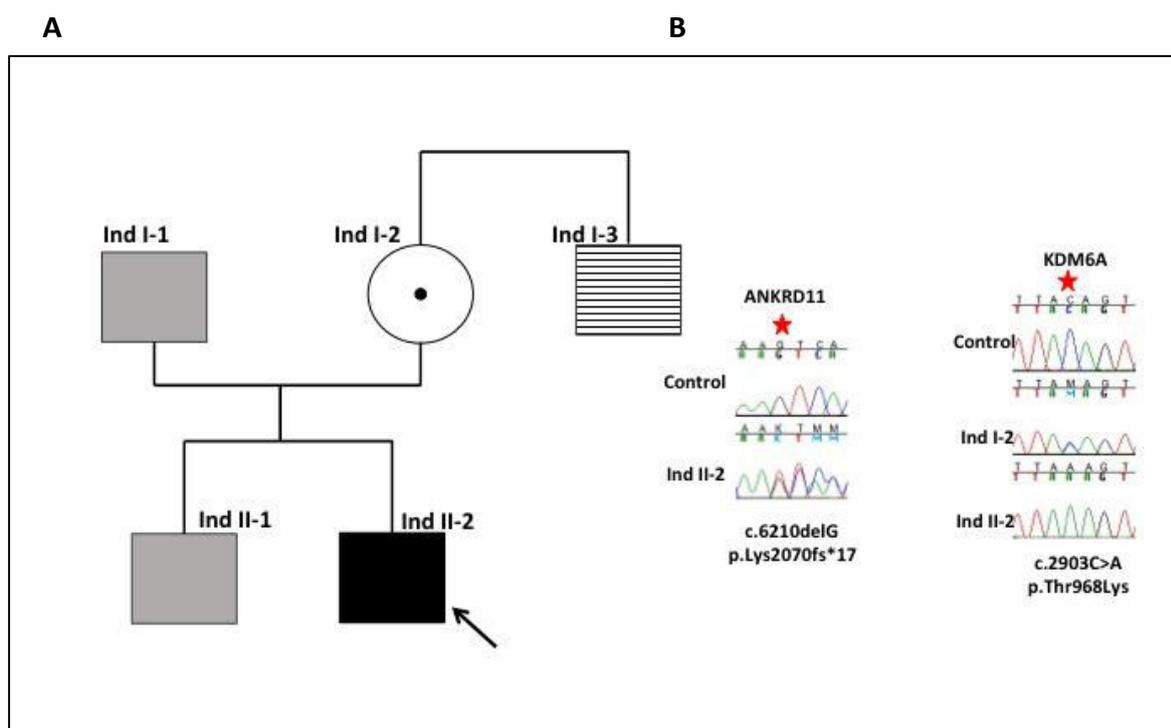


Figure 13 : Arbre généalogique et données du séquençage en Sanger de la famille du patient atteint de déficience intellectuelle

A : L'individu Ind II-2, cas index, a bénéficié d'une analyse en NGS avec un filtre sur les gènes ciblés pour la déficience intellectuelle. Il présente un phénotype atypique par rapport à son frère. Seul l'individu Ind I-3 (rayé gris) n'a pu être testé en Sanger pour les deux variants, mais il présente un phénotype compatible avec le syndrome de Kabuki. La mère asymptotique du cas index (Ind I-2) est porteuse de la variation dans le gène *KDM6A* (NM_021140 chrX (*hg19*) :44 937 715 C>A, c.2903 C>A : p.Thr968Lys), mais elle est associée à un biais total d'inactivation de l'un de ses chromosomes X. Les patients en gris clair (le père ind I-1 et le frère du cas index ind II-1) présentent un syndrome KBG compatible avec la présence du variant dans le gène *ANKRD11* (NM_013275 Chr16 (*hg19*): 89 346 740 delC, c.6210delG: p.Lys2070fs*17)

B : Les électrophérogrammes des séquences nucléotidiques sont présentés pour les contrôles (séquence supérieure) et pour les individus II-2 et I-2. Les étoiles rouges indiquent les positions des mutations sur la séquence de référence dans le gène *ANKRD11* et le gène *KDM6A*. Le cas index (Ind II-2) présente les deux variations nucléotidiques.

1.5.3. Caractérisation d'un nouveau gène responsable d'une forme d'anémie hémolytique héréditaire avec globules rouges déshydratés.

Le canal Gardos est un canal potassium spécifique, activé par le calcium. Il est exprimé dans plusieurs tissus et cellules, y compris les érythrocytes où il est impliqué dans la régulation du volume cellulaire. Nous rapportons ici le cas de deux familles non apparentées, chez qui l'analyse d'exome en NGS a été réalisée pour une indication d'anémie hémolytique sans étiologie connue. L'analyse d'exome a permis l'identification d'une mutation héréditaire dominante située dans le gène *KCNN4* codant pour la sous unité du canal Gardos. Les personnes atteintes présentaient une anémie chronique à globules rouges déshydratés appelée Xérocytose Héréditaire (HX). Leurs globules rouges présentent diverses anomalies de forme. La mutation faux-sens mise en évidence par NGS et confirmée par la méthode de Sanger se situe sur un résidu hautement conservé lors de l'évolution au niveau de la région d'interaction avec la calmoduline et responsable de l'ouverture du canal et de l'efflux de K⁺. Nous avons vérifié par l'étude des transcrits que cette mutation n'entraînait pas d'anomalie d'épissage et par Western blot, pas d'anomalie du niveau d'expression de la chaîne protéique mutée.

Le canal Gardos n'avait jamais été impliqué dans une pathologie humaine auparavant, mais il y avait de forts arguments génétiques et cliniques pour penser que la mutation caractérisée était bien à l'origine du phénotype. Nous avons donc établi une collaboration avec une équipe de recherche afin de réaliser des études fonctionnelles et démontrer la pathogénicité du mutant.

Grâce à une collaboration avec l'équipe d'Hélène Guizouarn à Nice (UMR_7277, institut de Biologie, Valrose), il a été montré par des expériences d'électrophysiologie sur des ovocytes de Xénope et sur des cellules humaines de la lignée HEK293, que le canal muté présente une activité plus élevée et prolongée par rapport au canal de type sauvage. L'identification d'une mutation dans le gène *KCNN4* associée à une hémolyse chronique constitue le premier rapport d'une maladie humaine causée par un défaut du canal Gardos.

L'élucidation des bases moléculaires de la HX ouvre des perspectives thérapeutiques intéressantes. En effet, le canal muté reste sensible à l'inhibition par des inhibiteurs spécifiques, suggérant que le traitement de ce type de HX par un inhibiteur spécifique du canal Gardos pourrait être envisagé. Des expériences réalisées *in vitro*, sur des cellules HEK transfectées, mais aussi sur des globules rouges de patients mutés, montrent que les

inhibiteurs agissent efficacement *in vitro* pour diminuer la déshydratation et que l'efficacité de l'inhibiteur dépend de la mutation en cause.

Par ailleurs, plusieurs descriptions d'accidents thrombotiques sévères chez des patients atteints de HX et ayant été splénectomisés, ont conduit à la contre-indication de la splénectomie en cas de HX. Toutefois, ce risque accru de thrombose n'est documenté que chez des patients présentant des mutations de *PIEZO1* (l'autre gène responsable de XH) ou dans des articles antérieurs à la description des causes moléculaires. Dans 2 des familles rapportées avec mutations du canal Gardos, aucun accident thrombotique n'a été observé chez 3 patients splénectomisés avec un recul d'une quinzaine d'années. Il est donc possible que ce sur-risque soit dû non pas à la déshydratation des GR en elle-même, mais plutôt à une anomalie en lien avec *PIEZO1*. D'autres descriptions de cas de HX dues à des mutations du canal Gardos permettront de confirmer ce point et donc de limiter éventuellement la contre-indication de la splénectomie aux formes mutées sur le canal *PIEZO1*.

On voit donc dans cet exemple que la description du gène et de la mutation causale caractérisée en NGS permet d'une part d'envisager un nouveau traitement pharmacologique, et d'autre part de préciser les conditions de contre-indication de la splénectomie, seul traitement disponible jusqu'à présent, deux points qui vont contribuer de façon significative à l'amélioration de la prise en charge des patients atteints

Ce travail collaboratif a fait l'objet d'une publication dans *Blood* présentée ci-après.

ARTICLE 3

A mutation in the Gardos channel is associated with hereditary xerocytosis

Plenary Paper

RED CELLS, IRON, AND ERYTHROPOIESIS

A mutation in the Gardos channel is associated with hereditary xerocytosis

Raphael Rapetti-Mauss,¹ Caroline Lacoste,^{2,3} Véronique Picard,^{4,5} Corinne Guitton,⁶ Elise Lombard,^{2,7} Marie Loosveld,⁸ Vanessa Nivaggioni,⁸ Nathalie Dasilva,² David Salgado,² Jean-Pierre Desvignes,² Christophe Bérout,^{2,3} Patrick Viout,⁹ Monique Bernard,⁹ Olivier Soriani,¹ Henri Vinti,¹⁰ Valérie Lacroze,¹¹ Madeleine Feneant-Thibault,¹² Isabelle Thuret,¹³ Hélène Guizouarn,¹ and Catherine Badens^{2,3,7}

¹Institut de Biologie Valrose, Unité Mixte de Recherche 7277, Centre National de Recherche Scientifique, Université de Nice–Sophia Antipolis, INSERM U1091, Nice, France; ²Aix-Marseille Université, INSERM, Unité Mixte de Recherche S910, Marseille, France; ³Assistance Publique–Hôpitaux de Marseille, Département de Génétique Médicale, Hôpital d'Enfants de la Timone, Marseille, France; ⁴Assistance Publique–Hôpitaux de Paris, Service d'Hématologie Biologique, Centre de Référence des Maladies du Globule Rouge et de l'Erythropoïèse Hôpital Bicêtre, Le Kremlin-Bicêtre, France; ⁵Laboratoire d'Hématologie, Université Paris Sud, Faculté de Pharmacie, Chatenay-Malabry, France; ⁶Assistance Publique–Hôpitaux de Paris, Service de Pédiatrie, Centre de Référence des Maladies du Globule Rouge et de l'Erythropoïèse Hôpital Bicêtre, Le Kremlin-Bicêtre, France; ⁷Assistance Publique–Hôpitaux de Marseille, Laboratoire de Biochimie, Hôpital Conception, Marseille, France; ⁸Assistance Publique–Hôpitaux de Marseille, Laboratoire d'Hématologie, Hôpital de la Timone, Marseille, France; ⁹Aix-Marseille Université, Centre National de Recherche Scientifique, Centre de Résonance Magnétique Biologique et Médicale, Unité Mixte de Recherche 7339, Marseille, France; ¹⁰Service d'Hématologie Clinique, Hôpital de l'Archet, Nice, France; ¹¹Assistance Publique–Hôpitaux de Marseille, Service de Néonatalogie, Hôpital d'Enfants de la Timone, Marseille, France; ¹²Assistance Publique–Hôpitaux de Paris, Service de Biochimie, Centre de Référence des Maladies du Globule Rouge et de l'Erythropoïèse Hôpital Bicêtre, Le Kremlin-Bicêtre, France; ¹³Assistance Publique–Hôpitaux de Marseille, Centre de Référence des Thalassémies, Service d'Hématologie Pédiatrique, Hôpital d'Enfants de la Timone, Marseille, France

Key Points

- The Gardos channel is a potassium channel involved in red cell volume modification.
- A mutation in *KCNN4* encoding the Gardos channel is presented as the genetic basis for a new type of hereditary xerocytosis.

