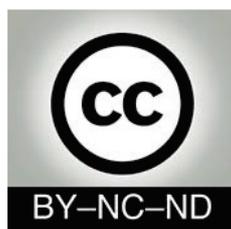




<http://portaildoc.univ-lyon1.fr>

Creative commons : Paternité - Pas d'Utilisation Commerciale -
Pas de Modification 2.0 France (CC BY-NC-ND 2.0)



<http://creativecommons.org/licenses/by-nc-nd/2.0/fr>



Université Claude Bernard Lyon 1



ANNÉE 2019 N°304

Le séquençage de l'exome pour le diagnostic des anomalies du développement chez 110 patients dans le cadre du protocole FIND

THESE D'EXERCICE EN MEDECINE

Présentée à l'Université Claude Bernard Lyon 1
Et soutenue publiquement le **25 octobre 2019**
En vue d'obtenir le titre de Docteur en Médecine
Par

Marie FAUCHER

Née le 20 mars 1990 à Rennes

Sous la direction du Docteur Gaëtan LESCA

UNIVERSITE CLAUDE BERNARD – LYON 1

2018-2019

Président de l'Université

Frédéric FLEURY

Président du Comité de Coordination des Etudes Médicales

Pierre COCHAT

Directeur Général des Services

Damien VERHAEGUE

SECTEUR SANTE

UFR DE MEDECINE LYON EST

Doyen : Gilles RODE

UFR DE MEDECINE ET DE MAIEUTIQUE
LYON SUD - CHARLES MERIEUX

Doyen : Carole BURILLON

INSTITUT DES SCIENCES PHARMACEUTIQUES
ET BIOLOGIQUES (ISPB)

Directeur : Christine
VINCIGUERRA

UFR D'ODONTOLOGIE

Doyen : Dominique SEUX

INSTITUT DES SCIENCES ET TECHNIQUES DE
READAPTATION (ISTR)

Directeur : Xavier PERROT

DEPARTEMENT DE FORMATION ET CENTRE
DE RECHERCHE EN BIOLOGIE HUMAINE

Directeur : Anne-Marie
SCHOTT

SECTEUR SCIENCES ET TECHNOLOGIE

UFR DE SCIENCES ET TECHNOLOGIES

Directeur : Fabien DE
MARCHI

UFR DE SCIENCES ET TECHNIQUES DES
ACTIVITES PHYSIQUES ET SPORTIVES (STAPS)

Directeur : Yannick
VANPOULLE

POLYTECH LYON

Directeur : Emmanuel
PERRIN

I.U.T. LYON 1

VITON

Directeur : Christophe

INSTITUT DES SCIENCES FINANCIERES
ET ASSURANCES (ISFA)

Directeur : Nicolas LEBOISN

OBSERVATOIRE DE LYON

Directeur : Isabelle DANIEL

ECOLE SUPERIEUR DU PROFESSORAT

Directeur Alain
MOUGNIOTTE

ET DE L'EDUCATION (ESPE)

PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS (Classe exceptionnelle)

| | |
|---------------------------|--|
| ADHAM Mustapha | Chirurgie Digestive |
| BONNEFOY Marc | Médecine Interne, option Gériatrie |
| BERGERET Alain | Médecine et Santé du Travail |
| BROUSSOLLE Emmanuel | Neurologie |
| BURILLON-LEYNAUD Carole | Ophthalmologie |
| CHIDIAC Christian | Maladies infectieuses ; Tropicales |
| FLOURIE Bernard | Gastroentérologie ; Hépatologie |
| FOUQUE Denis | Néphrologie |
| GEORGIEFF Nicolas | Pédopsychiatrie |
| GILLY François-Noël | Chirurgie générale |
| GLEHEN Olivier | Chirurgie Générale |
| GOLFIER François | Gynécologie Obstétrique ; gynécologie médicale |
| GUEUGNIAUD Pierre-Yves | Anesthésiologie et Réanimation urgence |
| LAVILLE Martine | Nutrition - Endocrinologie |
| LAVILLE Maurice | Thérapeutique - Néphrologie |
| LINA Gérard | Bactériologie |
| MALICIER Daniel | Médecine Légale et Droit de la santé |
| MION François | Physiologie |
| MORNEX Françoise | Cancérologie ; Radiothérapie |
| MOURIQUAND Pierre | Chirurgie infantile |
| NICOLAS Jean-François | Immunologie |
| PIRIOU Vincent | Anesthésiologie et réanimation chirurgicale |
| RODRIGUEZ-LAFRASSE Claire | Biochimie et Biologie moléculaire |
| SALLES Gilles | Hématologie ; Transfusion |
| SIMON Chantal | Nutrition |
| THIVOLET Charles | Endocrinologie et Maladies métaboliques |
| THOMAS Luc | Dermato –Vénérologie |

| | |
|--------------------------|---------------------------------|
| TRILLET-LENOIR Véronique | Cancérologie ; Radiothérapie |
| VALETTE Pierre Jean | Radiologie et imagerie médicale |
| VIGHETTO Alain | Neurologie |

PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS (1ère Classe)

| | |
|--------------------------------|--|
| ALLAOUCHICHE Bernard | Anesthésie-Réanimation Urgence |
| ANDRE Patrice | Bactériologie – Virologie |
| BERARD Frédéric | Immunologie |
| BONNEFOY- CUDRAZ Eric | Cardiologie |
| BOULETREAU Pierre | Chirurgie maxillo-faciale et stomatologie |
| CERUSE Philippe | O.R.L |
| CHAPET Olivier | Cancérologie, radiothérapie |
| DES PORTES DE LA FOSSE Vincent | Pédiatrie |
| DORET Muriel | Gynécologie-Obstétrique ; gynécologie médicale |
| FARHAT Fadi | Chirurgie thoracique et cardiovasculaire |
| FESSY Michel-Henri | Anatomie – Chirurgie Ortho |
| FEUGIER Patrick | Chirurgie Vasculaire |
| FRANCK Nicolas | Psychiatrie Adultes |
| FREYER Gilles | Cancérologie ; Radiothérapie |
| GIAMMARILE Francesco | Biophysique et Médecine nucléaire |
| JOUANNEAU Emmanuel | Neurochirurgie |
| KASSAI KOUPAI Behrouz | Pharmacologie Fondamentale, Clinique |
| LANTELME Pierre | Cardiologie |
| LEBECQUE Serge | Biologie Cellulaire |
| LIFANTE Jean-Christophe | Chirurgie Générale |
| LONG Anne | Médecine vasculaire |
| LUAUTE Jacques | Médecine physique et Réadaptation |
| PAPAREL Philippe | Urologie |
| PEYRON François | Parasitologie et Mycologie |
| PICAUD Jean-Charles | Pédiatrie |
| POUTEIL-NOBLE Claire | Néphrologie |

| | |
|------------------------|---|
| PRACROS J. Pierre | Radiologie et Imagerie médicale |
| RIOUFFOL Gilles | Cardiologie |
| RUFFION Alain | Urologie |
| SALLE Bruno | Biologie et Médecine du développement et de la reproduction |
| SANLAVILLE Damien | Génétique |
| SAURIN Jean-Christophe | Hépatogastroentérologie |
| SERVIEN Elvire | Chirurgie Orthopédique |
| SEVE Pascal | Médecine Interne, Gériatrique |
| THOBOIS Stéphane | Neurologie |
| TRONC François | Chirurgie thoracique et cardio |

PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS (2ème Classe)

| | |
|-----------------------|---|
| BARREY Cédric | Neurochirurgie |
| BELOT Alexandre | Pédiatrie |
| BOHE Julien | Réanimation urgence |
| BREVET-QUINZIN Marie | Anatomie et cytologie pathologiques |
| CHO Tae-hee | Neurologie |
| CHOTEL Franck | Chirurgie Infantile |
| COTTE Eddy | Chirurgie générale |
| COURAND Pierre-Yves | Cardiologie |
| COURAUD Sébastien | Pneumologie |
| DALLE Stéphane | Dermatologie |
| DEVOUASSOUX Gilles | Pneumologie |
| DISSE Emmanuel | Endocrinologie diabète et maladies métaboliques |
| DUPUIS Olivier | Gynécologie-Obstétrique ; gynécologie médicale |
| FRANCO Patricia | Physiologie - Pédiatrie |
| GHESQUIERES Hervé | Hématologie |
| GILLET Pierre-Germain | Biologie Cell. |
| HAUMONT Thierry | Chirurgie Infantile |

| | |
|---------------------------|---|
| LASSET Christine | Epidémiologie., éco. Santé |
| LEGA Jean-Christophe | Thérapeutique – Médecine Interne |
| LEGER FALANDRY Claire | Médecine interne, gériatrie |
| LUSTIG Sébastien | Chirurgie. Orthopédique, |
| MOJALLAL Alain-Ali | Chirurgie. Plastique., |
| NANCEY Stéphane | Gastro Entérologie |
| PASSOT Guillaume | Chirurgie Générale |
| PIALAT Jean-Baptiste | Radiologie et Imagerie médicale |
| REIX Philippe | Pédiatrie - |
| ROUSSET Pascal | Radiologie imagerie médicale |
| TAZAROURTE Karim | Médecine Urgence |
| THAI-VAN Hung | Physiologies - ORL |
| TRAVERSE-GLEHEN Alexandra | Anatomie et cytologie pathologiques |
| TRINGALI Stéphane | O.R.L. |
| VOLA Marco | Chirurgie thoracique cardiologie vasculaire |
| WALLON Martine | Parasitologie mycologie |
| WALTER Thomas | Gastroentérologie – Hépatologie |
| YOU Benoît | Cancérologie |