The Gardos channel is a Ca^{2+} -sensitive, intermediate conductance, potassium selective channel expressed in several tissues including erythrocytes and pancreas. In normal erythrocytes, it is involved in cell volume modification. Here, we report the identification of a dominantly inherited mutation in the Gardos channel in 2 unrelated families and its association with chronic hemolysis and dehydrated cells, also referred to as hereditary xerocytosis (HX). The affected individuals present chronic anemia that varies in severity. Their red cells exhibit a panel of various shape abnormalities such as elliptocytes, hemighosts, schizocytes, and very rare stomatocytic cells. The missense mutation concerns a highly conserved residue among species, located in the region interacting with Calmodulin and responsible for the channel opening and the K^+ efflux. Using 2-microelectrode experiments on *Xenopus* oocytes and patch-clamp electrophysiology on HEK293 cells, we demonstrated that the mutated channel exhibits a higher activity and a higher Ca^{2+} sensitivity compared with the wild-type (WT) channel. The mutated channel remains sensitive to inhibition suggesting that treatment of this type of HX by a specific inhibitor of the Gardos channel could be considered. The identification of a *KCNN4* mutation associated with chronic hemolysis constitutes the first report of a human disease caused by a defect of the Gardos channel. (*Blood*. 2015;126(11):1273-1280)

Introduction

Hereditary xerocytosis (HX) (Online Mendelian Inheritance in Man 194380) is an autosomal dominant congenital hemolytic anemia characterized by primary erythrocyte dehydration.¹ In HX patients, red blood cells exhibit altered intracellular cation content and cellular dehydration, which are responsible for an increased erythrocyte mean corpuscular hemoglobin concentration (MCHC) and a decreased erythrocyte osmotic fragility.² Under the microscope, blood films show various cell shape abnormalities, the most characteristic being a central pallor, straight or crescent shaped. These cells are called stomatocytes, and HX is also known as dehydrated hereditary stomatocytosis.³

HX has been associated with missense mutations in *FAM38A* encoding the red cell membrane mechanosensitive cation channel

PIEZO1.^{4,5} Functional studies have demonstrated that in PIEZO1, mutations slowed channel inactivation and introduced a pronounced latency for activation.⁶ More recently, another type of red cell ion exchange defect associated with pseudohyperkalemia has been linked to mutations in the ATP-binding cassette transporter ABCB6.⁷

The Gardos channel is a cation channel also referred to as KCa3.1 or *KCNN4*. It is a Ca^{2+} -sensitive, intermediate conductance, potassium selective channel, initially described in pancreas cells but present in many cell types including erythrocytes.⁸ A number of recent studies have described its role in a variety of physiologic events and identified it as an interesting therapeutic target in a large panel of human diseases.^{9,10} We report the first identification of a missense mutation

Submitted April 29, 2015; accepted June 28, 2015. Prepublished online as *Blood* First Edition paper, July 6, 2015; DOI 10.1182/blood-2015-04-642496.

H.G. and C. Badens contributed equally to this work.

There is an Inside *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2015 by The American Society of Hematology

located in one of the functional regions of the Gardos channel and its association with chronic hemolysis and dehydrated cells in 2 unrelated families, with 8 affected persons.

Material and methods

Hematologic tests

An osmotic fragility test in hypotonic saline solutions was performed on red blood cells immediately after sampling and after 24-hour incubation at 4° or 37°C.

Nuclear magnetic resonance

Nuclear magnetic resonance experiments were performed on a 400 AVANCE wide-bore spectrometer (Bruker Biospin, Billerica, MA), using stimulation by the ionophore A23187 (Sigma Aldrich).

Next-generation sequencing

Exome sequencing was performed after exome enrichment using Ion AmpliSeq (Thermo Fisher Scientific, Waltham, MA), template preparation using the Ion PI Template OT2 200 Kit v2 on the Ion OneTouch 2 System, and sequencing using the Ion PI Chip Kit v2 and Ion PI Sequencing 200 Kit v2 on the Ion Proton Sequencer (Thermo Fisher Scientific). Raw data were first aligned with the provided software suite to generate BAM files. The coverage and sequencing depth analysis were computed using the BEDtools suite v2.17¹¹ and in-house scripts. Variants were identified using the Torrent Browser Variant caller (version 4.0.2), annotated, and prioritized with the in-house VarAFT system that includes Annovar.¹²

The mutation was confirmed on DNA samples and *KCNN4* transcripts from fresh reticulocytes by Sanger sequencing (3500XL Genetic Analyzer; Life Technologies, Carlsbad, CA).

Expression in *Xenopus* oocytes

Plasmid pcDNA3KCNN4-HA was used to introduce the point mutation p.Arg352His by polymerase chain reaction. A hemagglutinin tag (HA) was present in the C-terminal end of KCNN4.¹³ Female *Xenopus laevis* were anesthetized with MS222 according to the procedure recommended by our ethics committee. Oocytes were harvested and injected as previously published.¹⁴

Current recording was performed as follows: a ramp protocol between –120 and +80 mV for 2 seconds, holding potential at –80 mV, was applied using Clampex (PCLamp; Molecular Devices Corporation). To avoid looking at chloride channel activation, currents were recorded in modified Barth's saline (MBS), where chloride was substituted by gluconate (Na-gluconate, 85 mM; K-gluconate, 1 mM). Junction potential was minimized using an agar bridge and 3 M KCl. Electrodes filled with 3 M KCl had a resistance of 0.5 MOhm. After equilibration in this gluconate MBS, KCNN4 was activated by the calcium ionophore A23187 at 1 μM in MBS gluconate. In control oocytes, no current was activated by ionophore addition.

Western blotting on oocyte

Oocyte membranes were prepared as previously described.¹⁵ Immunodetection of KCNN4-HA was done using an anti-HA antibody (1/1000; Sigma). To compare KCNN4 expression levels in different samples, the cell membrane marker β1 Na,K-ATPase was used (1/500; Sigma). Signals were detected by chemiluminescent reaction with Immobilon Western reagent (Millipore) and a Fusion FX7 (Vilber-Lourmat). The intensity of KCNN4 bands relative to the β1 Na,K-ATPase signal was quantified using ImageJ Version 1.44 software (National Center for Biotechnology Information).

HEK293 cell transfection

HEK293 cells were grown in Dulbecco's modified Eagle medium glutamax (Gibco), 10% fetal bovine serum, and penicillin-streptomycin. Cells were cotransfected with 1 μg WT or point mutated pcDNA3-KCNN4-HA and 0.5 μg plasmid internal ribosome entry site with enhanced yellow fluorescent protein

using CaPO₄. Sixteen hours later, cells were washed twice with phosphate-buffered saline, and patch clamp was performed on fluorescently labeled cells.

Patch-clamp electrophysiology

Glass pipettes (Brand, Wertheim, Germany) were made on a horizontal pipette puller (P-97; Sutter Instrument, Navato, CA) to give a final resistance ranging from 3 to 5 MΩ. For whole cell experiments, the bath solution was (in mM) NaCl 140, KCl 5, CaCl₂ 1, glucose 29, and HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) 25, pH 7.4, adjusted with NaOH. The intracellular solution was (in mM) KCl 30, KGluconate 100, EGTA 5, HEPES 10, pH 7.2, adjusted with NaOH, CaCl₂ 4.19 (corresponding to 1 μM free calcium), and MgATP 2. Currents were measured at room temperature using a ramp protocol from –120 to +80 mV from a holding potential of –60 mV (sampling frequency, 10 kHz; filtered, 1 kHz).

Calcium dependence of KCNN4 was studied with intracellular (bath) solutions (in mM): KCl 30, KGluconate 100, EGTA 5, HEPES 10, pH 7.2 adjusted with KOH, and CaCl₂ with varying concentrations 4.91, 4.19, 3.61, 2.82, and 1.7 (10^{–5}, 10^{–6}, 5 × 10^{–7}, 2.5 × 10^{–7}, and 10^{–7} M of free calcium, respectively). Maxchelator was used to calculate free Ca²⁺ concentration (<http://maxchelator.stanford.edu/CaEGTA-TS.htm>). Extracellular solution was (in mM) NaCl 140, KCl 5, CaCl₂ 1, glucose 29, and HEPES 25, pH 7.4, adjusted with NaOH. Currents were evoked by voltage ramps from –120 to 80 mV (150 ms), filtered at 1 kHz, and acquired with a sampling frequency of 10 kHz. All traces were corrected for liquid junction potential. For dose-response experiments, normalized values of currents at –45 mV were plotted against free Ca²⁺ concentration.

All patch-clamp experiments were performed with a PC-controlled EPC 9 patch-clamp amplifier (HEKA, Lambrecht/Pfalz, Germany). Currents were acquired and analyzed with Pulse and Pulsefit software (HEKA).

Immunohistochemistry

Immunodetection of KCNN4-HA in HEK293 cells was performed using anti-HA antibody (Sigma-Aldrich).

Red cell cation content and volume measurements

Fresh venous blood was obtained by venipuncture from an informed patient from family 1 and a healthy volunteer. For 24-hour incubation, blood samples were stored at 37°C or 4°C.

For vanadate experiments, blood was washed 4 times at room temperature in medium containing (in mM) NaCl 147, KCl 5, MgSO₄ 2, CaCl₂ 1, and HEPES/NaOH 10, pH 7.4. The red cell suspension was then incubated at 37°C and 30% hematocrit, and 5 mM vanadate was added alone or with 10 μM TRAM34. A few minutes before sampling time, 400 μL cell suspension was taken to fill 3 nylon tubes that were centrifuged for 10 minutes at 4°C and 20 000 g at the exact sampling time. The supernatant was collected for extracellular ion content measurements. The pellet of red cells was extracted and immediately weighed. Then, dry weight was measured after overnight heating (80°C). Water content was calculated with a correction of 3.64%, corresponding to trapped medium between packed cells. Intracellular ions were extracted from dried pellets by overnight incubation at 4°C in 5 mL milliRho water (Millipore). Na⁺ and K⁺ were measured by flame spectroscopy with an Eppendorf ELEX6361.

Results

We initially investigated a fetus for severe in utero anemia without edema, requiring 1 transfusion in utero at week 27 (hemoglobin [Hb], 30 g/L). After preterm birth, he received 3 additional transfusions, immediately after birth, at 2 weeks (Hb, 65 g/L), and at 6 weeks of age (Hb, 70 g/L), and was then treated with erythropoietin for 6 weeks. Under treatment, the reticulocyte count progressively increased, and Hb value stabilized at 90 g/L at 3 months of age. No further transfusion was necessary. Currently, at 4 years and 10 months of age,

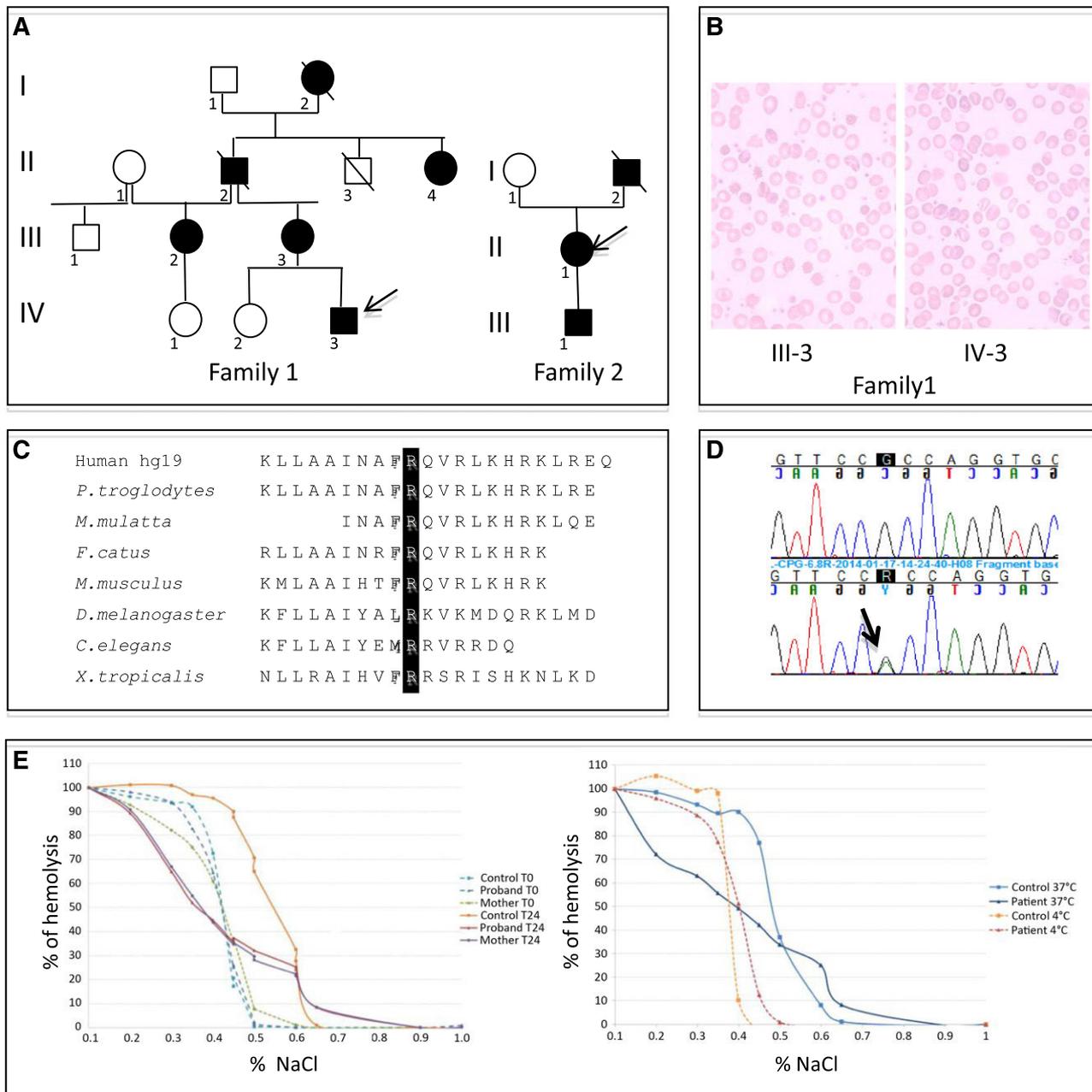


Figure 1. Red blood cell and DNA investigations. (A) Family pedigrees showing mutation segregation. (B) Blood film smears (MGG) for proband 1 and his mother. (C) Multiple interspecies protein sequence alignment of KCNN4 in the region of residue 352. (D) KCNN4 transcript sequencing: (upper) WT sequence and (lower) transcript with mutation c.1055G>A (p.Arg352His). (E) Red cell osmotic fragility test using osmotic gradient ranging from 0.1% to 1% of NaCl solution. (Left) At T0 and at T24-hour incubation at 37°C for a control and individuals III-3 and IV-3 from family 1. (Right) After T24-hour incubation at 4°C or 37°C for a control and individual III-2 from family 1.

the proband demonstrated mild anemia and splenomegaly. A clinical history revealed that the proband's mother was affected with chronic moderate hemolytic anemia of unknown origin from childhood. She was treated with a regular transfusion regimen from infancy to adolescence. Chelation therapy was started at 8 years of age, and a splenectomy was performed at 25 years old. During her adult life, she received 2 transfusions: one after a delivery and another one during an infection by the parvovirus B19. Four other members of this family originated from France and were also affected by chronic hemolytic anemia (Figure 1A). Three of 4 were splenectomized, and 2 of them have received regular transfusions and chelation therapy for periods of time. To date, none of them presented thrombotic complications.

In a second unrelated family (family 2; Figure 1A), the index case was a 25-year-old woman who has had moderate chronic hemolytic anemia since early childhood. She was never transfused and underwent a cholecystectomy because of biliary lithiasis. Her father originated from Poland and was reported to have severe hemolytic anemia treated by splenectomy and occasional transfusions. Her 2-year-old son was born after a normal pregnancy carried to term, and he also presented with well-tolerated chronic hemolytic anemia. The hematologic parameters of individuals IV-3 and III-3 from family 1 and II-1 and III-3 from family 2 are summarized in Table 1. In addition to anemia, all 4 have a discrete increase in MCHC value.

Table 1. Hematologic parameters for 4 subjects carrying the *KCNN4* c.1055G>A mutation (representative values in steady-state conditions)

	Hb (g/L)	MCV (fL)	Reticulocytes count (G/L)	MCHC (g/L)	Platelets (G/L)	Ferritin (μ g/L)
Normal range for adults	130-160	80-100	20-80	310-350	150-400	22-322
Family 1: IV-3	98	87.9	263	356	319	116
Family 1: III-3	85	109	255	354	783	94
Family 2: II-1	110	93.1	249	361	230	ND
Family 2: III-1	104	86.9	363	365	464	121

ND, not determined.

A microscopic examination of blood smears from individual IV-3 showed mild anisopoikilocytosis with <1% of target cells, polychromatophilic red blood cells, teardrop cells, elliptocytes with sometimes abnormal hemoglobin distribution, hemighosts, bite cells, knizocytes, schizocytes, and rare stomatocytic red cells (Figure 1B). For his mother and for individual II-1 from family 2, anomalies were similar, with more significant anisopoikilocytosis and the presence of acanthocytes (Figure 1B). There was no basophilic stippling of red blood cells.

The eosin 5' maleimide test, electrophoresis of red cell membrane proteins, and hemoglobin study were normal for all of them. The diagnosis of xerocytosis was not retained initially as there was almost no stomatocytes on blood films, and repeated ektacytometry was considered normal for all 4 affected individuals tested.

Whole exome sequencing was performed for 3 subjects in family 1, the proband IV-3, his affected mother III-3, and his unaffected sister IV-2. Variants were filtered against dbSNP137 and for heterozygous exonic mutations present in only the affected individuals. Thirty-three genes were found carrying exonic, nonsynonymous heterozygous mutations, among which *KCNN4* encoding the Gardos channel was the most consistent candidate because of its expression in red cells. The missense mutation c.1055G>A (p.Arg352His), confirmed by Sanger sequencing, is located in the Calmodulin interacting region and involves a residue that is highly conserved among species (Figure 1C); it was predicted pathogenic by in silico analysis (Polyphen: <http://genetics.bwh.harvard.edu/pph2>), with a score of 0.992 from a maximum of 1. Mutation segregation was studied in 3 other members of family 1, 2 affected and 1 unaffected by chronic hemolysis, and was consistent with a dominant transmission of the phenotype linked to the mutation. By direct sequencing of *KCNN4* in the 2 affected subjects of family 2, we identified the same missense mutation c.1055G>A in the heterozygous condition for both of them. Splicing was not affected by the substitution, as only normal-sized transcripts were found (Figure 1D). Using the data of exome sequencing in individual IV-3, we confirmed that no pathogenic mutation was present in *FAM38*, encoding PIEZO1 and described as the major cause of HX until now.

Further investigations were then performed for individual IV-3 and his mother. Osmotic fragility was tested to check red cell dehydration. Both mother and son had an abnormal profile after 24 hours of incubation at 37°C: 50% red cell lysis was obtained with reduced salt concentration compared with a normal control (Figure 1E). The profiles were similar to a normal control when the same analysis was performed after 24 hours at 4°C, explaining why ektacytometry was normal as it was systematically performed after incubation at 4°C. Plasmatic K⁺ concentrations in various conditions of time and temperature after sampling were measured by potentiometry and were in normal ranges. Dynamic efflux of K⁺ under Ca²⁺ stimulation was assessed by ³⁹K NMR of erythrocytes suspensions using stimulation by the ionophore A23187.

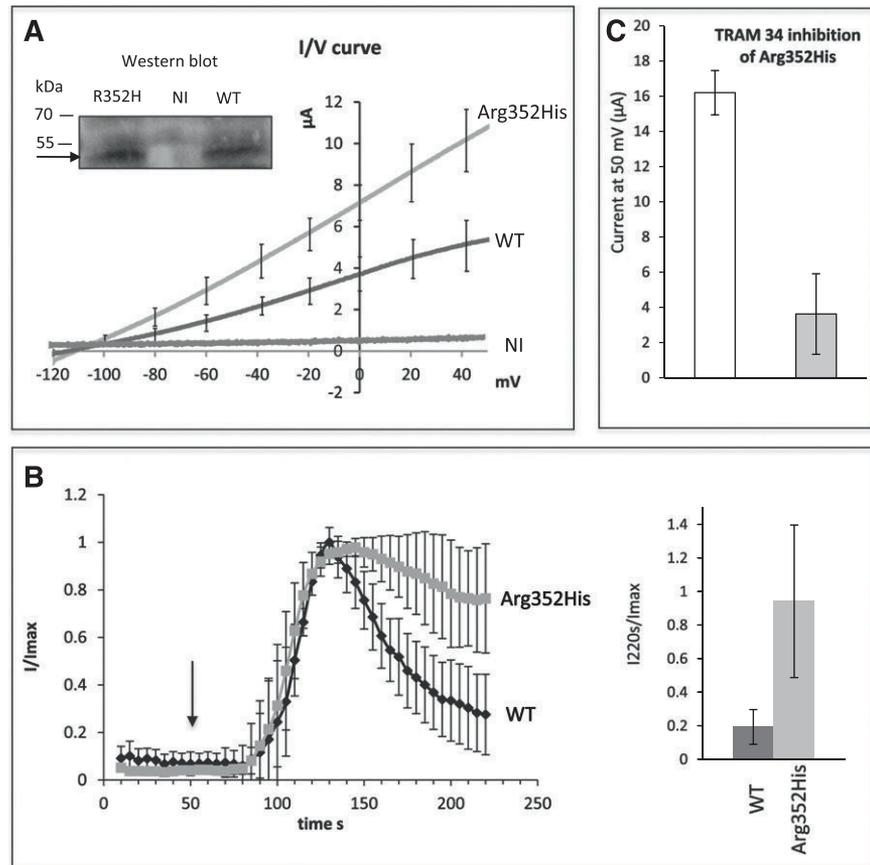
Except for a short delay in K⁺ exit following Ca²⁺ activation, no perturbation was observed (data not shown).

The function of the Gardos channel variant p.Arg352His was then investigated by the expression of a plasmid containing the mutated cDNA in *X laevis* oocytes. The current voltage curves showed that the missense mutation p.Arg352His does not prevent ion conductance through the channel (Figure 2A). The reversal potential (−120 mV) was similar between WT and the mutated channel, but the current elicited by p.Arg352His *KCNN4* expression in *Xenopus* oocytes was higher than observed with WT. Western blots confirmed that both proteins are expressed at similar levels, suggesting that the conductance increase observed for the mutated channel is directly associated with the mutation. The activation phase of the WT and p.Arg352His channel induced by calcium ionophores were similar (Figure 2B), but, whereas WT *KCNN4* activity decreased after reaching a maximum, the p.Arg352His mutant activity remained quite constant for several minutes. The high and sustained current with p.Arg352His made it difficult to record for >2 minutes after the peak. These data indicate that the mutation alters the regulation of channel activity favoring a longer activated state. In parallel, inhibition tests performed with TRAM34, the classical inhibitor of *KCNN4*, resulted in decreased current production, indicating that the p.Arg352His variant is sensitive to inhibition (Figure 2C).