PROFESSEUR ASSOCIE sur Contingent National

| | |
|----------------|-------------|
| PIERRE Bernard | Cardiologie |
|----------------|-------------|

PROFESSEURS - MEDECINE GENERALE (2^{ème} Classe)

| |
|--------------------|
| BOUSSAGEON Rémy |
| ERPELDINGER Sylvie |

PROFESSEUR ASSOCIE - MEDECINE GENERALE

DUPRAZ Christian

PERDRIX Corinne

MAITRES DE CONFERENCES DES UNIVERSITES - PRATICIENS HOSPITALIERS (Hors Classe)

| | |
|---|--|
| ARDAIL Dominique | Biochimie et Biologie moléculaire |
| CALLET-BAUCHU Evelyne | Hématologie ; Transfusion |
| DIJOURD Frédérique | Anatomie et Cytologie pathologiques |
| LORNAGE-SANTAMARIA Jacqueline reproduction | Biologie et Médecine du développement et de la |
| MASSIGNON Denis | Hématologie – Transfusion |
| RABODONIRINA Meja | Parasitologie et Mycologie |
| VAN GANSE Eric | Pharmacologie Fondamentale, Clinique |

MAITRES DE CONFERENCES DES UNIVERSITES – PRATICIENS HOSPITALIERS (1ère Classe)

| | |
|----------------------------|--|
| BRUNEL SCHOLTES Caroline | Bactériologie virologie ; Hygiène .hospitalière. |
| COURY LUCAS Fabienne | Rhumatologie |
| DECAUSSIN-PETRUCCI Myriam | Anatomie et cytologie pathologiques |
| DESESTRET Virginie | Cytologie – Histologie |
| FRIGGERI Arnaud | Anesthésiologie |
| DUMITRESCU BORNE Oana | Bactériologie Virologie |
| GISCARD D’ESTAING Sandrine | Biologie et Médecine du développement et de la reproduction |
| LOPEZ Jonathan | Biochimie Biologie Moléculaire |
| MAUDUIT Claire | Cytologie – Histologie |
| MILLAT Gilles | Biochimie et Biologie moléculaire |
| PERROT Xavier | Physiologie - Neurologie |
| PONCET Delphine | Biochimie, Biologie cellulaire |

| | |
|-----------------------------|--|
| RASIGADE Jean-Philippe | Bactériologie – Virologie ; Hygiène hospitalière |
| NOSBAUM ép ROSSIGNOL Audrey | Immunologie |
| SUJOBERT Pierre | Hématologie - Transfusion |
| VALOUR Florent | Mal infect. |
| VUILLEROT Carole | Médecine Physique Réadaptation |

MAITRES DE CONFERENCES DES UNIVERSITES - PRATICIENS HOSPITALIERS (2ème Classe)

| | |
|------------------------|----------------------------------|
| BOLZE Pierre-Adrien | Gynécologie Obstétrique |
| DEMILY Caroline | Psy-Adultes |
| HALFON DOMENECH Carine | Pédiatrie |
| JAMILLOUX Yvan | Médecine Interne - Gériatrie |
| KOPPE Laetitia | Néphrologie |
| PETER DEREK | Physiologie - Neurologie |
| PUTOUX DETRE Audrey | Génétique |
| RAMIERE Christophe | Bactériologie-virologie |
| SKANJETI Andréa | Biophysique. Médecine nucléaire. |
| SUBTIL Fabien | Bio statistiques |
| VISTE Anthony | Anatomie |

MAITRES DE CONFERENCES ASSOCIES – MEDECINE GENERALE

SUPPER Irène

PROFESSEURS EMERITES

Les Professeurs émérites peuvent participer à des jurys de thèse ou d'habilitation. Ils ne peuvent pas être président du jury.