To further characterize the p.Arg352His *KCNN4*, HEK293 cells were transfected with the WT or mutated channel. The mutation did not prevent the addressing of the channel to plasma membrane as illustrated on Figure 3A-B. Whole cell recording showed a different calcium-dependent activation kinetic for HEK293 cells expressing WT or p.Arg352His *KCNN4* (Figure 3B). For the former, the current progressively appeared while Ca²⁺ was diffusing from the pipette to the intracellular compartment. By contrast, in the latter case, the current was activated immediately after break-in for the mutant and further increased during the time of recording. As in oocyte experiments, the maximum current density was increased in p.Arg352His *KCNN4*-expressing cells (Figure 3B-D). These results suggest that the mutation increases the sensitivity of the channel to calcium. The leftward shift in reversal potential observed between break-in and steady state for WT confirms the delay due to Ca²⁺ diffusion to activate the channel. This delay is not observed for p.Arg352His *KCNN4*. The calcium sensitivity of WT vs mutated *KCNN4* was further explored by performing giant excised inside-out patch-clamp experiments. Figure 3E shows representative traces of K⁺ currents as a function of voltage and Ca²⁺ concentrations applied to the internal face of the membrane. In Figure 3F, currents at −45 mV are plotted as a function of Ca²⁺ concentrations. Quantitative analysis showed the calcium dependence of the WT *KCNN4* to have an apparent K_d of 0.95 ± 0.09 μ M (which is comparable to measurements done on intact red cells¹⁶), whereas the apparent K_d is 0.21 ± 0.02 μ M for p.Arg352His mutant. The Hill coefficients were not statistically different between WT and mutant *KCNN4* (3.75 ± 1.45 and 3.3 ± 0.85, respectively, n = 4).

According to electrophysiologic data, *KCNN4* should be activated by lower calcium concentration in patient red cells compared with controls. To assess the effect of an increase in intracellular Ca²⁺ on the kinetic of Gardos channel activation in control or patient red cells, the net potassium flux was measured in red cells treated by vanadate. Vanadate increases intracellular Ca²⁺ concentration in red cells by inhibiting the calcium pump and also by activating the calcium influx.^{17,18} Figure 4 illustrates the K⁺ content of control or patient red cells (individual III-2) in the presence of 5 mM vanadate with or without 10 μ M TRAM34. Whereas vanadate did not significantly change intracellular K⁺ content in control red cells over 1 hour, a significant decrease in intracellular K⁺ was observed in patient red cells, and this decrease

Figure 2. Functional analysis of the Gardos channel variant p.Arg352His. (A) Current-voltage curves of WT and mutated KCNN4 in oocyte membranes with quantification. *I/V* curves correspond to the maximal current recorded for WT or p.Arg352His KCNN4-expressing oocytes (around 135 seconds) in gluconate medium with 1 μ M A23187. Data are means of ramps recorded on 15 (WT) or 22 (p.Arg352His) oocytes. NI are control (noninjected) oocytes ($n = 4$). (Inset) Western blot detection of WT and mutated KCNN4 indicated by the arrow (around 50 kDa). (B) Activation kinetic. For oocytes expressing WT KCNN4 or p.Arg352His KCNN4, the current measured in gluconate medium at 0 mV was plotted as a function of time (left). The maximal intensity of the current being different between WT and mutated KCNN4, a ratio between *I* at different times and the I_{max} was calculated for each condition and plotted as a function of time. Data are means of 15 (WT) or 22 (p.Arg352His) oocytes coming from 3 different batches. The arrow indicates the opening of calcium ionophore perfusion. The addition of A23187 did not stimulate any current in noninjected oocytes; for clarity, the NI trace was not plotted on this graph. The bar graph (right) quantifies the remaining current at 220 seconds in oocytes expressing WT or p.Arg352His KCNN4. The current at 220 seconds was divided by I_{max} (at about 135 seconds) for each recording. Data are means \pm standard error of the mean (SEM) of 15 (WT) or 22 (p.Arg352His) oocytes. Statistical analysis was done using the Mann-Whitney test; the 2 bars are different with a risk of 0.2% (bidirectional). (C) TRAM34 inhibition: once the maximal current was reached in oocytes expressing p.Arg352His mutant, 10 μ M TRAM34 was added. This induced a rapid current decrease. The mean value of maximal currents at 50 mV was calculated (white bar) and compared with the mean value of minimal currents at 50 mV after TRAM34 addition (gray bar). Data are means of 4 oocytes \pm SEM.



was blocked by TRAM34. The K^+ efflux was correlated to cell volume decrease as illustrated on Figure 4B. No significant change in Na^+ contents was observed in control or in patient red cells at the same time (Figure 4C).

The K^+ content of red cells in blood stored for 24 hours at 37°C or 4°C is given in Table 2. In control red cells, the K^+ content is decreased by 21.1 μ mol/g dry weight after 24 hours at 37°C. This variation is similar for blood stored for 24 hours at 4°C (−24.1 μ mol/g dry weight). In contrast, there is a K^+ loss of 93.8 μ mol/g dry weight in patient red cells stored at 37°C compared with 35.3 μ mol/g dry weight for patient blood stored for 24 hours at 4°C.

Discussion

Water and solute homeostasis is essential for the maintenance of erythrocyte integrity and is controlled via the regulation of monovalent cation content. Several primary disorders of erythrocyte hydration exist and are characterized by an abnormal permeability of the erythrocyte membrane to sodium and potassium, resulting either in swelling or shrinkage of red cells.¹⁹ Clinically, these inherited disorders are associated with chronic hemolytic anemia and are due to defects in various transmembrane ion channels or transporters.³

The Gardos channel, a K^+ channel, is made of 4 identical subunits; each subunit is encoded by a single gene, *KCNN4*, and comprises 6 transmembrane domains and a pore region between the fifth and the sixth transmembrane domains (Figure 4D).⁸ In steady-state conditions, the Gardos channel is inactive. Its function is not fully elucidated in mature normal erythrocytes, but it is probably one of the major players

in red blood cell dehydration.²⁰ Under external stimulation, intracellular Ca^{2+} increases and then interacts with Calmodulin molecules that are bound tightly on each of the 4 channel subunits of the Gardos channel. Ca^{2+} binding to Calmodulin results in the opening of the channel and rapid K^+ and water efflux leading to erythrocyte dehydration and shrinkage, a mechanism referred to as the Gardos effect.^{8,21} Red blood cells are in constant movement during blood circulation where they experience mechanical stress on their membrane. Using on-cell patch-clamp experiments, it has been shown that local membrane deformation can act as a stimulating event in red cells and lead to Gardos activation, suggesting that this mechanosensory mechanism may allow erythrocytes to adapt their volume and shape to pass through the narrow capillaries of the microvasculature.²²

In the present study, the functional experiments performed on *Xenopus* oocytes showed that the channel mutated on residue 352 is normally activated by Ca^{2+} influx but permits an increased efflux of K^+ compared with the WT channel and remains open and active during a prolonged period. It is likely that the mutation, which removes a positive charge in the Calmodulin binding domain of the Gardos channel, modifies interactions with this activating partner, resulting in a more active channel. Interestingly, Arg352 has already been shown to play a crucial role in the regulation of KCNN4 activity through Calmodulin interaction in 2 different studies. First, in an analysis of the crystal structure of Calmodulin bound to the Calmodulin binding domain (CaMBD) of KCNN2, another Ca^{2+} -activated potassium channel, it is observed that the residue Arg464 in CaMBD interacts with Glu84 in the C-lobe of Calmodulin and is involved in the Ca^{2+} independent interactions between the CaMBD and Calmodulin.²³ This Arg464 residue corresponds to Arg352 in KCNN4 (Figure 4D). Second, in an extensive study analyzing the contribution of the

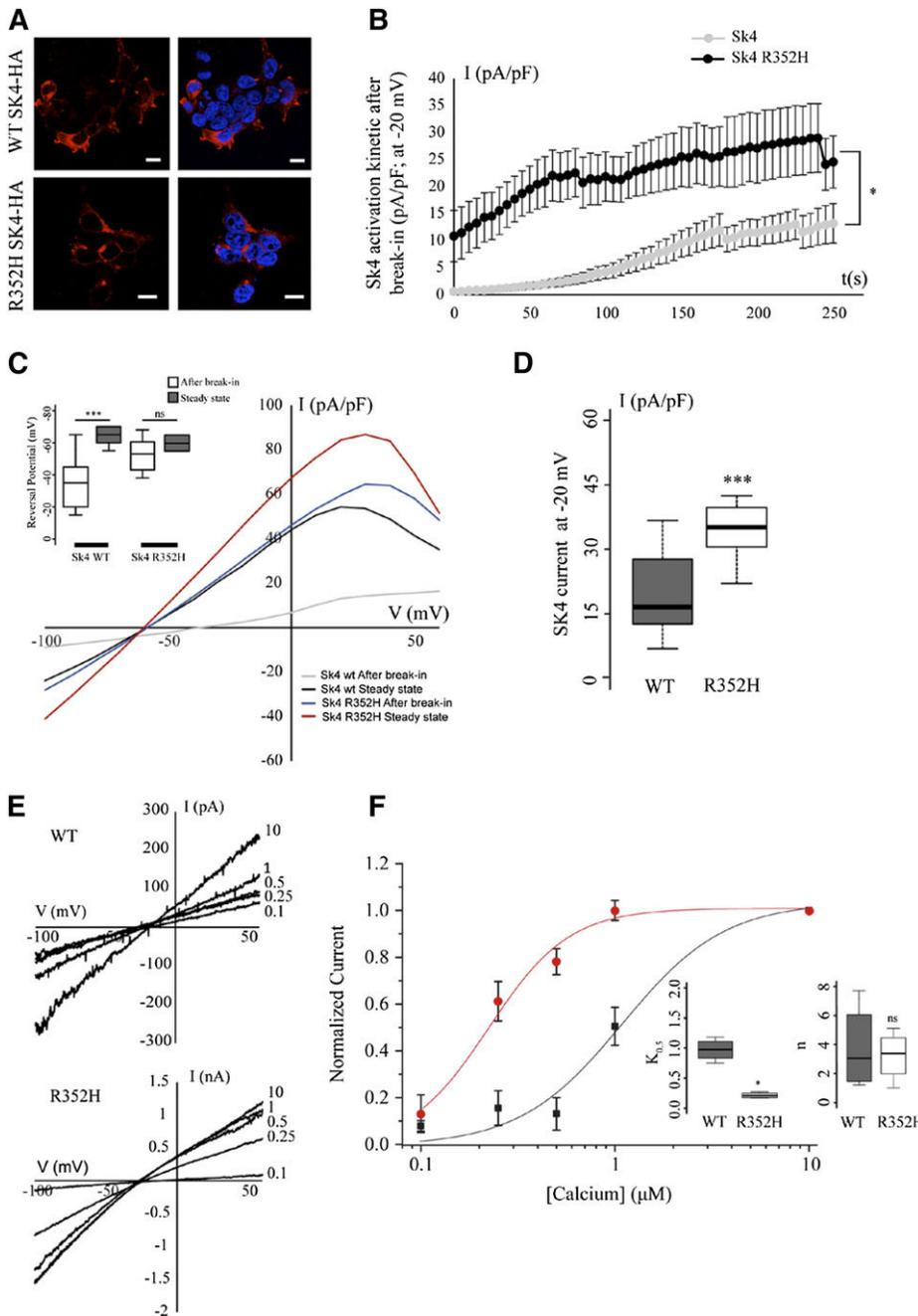


Figure 3. KCNN4 expression in HEK293 cells.