| | |
|----------------|--|
| ANNAT Guy | Physiologie |
| BERLAND Michel | Gynécologie-Obstétrique ; gynécologie médicale |

| | |
|-----------------------|---|
| CAILLOT Jean Louis | Chirurgie générale |
| CARRET Jean-Paul | Chirurgie Orthopédique |
| ECOCHARD René | Bio-statistiques |
| FLANDROIS Jean-Pierre | Bactériologie – Virologie; Hygiène hospitalière |
| DUBOIS Jean-Pierre | Médecine générale |
| LLORCA Guy | Thérapeutique |
| MATILLON Yves | Epidémiologie, Economie Santé et Prévention |
| PACHECO Yves | Pneumologie |
| PEIX Jean-Louis | Chirurgie Générale |
| SAMARUT Jacques | Biochimie et Biologie moléculaire |
| TEBIB Jacques | Rhumatologie |

Le Serment d'Hippocrate

Je promets et je jure d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans discrimination.

J'interviendrai pour les protéger si elles sont vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance.

Je donnerai mes soins à l'indigent et je n'exigerai pas un salaire au-dessus de mon travail.

Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement la vie ni ne provoquerai délibérément la mort.

Je préserverai l'indépendance nécessaire et je n'entreprendrai rien qui dépasse mes compétences. Je perfectionnerai mes connaissances pour assurer au mieux ma mission.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé si j'y manque.

COMPOSITION DU JURY

Président du jury

Monsieur le Professeur Damien SANLAVILLE

Professeur des Universités – Praticien Hospitalier – Génétique

UFR Lyon Sud

Membres assesseurs

Monsieur le Professeur Vincent DES PORTES DE LA FOSSE

Professeur des Universités – Praticien Hospitalier – Neuropédiatrie

UFR Lyon Sud

Monsieur le Professeur Patrick EDERY

Professeur des Universités – Praticien Hospitalier – Génétique

UFR Lyon Est

Monsieur le Docteur Gaetan LESCA

Maître de Conférences des Universités – Praticien Hospitalier – Génétique

UFR Lyon Est

Madame la Professeure Véronique DAVID

Professeur des Universités – Praticien Hospitalier – Génétique Moléculaire

Faculté de Médecine - Université Rennes 1

Remerciements

Aux membres du jury,

Monsieur le Professeur Damien SANLAVILLE

Merci pour ta pédagogie et ta disponibilité tout au long de ces années d'internat. Merci de m'avoir donné l'occasion de travailler sur différents projets très intéressants et de m'avoir fait confiance. Merci pour les fameuses soirées jeux et pour l'ambiance de ce super service de cytogénétique dans lequel je suis venue travailler pendant 1 an et demi avec un grand plaisir tous les jours.

Monsieur le Professeur Vincent DES PORTES DE LA FOSSE

Je vous remercie chaleureusement de votre présence au sein de mon jury et du temps que vous m'avez accordé. En espérant que cette thèse vous aura intéressé.

Monsieur le Professeur Patrick EDERY

Merci pour votre accueil au sein du service de génétique quand je suis arrivée à Lyon et lors de mon stage en génétique clinique. Merci du temps que vous m'avez accordé et d'avoir accepté d'être membre de mon jury.

Monsieur le Docteur Gaetan LESCA

Un grand merci pour tout le temps que tu m'as consacré tout au long de mon internat, pour les relectures de mémoires, d'article ou de thèse dans l'urgence, d'avoir partagé tes connaissances avec moi, de m'avoir appris à lire des panels et des exomes (j'en ferais bon usage !) et de m'avoir donné l'occasion de m'impliquer dans des cas très intéressants. Merci encore pour tout ça.

Madame la Professeure Véronique DAVID

Merci chaleureusement d'avoir accepté de juger mon travail de thèse et de participer à ma soutenance malgré la distance. J'ai hâte de travailler dans votre service et de mettre en pratique tout ce que j'ai pu apprendre pendant mon internat. Merci de votre confiance pour la suite.

Au service de cytogénétique,

Merci à l'ensemble du service de cytogénétique où j'ai passé 1 an et demi de mon internat tellement il y avait de chose à apprendre ici. Ces semestres m'ont tellement apportés aussi bien sur le plan technique qu'humain. Merci à toutes les techniciennes, **Christelle, Audrey, Isabelle, Alexia, Cathy, Brigitte, Caroline, Chantal, Amélie, Sylvie, Marie, Hélène, Amina, Laura, Flavie, Françoise** de m'avoir appris toutes les bases et les manips au début de mon internat quand je savais à peine tenir une pipette. Merci **Jessica** et merci **Laurence** et aux secrétaires. Merci pour votre disponibilité et votre gentillesse.