(A) Immuno-staining showing expression of WT and mutated KCNN4 at the plasma membrane. WT SK4-HA or SK4 p.Arg352His-HA were transfected in HEK293 cells and stained with an anti-HA antibody (Red). Nuclei were stained with Hoechst (blue). Images are representative of 3 independent experiments. Scale bars, 10 μ m. (B) Activation kinetic of WT and p.Arg352His KCNN4 recorded in whole cell configuration. WT KCNN4 or p.Arg352His KCNN4 was expressed in HEK293 cells and then subjected to patch-clamp experiments in whole cell configuration. Currents were recorded immediately after break-in using a 150-ms voltage ramp protocol from -120 to $+80$ mV from a holding potential of -60 mV. The current at -20 mV was plotted as a function of time. Values are mean \pm SEM of 8 to 12 experiments. (C) Representative current/voltage curves for HEK293 expressing WT or p.Arg352His KCNN4. Inset (upper left) represents reversal potentials just after break-in and at the steady state. Values are represented as a Tukey's plot ($n = 8-12$). Statistical analyses were done using the Kruskal-Wallis test followed by a Tukey post hoc test. (D) Tukey's plots showing current density at -20 mV in WT and mutated SK4 ($n = 8-12$; $***P < .001$). Statistical analysis was performed using a Mann-Whitney test. (E) Representative traces of Ca^{2+} -dependent activation of WT and p.Arg352His KCNN4 current recorded in an inside-out macropatch configuration. Currents were elicited by 150-ms voltage ramps from -120 to $+80$ mV. Each trace corresponds to a different concentration of Ca^{2+} indicated on the right side of the I/V (in micromolar). (F) Normalized K^+ current measured at -45 mV in response to $[Ca^{2+}]_i$ was plotted as a function of $[Ca^{2+}]_i$ for WT (black squares) and p.Arg352His-mutated KCNN4 (red circles). The experimental values (mean \pm SEM) were fitted with the Hill equation (Origin Software, Northampton, MA): $Y = \frac{Y_{max}[Ca^{2+}]_i^n}{(K_{0.5})^n + [Ca^{2+}]_i^n}$, where Y is the relative KCNN4 current at -45 mV (I/I_{max}), Y_{max} is the maximum current (I_{max}), $K_{0.5}$ is the apparent dissociation constant, and n is the Hill coefficient. Inset (lower right) shows Hill equation parameters $K_{0.5}$ and n . Values are represented as a Tukey plot ($n = 4$; $*P < .05$). Statistical analysis was performed using a Mann-Whitney test.

KCNN4-Calmodulin interactions to the regulation of the KCNN4 gating process, it is shown that the interaction between 2 KCNN4 monomers involves electrostatic interactions of Arg352 from 1 monomer with Glu363 of the adjacent monomer and that electrostatic interactions at Glu363 contribute to control the open probability of KCNN4.²⁴ Furthermore, in the same model, the substitution of Arg352 by Cys increases KCNN4 deactivation time. It is thus likely that Arg352 substitution by His also changes electrostatic interactions with Glu363 and therefore affects KCNN4 kinetic properties leading to a prolonged activated state of p.Arg352His KCNN4 as it is observed in our experiments involving this mutant channel.

Experiments on the human cell line HEK293 confirmed the higher current density for mutated KCNN4. This high current density could result from changes in the open probability, in the trafficking, or in the unitary conductance of the channel. The association

Calmodulin-KCNN4 is required for the trafficking and assembly of the functional channels in CHO cells.²⁵ Despite the fact that the Calmodulin binding site is mutated, the trafficking properties of the mutant are unaffected in cells with very different trafficking properties (HEK293 and *Xenopus* oocytes). Thus, the mutation is proposed to affect ion channel functional properties rather than its trafficking. Finally, these experiments showed that p.Arg352His mutation changes the Ca^{2+} sensitivity of the channel that is activated by 10 times lower Ca^{2+} concentration.

The anomaly in the kinetic of activation combined with a higher sensitivity to Ca^{2+} confers pathogenicity to p.Arg352His KCNN4. This is reminiscent of the pathogenic mechanism associated with several of the mutations affecting PIEZO1. Indeed, 3 different studies report that mutations in PIEZO1 are associated with a slowing of inactivation kinetics of this channel.^{5,6,26}

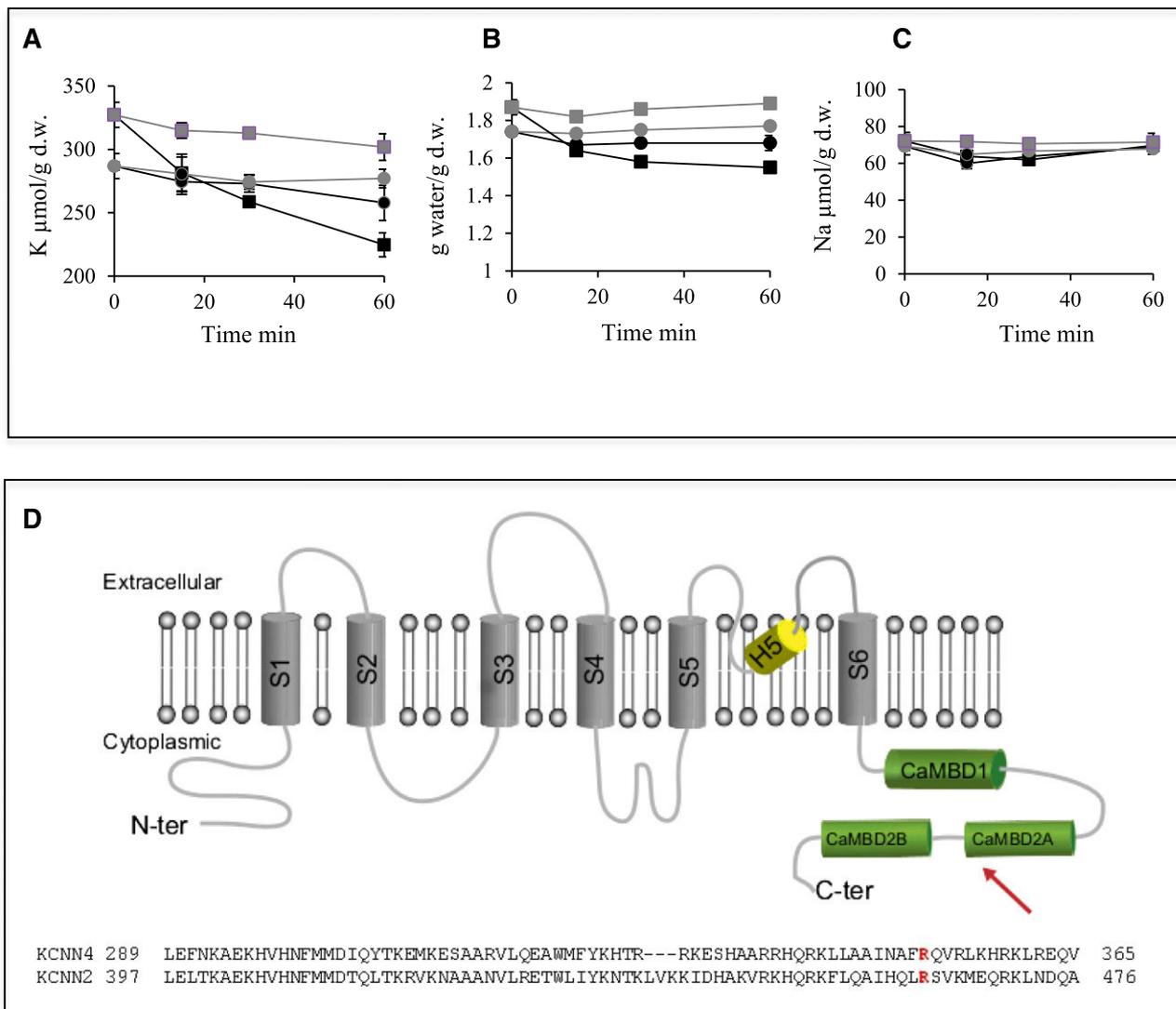


Figure 4. Red cell cation content and volume as a function of incubation time with vanadate. (A) K⁺ content, (B) cell water, and (C) Na⁺ content in red cells incubated with 5 mM vanadate (black circles for control red cells; black squares for patient red cells) or 5 mM vanadate with 10 μM TRAM34 (gray circles for control red cells and gray squares for patient red cells). Data are in μmoles per gram of dry weight and are mean ± standard deviation. n = 3. (D) (Upper) On the scheme of the Gardos channel subunit (adapted from Morales et al²⁴ and UniProt accession O15554), the arrow shows the position of the mutated residue and the 3 boxes in green represent the domain interacting with Calmodulin: the Calmodulin is constitutively bound to the 312 to 329 segment of the channel C-terminal region (CaMBD1); the Ca²⁺-dependent binding of the Calmodulin to the channel involves the segment 344 to 353 (CaMBD2A) and a stretch of 13 amino acids from 360 to 373 (CaMBD2B). (Lower) Protein sequence alignment between KCNN4 and KCNN2 in the region of the CaMBD2A.

Dehydrated red cells are usually associated with hemolytic anemia because shrinkage stimulates phosphatidylserine exposure as previously shown in both normal red blood cells, glucose-6-phosphate-deficient cells, and hemoglobin S cells.^{27,28} A relationship between cation leakage and hemolytic anemia has also been observed for other membrane proteins mutations including Band 3 mutations.²⁹ Just like in patients with PIEZO1 dysfunction, we observed in the patients with mutation in the Gardos channel, a variability in disease severity, with a level

of anemia rather severe in family 1, whereas individuals of family 2 present normal or subnormal Hb levels.^{30,31} Furthermore, the clinical pattern of expression is similar in the 2 types of channel defects with an expression in erythroid cells only, whereas both channels are expressed in various tissues. This could be explained by the fact that volume is crucial for a circulating cell such as the red cell which has to face circulatory shear stress associated with changes in oxygenation/deoxygenation. In addition, this point could shed light on PIEZO1 function in erythrocyte, suggesting that PIEZO1 and the Gardos channel might act in the same stretch-induced cation pathway involved in cell volume changes.

Table 2. K⁺ content in red cells as a function of blood temperature

	T0	T24 hours at 37°C	T24 hours at 4°C
Control	281.1 ± 4.9	260.0 ± 1.7	257.0 ± 3.4
Family 1: individual III-2	316.9 ± 5.5	223.1 ± 10.2	281.6 ± 4.6

Data are expressed in μmoles per gram of dry weight ± standard deviation for 3 samples. Blood samples were divided into 2 batches: one stored at 37°C and the other at 4°C.

In regard to iron overload due to chronic anemia, it is difficult to evaluate in family 1 as individual III-3 has regularly been transfused and treated by deferriprox, and IV-3 is too young to have iron overload. In the second family, the patients exhibit moderate iron overload as observed in chronic hemolytic anemia.

This report will probably lead to additional descriptions of this disorder, which is difficult to diagnose because biologic tests are only

slightly altered, although the anemia can be severe. The fact that the osmotic resistance test reveals abnormal profiles only after 24-hour incubation at 37°C is of importance for the diagnosis procedure but also for the mechanistic investigation of the pathogenic effect. Indeed, it suggests that ATP, which is consumed at 37°C but not at 4°C, has to be depleted for abnormal red cell resistance to be observed. At 37°C, the Na⁺/K⁺ ATPase and the Ca²⁺ pump are acting to maintain low intracellular calcium concentration and Na⁺ and K⁺ gradients, consuming ATP. With time, ATP is depleted, pump efficiency decreases, and intracellular calcium increases. In the patient red cells, a small Ca²⁺ increase is sufficient to activate the mutated Gardos channel and to induce red cell dehydration as observed in our experiments.

This identification of a *KCNN4* mutation associated with chronic hemolysis and dehydrated cells constitutes the first report of a human disease caused by a defect of the Gardos channel. The description of such a molecular defect provides a rationale for suggesting the treatment of this type of HX by specific inhibitors of the Gardos channel such as the Senicapoc, which was tested in the past in a phase 3 study for the treatment of sickle cell disease and was proven, on this occasion, to be nontoxic.³²

References

- Miller DR, Rickles FR, Lichtman MA, La Celle PL, Bates J, Weed RI. A new variant of hereditary hemolytic anemia with stomatocytosis and erythrocyte cation abnormality. *Blood*. 1971;38(2):184-204.
- Archer NM, Shmukler BE, Andolfo I, et al. Hereditary xerocytosis revisited. *Am J Hematol*. 2014;89(12):1142-1146.
- Da Costa L, Galimand J, Fenneteau O, Mohandas N. Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. *Blood Rev*. 2013;27(4):167-178.
- Zarychanski R, Schulz VP, Houston BL, et al. Mutations in the mechanotransduction protein PIEZO1 are associated with hereditary xerocytosis. *Blood*. 2012;120(9):1908-1915.
- Andolfo I, Alper SL, De Franceschi L, et al. Multiple clinical forms of dehydrated hereditary stomatocytosis arise from mutations in PIEZO1. *Blood*. 2013;121(19):3925-3935, S1-S12.
- Bae C, Gnanasambandam R, Nicolai C, Sachs F, Gottlieb PA. Xerocytosis is caused by mutations that alter the kinetics of the mechanosensitive channel PIEZO1. *Proc Natl Acad Sci USA*. 2013;110(12):E1162-E1168.
- Andolfo I, Alper SL, Delaunay J, et al. Missense mutations in the ABCB6 transporter cause dominant familial pseudohyperkalemia. *Am J Hematol*. 2013;88(1):66-72.
- Maier AD, Kuchel PW. The Gardos channel: a review of the Ca²⁺-activated K⁺ channel in human erythrocytes. *Int J Biochem Cell Biol*. 2003;35(8):1182-1197.
- Wulff H, Köhler R. Endothelial small-conductance and intermediate-conductance K_{Ca} channels: an update on their pharmacology and usefulness as cardiovascular targets. *J Cardiovasc Pharmacol*. 2013;61(2):102-112.
- Wulff H, Castle NA. Therapeutic potential of KCa3.1 blockers: recent advances and promising trends. *Expert Rev Clin Pharmacol*. 2010;3(3):385-396.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Joiner WJ, Wang LY, Tang MD, Kaczmarek LK. hSK4, a member of a novel subfamily of calcium-activated potassium channels. *Proc Natl Acad Sci USA*. 1997;94(20):11013-11018.
- Barneaud-Rocca D, Borgese F, Guizouarn H. Dual transport properties of anion exchanger 1: the same transmembrane segment is involved in anion exchange and in a cation leak. *J Biol Chem*. 2011;286(11):8909-8916.
- Martial S, Guizouarn H, Gabillat N, Pellissier B, Borgese F. Importance of several cysteine residues for the chloride conductance of trout anion exchanger 1 (tAE1). *J Cell Physiol*. 2007;213(1):70-78.
- Grygorczyk R, Schwarz W. Properties of the CA2+-activated K⁺ conductance of human red cells as revealed by the patch-clamp technique. *Cell Calcium*. 1983;4(5-6):499-510.
- Varecka L, Carafoli E. Vanadate-induced movements of Ca²⁺ and K⁺ in human red blood cells. *J Biol Chem*. 1982;257(13):7414-7421.
- Bennekou P, Harbak H, Simonsen LO. Vanadate-induced Ca(2+) and Co(2+) uptake in human red blood cells. *Blood Cells Mol Dis*. 2012;48(2):102-109.
- Rinehart J, Gulcicek EE, Joiner CH, Lifton RP, Gallagher PG. Determinants of erythrocyte hydration. *Curr Opin Hematol*. 2010;17(3):191-197.
- Thomas SL, Bouyer G, Cueff A, Egée S, Glogowska E, Ollivaux C. Ion channels in human red blood cell membrane: actors or relics? *Blood Cells Mol Dis*. 2011;46(4):261-265.
- Fanger CM, Ghanshani S, Logsdon NJ, et al. Calmodulin mediates calcium-dependent activation of the intermediate conductance K_{Ca} channel, IKCa1. *J Biol Chem*. 1999;274(9):5746-5754.
- Dyrda A, Cytlak U, Ciuraszkiewicz A, et al. Local membrane deformations activate Ca²⁺-dependent K⁺ and anionic currents in intact human red blood cells. *PLoS One*. 2010;5(2):e9447.
- Schumacher MA, Rivard AF, Bächinger HP, Adelman JP. Structure of the gating domain of a Ca²⁺-activated K⁺ channel complexed with Ca²⁺/calmodulin. *Nature*. 2001;410(6832):1120-1124.
- Morales P, Garneau L, Klein H, Lavoie MF, Parent L, Sauvé R. Contribution of the KCa3.1 channel-calmodulin interactions to the regulation of the KCa3.1 gating process. *J Gen Physiol*. 2013;142(1):37-60.
- Joiner WJ, Khanna R, Schlichter LC, Kaczmarek LK. Calmodulin regulates assembly and trafficking of SK4/IK1 Ca²⁺-activated K⁺ channels. *J Biol Chem*. 2001;276(41):37980-37985.
- Albuisson J, Murthy SE, Bandell M, et al. Dehydrated hereditary stomatocytosis linked to gain-of-function mutations in mechanically activated PIEZO1 ion channels. *Nat Commun*. 2013;4:1884.
- Lang F, Birka C, Myssina S, et al. Erythrocyte ion channels in regulation of apoptosis. *Adv Exp Med Biol*. 2004;559:211-217.
- Weiss E, Rees DC, Gibson JS. Role of calcium in phosphatidylserine externalisation in red blood cells from sickle cell patients. *Anemia*. 2011;2011:379894.
- Bruce LJ, Robinson HC, Guizouarn H, et al. Monovalent cation leaks in human red cells caused by single amino-acid substitutions in the transport domain of the band 3 chloride-bicarbonate exchanger, AE1. *Nat Genet*. 2005;37(11):1258-1263.
- Carella M, Stewart G, Ajetunmobi JF, et al. Genomewide search for dehydrated hereditary stomatocytosis (hereditary xerocytosis): mapping of locus to chromosome 16 (16q23-qter). *Am J Hum Genet*. 1998;63(3):810-816.
- Houston BL, Zelinski T, Israels SJ, et al. Refinement of the hereditary xerocytosis locus on chromosome 16q in a large Canadian kindred. *Blood Cells Mol Dis*. 2011;47(4):226-231.
- Ataga KI, Reid M, Ballas SK, et al; ICA-17043-10 Study Investigators. Improvements in haemolysis and indicators of erythrocyte survival do not correlate with acute vaso-occlusive crises in patients with sickle cell disease: a phase III randomized, placebo-controlled, double-blind study of the Gardos channel blocker senicapoc (ICA-17043). *Br J Haematol*. 2011;153(1):92-104.

Acknowledgments

The authors thank all the members from the 2 families for their contribution to this study. The pcDNA3KCNN4-HA plasmid was a kind gift of Leonard Kaczmarek's laboratory (Yale University, New Haven, CT).

Authorship

Contribution: R.R.-M., C.L., E.L., N.D., V.N., M.L., M.B., O.S., P.V., J.-P.D., C. Bérout, D.S., V.P., M.F.-T., H.G., and C. Badens participated in experimental design, execution, and interpretation; H.V., V.L., C.G., and I.T. selected and characterized patients; C.L., H.G., and C. Badens wrote the paper; and all authors reviewed the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Catherine Badens, INSERM UMR_910, Faculté de Médecine, 27 Blvd Jean Moulin, 13385 Marseille Cedex 5, France; e-mail: catherine.badens@univ-amu.fr.

2. CONCLUSIONS GENERALES et PERSPECTIVES

Les techniques de NGS ont bouleversé le monde de la génétique médicale en entraînant un changement d'échelle de l'offre diagnostique. Le transfert de ces approches du secteur recherche vers les laboratoires de diagnostic a nécessité une adaptation aussi bien des biologistes que des cliniciens. Depuis plus de 10 ans, les performances des nouveaux équipements et l'engouement qu'ils suscitent laissent envisager à moyen terme une utilisation très importante et une diffusion dans des domaines comme le dépistage et la prévention des maladies ou la médecine personnalisée. Un grand choix de plateformes est proposé mais toutes ne sont pas accessibles à un laboratoire de diagnostic en terme de rendement et de coût (Goodwin S et al., 2016).

2.1. LA PLATEFORME ET LA STRATEGIE EN DIAGNOSTIC

Grâce aux financements de la DHOS, notre laboratoire a été équipé d'une plateforme NGS capable de générer des données d'exome. Depuis plus de 3 ans, les performances de la plateforme n'ont cessé de s'améliorer suite aux modifications apportées aux réactifs et à l'acquisition d'automates augmentant rendement et reproductibilité de la plateforme tout en diminuant son coût de fonctionnement.

Au niveau des stratégies d'enrichissement, l'option choisie initialement a été de traiter par un même processus tous les échantillons que nous recevions quelle que soit la pathologie pour laquelle ils nous avaient été adressés. Cette stratégie était en opposition avec la plupart des autres centres qui ont plutôt développé une approche par panel de gènes. Après quelques mois de pratique, il semblerait que les stratégies se rejoignent et que la solution la plus adaptée actuellement soit une approche intermédiaire par large panel de gènes, offrant la meilleure balance entre coût et efficacité de séquençage.

Les prochaines étapes pour le laboratoire sont d'une part, la finalisation des documents et procédures nécessaires pour l'accréditation des analyses réalisées par cette technologie NGS et d'autre part, la mise au point d'un module d'analyse complémentaire à notre logiciel d'analyse VarAFT® pour détecter les CNVs (*Copy Number Variations*) à partir des données issues de la plateforme. Enfin, pour l'identito-vigilance, une réflexion a été lancée pour

trouver rapidement un panel de SNPs (*Single Nucleotide Polymorphisms*) qui permettraient de vérifier et confirmer la bonne correspondance avec l'identité du patient séquencé par NGS.

À plus long terme, l'évolution de ces stratégies de NGS se fera probablement vers le séquençage du génome complet. Cette stratégie ne nécessite pas de technique d'enrichissement puisqu'elle consiste à séquencer l'intégralité du génome. Toutefois, elle est encore loin d'être utilisée en routine diagnostique et reste pour l'instant limitée à la recherche.

D'autres applications vont connaître un développement important dans les mois qui viennent comme le dépistage prénatal non invasif (DPNI) de la trisomie 21 sur l'ADN fœtal circulant dans le sang maternel. La technologie NGS permet de séquencer cet ADN fœtal présent en très petites quantités dans le sang maternel dès la 6^{ème} semaine de grossesse. Les différentes études de recherche de trisomie 21 sur les ADN fœtaux circulants semblent cohérentes en termes de sensibilité (plus de 99,9% sur les données interprétables) mais on note également que 5% des prélèvements ne sont pas interprétables pour des raisons techniques. De même, le taux de faux positifs observé est autour de 0,2%, ce qui oblige dans ce cas à avoir recours à un test invasif (Bianchi DW, 2012). Mais ce test réalisé sur les femmes à risque permettrait de limiter une grande partie des pertes dues au prélèvement de cellules fœtales (estimées entre 80 et 240 fausses couches par an en France). La mise en place d'un dépistage non invasif est en cours de validation par une autre équipe au sein du département de Génétique de la Timone et pourrait rapidement intégrer la plateforme NGS de diagnostic du laboratoire.

2.2. LES SEQUENCEURS 3^{EME} GENERATION

Une troisième génération d'automates a vu le jour en 2011-2012, avec le séquençage sur molécule unique, qui combine les avantages du séquençage massif en parallèle tout en évitant l'étape d'enrichissement préalable puisque chaque molécule d'ADN est « lue » indépendamment et en temps réel dans un micropuits de séquence par une polymérase, le but étant de produire plus rapidement des reads plus longs et moins chers.

La société **Pacific Bioscience** travaille depuis 2010 à un séquenceur 3^{ème} génération capable de lire une seule molécule en utilisant des nanopores au travers desquelles la molécule d'ADN passe et est lue grâce à la détection d'un signal par des senseurs de lumière. Avec le même principe, la technologie **GridION** de **Oxford Nanopore Technologies** utilise des nanopores de nature protéique fixés sur une membrane artificielle à travers lesquels passe la molécule d'ADN en cours de synthèse. C'est la signature ionique spécifique de chaque nucléotide qui détermine la séquence. La miniaturisation est telle que cette société annonce un séquenceur de la taille d'une clé USB, directement connecté à un ordinateur qui en extrait et en analyse les données brutes. D'autres sociétés comme **Reveo** mettent au point le séquençage par microscopie TUNEL qui permet d'établir le spectre d'une séquence d'ADN donnée ou encore comme **IBM**, qui fait passer un brin d'ADN dans un nanotube et lit la séquence par mesure des paramètres physiques de chaque nucléotide.

Pour l'instant, ces nouvelles technologies ne sont encore pas utilisées en routine dans les laboratoires. Elles semblent réservées à des gros centres de séquençage ou des prestataires de service capables de supporter l'investissement dans de telles technologies. Le transfert en diagnostic de routine est encore loin (Goodwin S et al., 2016).

2.3. LE DIAGNOSTIC DE DEMAIN : VERS UNE MEDECINE PERSONNALISEE PAR L'ANALYSE DU GENOME ?

En séquençant deux génomes d'individus pris au hasard, quelle que soit leur origine géographique, on trouve en moyenne 99.9% d'identité de séquence (Jorde LB et al., 2004). Cela peut paraître élevé, mais cela implique qu'il subsiste environ 3 millions de différences ponctuelles ou SNP (*Single nucleotide polymorphism*) dont il faudrait pouvoir analyser l'effet. Ces variations de 0,1% de l'ADN permettent aux populations de s'adapter face aux facteurs environnementaux, mais influencent aussi les risques de développer une maladie ou les sensibilités aux médicaments. Ces différences de réponse aux thérapies médicamenteuses sont un point essentiel et un défi important dans la pratique clinique actuelle et le développement de nouveaux médicaments (Evans WE et al., 2003) ; (Ma Q et al., 2011) . À cela s'ajoutent les modifications épigénétiques (Feinberg AP et al., 2004) qui témoignent du fait que des facteurs non génétiques jouent un rôle clé dans l'expression des gènes.

C'est par la technologie NGS et la stratégie du séquençage du génome que nous allons pouvoir accroître nos connaissances sur l'effet des variations nucléotidiques ainsi que sur les modifications épigénétiques, afin de comprendre les maladies complexes et aboutir à une médecine personnalisée. De nombreux projets nationaux et internationaux sont développés pour soutenir cette démarche (Apellaniz-Ruiz M et al., 2016).

2.4. LE POSITIONNEMENT DE LA FRANCE AU NIVEAU INTERNATIONAL

Les États-Unis, la Chine, le Royaume Uni principalement, mais aussi d'autres pays, ont déjà commencé à intégrer la médecine génomique dans leur système de santé soutenu fortement par leur gouvernement respectif. Cette médecine génomique n'est plus un mythe mais est bien réelle et entraîne une forte compétition internationale justifiée par des enjeux majeurs.

L'accès à cette médecine génomique est en premier lieu un enjeu de Santé Publique. Par cette technologie, le parcours de soins et son organisation vont être révolutionnés dans les domaines des maladies rares, de l'oncologie mais aussi dans les maladies plus communes. Un grand nombre de patients bénéficiera de cette médecine en routine. Les données du génome couplées aux données cliniques auront un impact important sur la prise en charge diagnostique et les orientations thérapeutiques.

Les applications en médecine prédictive sont régulièrement mises en avant pour justifier le recours à ces stratégies par une diminution des coûts pour le système de santé.

Au-delà de l'enjeu en diagnostic, l'autre aspect économique important est le développement de nouvelles filières industrielles innovantes dédiées à la médecine génomique.

Il est primordial que la France se positionne au niveau international pour ne pas prendre de retard. C'est pourquoi en 2016, l'état français, représenté par son premier ministre, lance le plan **France Médecine Génomique 2025** (téléchargeable sur le lien suivant : <http://presse.inserm.fr/wp-content/uploads/2016/06/Plan-Francemedecine-genomique-2025.pdf>). Ce plan ambitionne de mettre en place un parcours de soins générique en médecine génomique commun à tous les patients quelle que soit leur pathologie (cancer, maladies rares et communes).

Les objectifs opérationnels dès 2020 seront de traiter un petit volume de patients ciblés. Pour les maladies rares, 20 000 patients et leurs parents pourraient être pris en charge (soit 60 000 génomes) et pour l'oncologie, 50 000 patients métastatiques et réfractaires aux traitements (soit 175 000 génomes).

Plusieurs mesures ont été lancées pour atteindre les objectifs de ce plan. Parmi ces mesures, on notera la mise en place de très grandes plateformes de séquençage à très haut débit mais aussi les ressources nécessaires pour stocker et analyser ces données afin de les intégrer dans le parcours de soins. Tout cela en restant dans un cadre technique et éthique sécurisés.

2.5. DIFFERENCIER LE DIAGNOSTIC ET LA RECHERCHE

La diffusion du NGS dans les différents centres hospitaliers renforce encore le lien étroit entre la recherche et le diagnostic moléculaire des maladies génétiques, mais il met aussi à jour la confusion qui peut exister entre ces deux domaines (Matthijs G et al., 2016). Dans le cadre de la mise en place du diagnostic par NGS, la limite doit être clarifiée. Le diagnostic doit offrir à chaque patient la même analyse avec un niveau de qualité maîtrisé, le but étant d'identifier la cause précise de la maladie individuelle présentée par le patient. Ceci aura un impact direct pour le patient en termes de prise en charge. Une analyse réalisée dans le cadre de la recherche tente de répondre à une question et il n'y a pas d'obligation de résultats, mais concerne plutôt l'amélioration des connaissances générales sur la pathologie. Une zone grise existe dans la mesure où les analyses faites en NGS dans le cadre du diagnostic, en particulier avec des données d'exome, peuvent conduire à émettre des hypothèses qui ne peuvent être confirmées que dans un cadre de recherche. De même, une fois le résultat rendu au patient dans un cadre diagnostic, les données non exploitées sont des données supplémentaires pouvant être utilisées par la recherche pour favoriser de nouvelles connaissances. À leur tour, des résultats pertinents obtenus en recherche doivent être confirmés dans un cadre diagnostic avant d'être portés au dossier du patient, ou même pour lui être communiqués. Il faut néanmoins se poser la question des modalités d'échanges de ces données qui sont des données de santé confidentielles et précieuses. D'un point de vue éthique, peut-on rendre ces données anonymes pour préserver la confidentialité du dossier du patient sans se priver des informations indispensables pour la recherche (Malzac P, 2016) ?

Grâce au déploiement des plateformes de NGS, de nouvelles perspectives s'ouvrent aux laboratoires de diagnostic mais aussi aux laboratoires de recherche, malgré un coût encore élevé. Le NGS permet en effet de raccourcir considérablement le temps d'analyse et de rendu de résultat tout en élargissant le nombre de gènes testés. Il augmente la proportion de diagnostics positifs.

Ces nouvelles technologies sont porteuses d'un réel espoir pour les patients atteints de maladie génétique qui subissent souvent de nombreux examens, dont certains invasifs, au cours d'une errance diagnostique encore trop longue. Enfin, le NGS permet d'identifier de nouveaux variants et de nouveaux gènes impliqués en pathologie, et ainsi d'améliorer notre connaissance du génome humain, de ses fonctions et d'identifier de nouvelles cibles thérapeutiques. Toutefois, les nouveaux problèmes éthiques soulevés par les découvertes fortuites, la propriété intellectuelle des séquences et l'interprétation délicate des données, rendent plus que jamais indispensables l'interface entre les cliniciens et les biologistes et la collaboration avec les bio-informaticiens.

BIBLIOGRAPHIE

- Altmüller J, Budde BS, Nürnberg P. Enrichment of target sequences for next-generation sequencing applications in research and diagnostics. *Biol Chem.* 2014 Feb;395(2):231-7.
- Apellaniz-Ruiz M, Gallego C, Ruiz-Pinto S, Carracedo A, Rodríguez-Antona C. Human genetics: international projects and personalized medicine. *Drug Metab Pers Ther.* 2016 Mar 1;31(1):3-8.
- Bai X, Edwards J, Ju J. Molecular engineering approaches for DNA sequencing and analysis. *Expert Rev Mol Diagn.* 2005 Sep;5(5):797-808.
- Bartoli M, Desvignes JP, Nicolas L, Martin K. Exome sequencing as a second-tier diagnostic approach for clinically suspected dysferlinopathy patients. *Muscle Nerve.* 2014 Dec;50(6):1007-10.
- Bianchi DW. Prenatal diagnostics: Fetal genes in mother's blood. *Nature.* 2012 Jul 18;487(7407):304-5.
- Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW, Kingsmore SF. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med.* 2011 Jan 12;3(65):65ra4.
- Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med.* 2014 Sep 18;371(12):1170.
- Boland JF, Chung CC, Roberson D, Mitchell J, Zhang X, Im KM, He J, Chanock SJ, Yeager M, Dean M. The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum Genet.* 2013 Oct;132(10):1153-63.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA.* 2009 Nov 10;106(45):19096-101.
- Comité Consultatif National d'Éthique (CCNE) pour les sciences de la vie et de la santé. Avis N° 124 - Réflexion éthique sur l'évolution des tests génétiques liée au séquençage de l'ADN humain à très haut débit. Disponible sur internet : URL : <http://www.ccne-ethique.fr/fr/publications>
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics.* 2011 Aug 1;27(15):2156-8.

- Dias C, Sincan M, Cherukuri PF, Rupps R, Huang Y, Briemberg H, Selby K, Mullikin JC, Markello TC, Adams DR, Gahl WA, Boerkoel CF. An analysis of exome sequencing for diagnostic testing of the genes associated with muscle disease and spastic paraplegia. *Hum Mutat.* 2012 Apr;33(4):614-26.
- Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. *N Engl J Med.* 2003 Feb 6;348(6):538-49.
- Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer.* 2004 Feb;4(2):143-53.
- Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, Carlson-Donohoe H, Gropman A, Pierson TM, Golas G, Wolfe L, Groden C, Godfrey R, Nehrebecky M, Wahl C, Landis DM, Yang S, Madeo A, Mullikin JC, Boerkoel CF, Tifft CJ, Adams D. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med.* 2012 Jan;14(1):51-9.
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet.* 2012 May;20(5):490-7.
- Goldenberg A, Riccardi F, Tessier A, Pfundt R, Busa T, Cacciagli P, Capri Y, Coutton C, Delahaye-Duriez A, Frebourg T, Gatinois V, Guerrot AM, Genevieve D, Lecoquierre F, Jacquette A, Khau Van Kien P, Leheup B, Marlin S, Verloes A, Michaud V, Nadeau G, Mignot C, Parent P, Rossi M, Toutain A, Schaefer E, Thauvin-Robinet C, Van Maldergem L, Thevenon J, Satre V, Perrin L, Vincent-Delorme C, Sorlin A, Missirian C, Villard L, Mancini J, Saugier-veber P, Philip N. Clinical and molecular findings in 39 patients with KBG syndrome caused by deletion or mutation of ANKRD11. *Am J Med Genet A.* 2016 Sep 8.
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016 May 17;17(6):333-51.
- Gorokhova S, Biancalana V, Lévy N, Laporte J, Bartoli M, Krahn M. Clinical massively parallel sequencing for the diagnosis of myopathies. *Rev Neurol (Paris).* 2015 Jun-Jul;171(6-7):558-71.
- Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 2013 Jul;15(7):565-74.
- Howell KB, McMahon JM, Carvill GL, Tambunan D, Mackay MT, Rodriguez-Casero V, Webster R, Clark D, Freeman JL, Calvert S, Olson HE, Mandelstam S, Poduri A, Mefford HC, Harvey AS, Scheffer IE. SCN2A encephalopathy: A major cause of epilepsy of infancy with migrating focal seizures. *Neurology.* 2015 Sep 15;85(11):958-66.
- Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet.* 2004 Nov;36(11 Suppl):S28-33.

- Kingsmore SF, Saunders CJ. Deep sequencing of patient genomes for disease diagnosis: when will it become routine? *Sci Transl Med.* 2011 Jun 15;3(87):87ps23
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell.* 2013 Sep 26;155(1):27-38.
- Lacoste C, Desvignes JP, Salgado D, Pecheux C, Villard L, Bartoli M, Beroud C, Levy N, Badens C, Krahn M. Coverage Analysis of Lists of Genes Involved in Heterogeneous Genetic Diseases following Benchtop Exome Sequencing using the Ion Proton. *J Genet.* 2016 Mar;95(1):203-8.
- Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001 Feb 15;409(6822):860-921. Erratum in: *Nature* 2001 Jun 7;411(6838):720. Szustakowski, J [corrected to Szustakowski, J]. *Nature* 2001 Aug 2;412(6846):565.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium.. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016 Aug 17;536(7616):285-91.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15;25(16):2078-9.
- Lupski JR. Digenic inheritance and Mendelian disease. *Nat Genet.* 2012 Dec;44(12):1291-2.
- Ma Q, Lu AY. Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacol Rev.* 2011 Jun;63(2):437-59.
- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet.* 2011 Sep;48(9):580-9.
- Malzac P. Les techniques de séquençage nouvelle génération : Enjeux Ethiques en pratique Clinique. *LAENNEC* 2016 N°3-6-17.
- Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9:387-402.

- Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, Race V, Sistermans E, Sturm M, Weiss M, Yntema H, Bakker E, Scheffer H, Bauer P. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet.* 2016 Jan;24(1):2-5.
- Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics.* 2011 Nov;10(6):374-86.
- Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, Rodenburg RJ, Yntema HG, Spruijt L, Vermeer S, Rinne T, van Gassen KL, Bodmer D, Lugtenberg D, de Reuver R, Buijsman W, Derks RC, Wieskamp N, van den Heuvel B, Ligtenberg MJ, Kremer H, Koolen DA, van de Warrenburg BP, Cremers FP, Marcelis CL, Smeitink JA, Wortmann SB, van Zelst-Stams WA, Veltman JA, Brunner HG, Scheffer H, Nelen MR. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat.* 2013 Dec;34(12):1721-6.
- O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE, Wei Z, Wang K, Lyon GJ. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med.* 2013 Mar 27;5(3):28.
- Petrikin JE, Willig LK, Smith LD, Kingsmore SF. Rapid whole genome sequencing and precision neonatology. *Semin Perinatol.* 2015 Dec;39(8):623-31.
- Piton A, Redin C, Mandel JL. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet.* 2013 Aug 8;93(2):368-83.
- Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet.* 2014 Jan;59(1):5-15.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature.* 2011 Jul 20;475(7356):348-52.
- Salgado D, Desvignes JP, Rai G, Blanchard A, Miltgen M, Pinard A, Lévy N, Collod-Bérout G, Bérout C. UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. *Hum Mutat.* 2016 May;37(5):439-46.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977 Dec;74(12):5463-7.

- Schwarz JM, Rödelberger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010 Aug;7(8):575-6.
- Smith LD, Willig LK, Kingsmore SF. Whole-Exome Sequencing and Whole-Genome Sequencing in Critically Ill Neonates Suspected to Have Single-Gene Disorders. *Cold Spring Harb Perspect Med*. 2015 Dec 18;6(2):a023168.
- Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ; NISC Comparative Sequencing Program, Margulies EH, Green ED, Collins FS, Mullikin JC, Biesecker LG. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res*. 2010 Oct;20(10):1420-31.
- Tunovic S, Barkovich J, Sherr EH, Slavotinek AM. De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome. *Am J Med Genet A*. 2014 Jul;164A(7):1744-9.
- Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med*. 2015 Jun;17(6):444-51.
- Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, Phillips AD, Shaw K, Stenson PD, Cooper DN, Tyler-Smith C; 1000 Genomes Project Consortium. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet*. 2012 Dec 7;91(6):1022-32.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013 Oct 17;369(16):1502-11.
- Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014 Nov 12;312(18):1870-9.

RAPPEL DES LIENS

- Fondation Maladies Rares. Disponible URL : <http://fondation-maladiesrares.org>
- National Center for Biology Information NCBI. Disponible URL : <http://www.ncbi.nlm.nih.gov>
- University of California Santa Cruz UCSC. Disponible URL : <https://genome.ucsc.edu>
- Association des Praticiens de Génétique Moléculaire (ANPGM) Disponible URL : <http://www.anpgm.fr/index.php/recommandations-professionnelles/29-recommandations-professionnelles-anpgm>
- Comité Consultatif National d'Éthique (CCNE) Disponible URL : <http://www.ccne-ethique.fr/fr/publications>
- France Médecine Génomique 2025- Téléchargeable sur le lien suivant : <http://presse.inserm.fr/wp-content/uploads/2016/06/Plan-Francemedecine-génomique-2025.pdf>
- Logiciel de prédiction UMD-Predictor Disponible URL : <http://umd-predictor.eu/index.php>
- Logiciel de prédiction Mutation Taster. Disponible URL : <http://www.mutationtaster.org>
- Base de données 1000 génomes. Disponible URL : <http://www.1000genomes.org>
- Base de données ExAC (Exome Aggregation Consortium). Disponible URL : <http://exac.broadinstitute.org>
- Base de données gnomAD (Genome Aggregation Database). Disponible URL : <http://gnomad.broadinstitute.org>
- Logiciel Primer 3 (v0.4.0). Disponible URL : <http://bioinfo.ut.ee/primer3-0.4.0/>
- Logiciel VarAFT® (pour *Variant Annotation and Filtering Tool*). Disponible URL : <http://varaft.eu/index.php>
- Logiciel PolyPhen. Disponible URL : <http://genetics.bwh.harvard.edu/pph2/>
- Logiciel SIFT. Disponible URL : <http://sift.jcvi.org/>

ABBREVIATIONS

ADN : Acide DesoxyriboNucléique

ANPGM : Association des Praticiens de Génétique Moléculaire

ACMG : American College of Medical Genetics (ACMG)

b : bases - **Kb** : kilobases - **Mb** : mégabases - **Gb** : gigabases – **Tb** : Térabases

BAM : *Binary Alignment Map*

CCNE : Comité Consultatif National d'Éthique

CNG : Centre National de Génotypage

CNV : Copy Number Variation

CRB : Centre de Ressources Biologiques

DPni : Dépistage Prénatal non invasif

DHOS : Direction Hospitalière de l'Offre de Soins

EEP : Encéphalopathie Epileptique Précoce

ESHG : European Society of Human Genetic

GCHh 38 : Genome Consortium Reference human version 38

Hi-Q : High Qualite

Hg19 : Human Genome version 19

Hg38 : Human Genome version 38

ISP : Ion Sphere Particule

NCBI : National Center of Biology Information

NGS : Next Generation Sequencing

OMIM: Online Mendelian Inheritance in Man

OT2/ES : One Touch 2 / Enrichissement System

PCR : Polymerase Chain Reaction

SNP : Single nucleotide polymorphism

UCSC : University of California Santa Cruz

VarAFT[®] : Variant Annotation and Filtering Tool

VCF : *Variant Call Format*

WES : pour Whole Exome Sequencing

WGS : pour Whole genome sequencing

COMMUNICATIONS ISSUES DE CE TRAVAIL DE THESE

- **Communications affichées**

- Analyse de la couverture de séquençage de listes de gènes impliqués dans des maladies génétiques hétérogènes sur une plateforme Ion Proton™ après enrichissement par capture de séquence *AgilentSure Select Inherited Diseases Panel™*.
- A novel *SCN2A* mutation responsible for dentato-olivary dysplasia in a fetus manifesting *in utero* seizures

8^{ème} Assises de Génétique Humaine – Février 2016 – LYON

- **Articles**

- **Lacoste C**, Desvignes JP, Salgado D, Pecheux C, Villard L, Bartoli M, Beroud C, Levy N, Badens C, Krahn M. **Coverage Analysis of Lists of Genes involved in Heterogeneous Genetic Diseases following Benchtop Exome Sequencing using the Ion Proton**. J Genet. 2016 Mar;95(1):203-8. PubMed PMID: 27019452.
- Rapetti-Mauss R, **Lacoste C**, Picard V, Guitton C, Lombard E, Loosveld M, Nivaggioni V, Dasilva N, Salgado D, Desvignes JP, Bérout C, Viout P, Bernard M, Soriani O, Vinti H, Lacroze V, Feneant-Thibault M, Thuret I, Guizouarn H, Badens C. **A mutation in the Gardos channel is associated with hereditary xerocytosis**. Blood. 2015 Sep 10;126(11):1273-80.
- **Lacoste C**, Fabre A, Pécheux C, Lévy N, Krahn M, Malzac P, Bonello-Palot N, Badens C, Bourgeois P. **Le séquençage d'ADN à haut débit en pratique diagnostique clinique**. Sous presse aux Archives de Pédiatrie

AUTRES PUBLICATIONS

- Devaux J, Abidi A, Roubertie A, Molinari F, Becq H, **Lacoste C**, Villard L, Milh M, Aniksztejn L. **A Kv7.2 mutation associated with early onset epileptic encephalopathy with suppression-burst enhances Kv7/M channel activity**. *Epilepsia*. 2016 May;57(5):e87-93.
- Di Meglio C, Lesca G, Villeneuve N, **Lacoste C**, Abidi A, Cacciagli P, Altuzarra C, Roubertie A, Afenjar A, Renaldo-Robin F, Isidor B, Gautier A, Husson M, Cances C, Metreau J, Laroche C, Chouchane M, Ville D, Marignier S, Rougeot C, Lebrun M, de Saint Martin A, Perez A, Riquet A, Badens C, Missirian C, Philip N, Chabrol B, Villard L, Milh M. **Epileptic patients with de novo STXBP1 mutations: Key clinical features based on 24 cases**. *Epilepsia*. 2015 Dec;56(12):1931-40
- Herbaux C, Duployez N, Badens C, Poret N, Gardin C, Decamp M, Eclache V, Daliphard S, Murati A, Cony-Makhoul P, Cheze S, Beve B, **Lacoste C**, Prebet T, Hunault-Berger M, Maloisel F, Renneville A, Figeac M, Stamatoullas-Bastard A, Bastard C, Fenaux P, Preudhomme C, Rose C; GFM (Groupe Francophone des Myelodysplasies). **Incidence of ATRX mutations in myelodysplastic syndromes, the value of microcytosis**. *Am J Hematol*. 2015 Aug;90(8):737-8.
- Abidi A, Devaux JJ, Molinari F, Alcaraz G, Michon FX, Sutera-Sardo J, Becq H, **Lacoste C**, Altuzarra C, Afenjar A, Mignot C, Doummar D, Isidor B, Guyen SN, Colin E, De La Vaissière S, Haye D, Trauffer A, Badens C, Prieur F, Lesca G, Villard L, Milh M, Aniksztejn L. **A recurrent KCNQ2 pore mutation causing early onset epileptic encephalopathy has a moderate effect on M current but alters subcellular localization of Kv7 channels**. *Neurobiol Dis*. 2015 Aug;80:80-92.
- Milh M, **Lacoste C**, Cacciagli P, Abidi A, Sutera-Sardo J, Tzelepis I, Colin E, Badens C, Afenjar A, Coeslier AD, Dailland T, Lesca G, Philip N, Villard L. **Variable clinical expression in patients with mosaicism for KCNQ2 mutations**. *Am J Med Genet A*. 2015 Oct;167A(10):2314-8. doi: 10.1002/ajmg.a.37152. Epub 2015 May 10.
- **Lacoste C**, Leheup B, Agouti I, Mowat D, Giuliano F, Badens C. **Mutations of codon 2085 in the helicase domain of ATRX are recurrent and cause ATRX syndrome**. *Clin Genet*. 2014 Nov;86(5):502-3.
- Milh M, Boutry-Kryza N, Sutera-Sardo J, Mignot C, Auvin S, **Lacoste C**, Villeneuve N, Roubertie A, Heron B, Carneiro M, Kaminska A, Altuzarra C, Blanchard G, Ville D, Barthez MA, Heron D, Gras D, Afenjar A, Dorison N, Doummar D, Billette de Villemeur T, An I, Jacqueline A, Charles P, Perrier J, Isidor B, Vercueil L, Chabrol B, Badens C, Lesca G, Villard L. **Similar early characteristics but variable neurological outcome of patients with a de novo mutation of KCNQ2**. *Orphanet J Rare Dis*. 2013 May 22;8:80.