Merci **Marianne** pour ta disponibilité, ta bonne humeur et ta gentillesse. Merci d'avoir pris le temps de nous transmettre ton savoir et d'avoir rendu les choses difficiles à comprendre très claires après tes explications.

Merci **Caroline** pour m'avoir appris tout ce qu'il faut savoir sur la FISH et les points de cassure.

Merci **Marie-Noelle** pour ta gentillesse et ton aide pour les CGH pendant ce semestre chargé.

Merci **Nicolas** de m'avoir appris pleins de choses et de m'avoir montré les joies du NGS et de l'interprétation des variants (et de linux !!).

Merci à **Eudeline** d'avoir fait de moi une pro de la CGH et des plan de plaque de prénatal (enfin presque!).

Merci **Pierre-Antoine** pour avoir tenté de me faire comprendre la bio-info (vaste tâche !).

Merci **Audrey** pour ta disponibilité, ta patience. Tu m'as appris tellement de chose : lire une CGH, lire un panel et un exome, lire des CNV sur les exomes... Merci d'avoir répondu à toutes mes questions pas forcément toujours très pertinentes.

A mes co-internes,

Linda, ma co-interne pendant 2 stages dès mes débuts en génétique à Lyon. Merci pour ces moments partagés entre le laboratoire de cytogénétique et les consultations et avis de génétique mais surtout merci pour les découvertes culinaires marseillaises en échange de quelques crêpes bien fournies. Tu vas me manquer. Merci **Coline** pour ta bonne humeur et les chasses de pokémons. Merci à **Maud** et **Maude** avec qui j'ai partagé beaucoup de chocolat et de thés ce dernier semestre. **Maud**, je crois qu'après ton année recherche, tu as pu constater qu'il y a mieux que l'immuno, il y a la génétique (c'est fantastique !). **Maude**, tu as passé ta thèse quelque semaines avant moi alors merci pour avoir partagé les mêmes galères. **Domitille**, merci de m'avoir appris à prescrire de la ritaline (j'en ferais bon usage :) !) et pour ses discussions sérieuses interminables. **Béa**, merci d'avoir partagé ton bureau avec moi pendant 6 mois et de m'avoir montré qu'il n'y avait pas que la génétique constitutionnelle. J'ai passé un excellent semestre avec toi et **Pauline** où on a bien rigolé (sauf devant la pile de CGH !). Merci **Julie** et **Julia** pour avoir partagé mon premier semestre d'internat. **Julia**, tu m'as appris à faire des pachytènes à la vitesse de l'éclair et tu m'as très rapidement transmis ta passion pour la cytogénétique. Merci à mes co-internes de génétique, **Pauline**, **Idriss** et **Julie**. Quand j'ai commencé mon internat à Lyon, j'étais la seule interne de génétique et j'ai été ravie de votre arrivée. Merci à mes potos d'immuno, vous m'avez accepté à votre table le midi où on a bien ri. Merci **David**, **Benoît**, **Audrey**, **Emeline**, **Emma** et plus particulièrement **Léa** (j'espère que tu ne regrettes pas de t'être investi dans cette amitié :) !). Merci **Julie**, de m'avoir hébergé quelques jours quand j'ai débarqué à Lyon ne connaissant personne après notre rencontre improbable dans les montagnes albanaises.

A mes amis,

Merci **Clém**, ma plus vieille amie, les années passent mais notre complicité reste intacte malgré la distance. J'apprécie toujours beaucoup de te retrouver toi et **Adrien** autour d'une bonne bouteille de vin ;) . Merci **Coco**, pour toutes ces années de jeunesse et de folies passés avec toi et **Marine** aussi bien dans les moments joyeux que ceux un peu plus durs. Je suis tellement contente de pouvoir vous retrouver à Rennes rapidement. Merci **Solenne** pour tous ces moments géniaux partagés avec toi (sur le dancefloor, dans les boutiques ou sur la

plage !!) et je suis sûre qu'il y en aura encore bien d'autres. Merci **Alizée** pour avoir partagé tous ces fous rires, ces années d'études de médecine et de collocations festives mais quand même un peu studieuses. J'espère qu'on arrivera toujours à trouver des moments pour se voir malgré la distance. Merci **Julien** pour ces soirées étudiantes aussi bien que pour les sessions « boulot ». Les années passent mais on ne s'assagie pas quand on se retrouve. Merci au petit trio parisien, merci pour votre amitié et tous ces moments passés ensemble où on refait le monde. Merci **Alban**, tu as commencé ta thèse bien avant moi et tu l'as fini en même temps. Y en a qui s'on plus efficace que d'autres. Plus sérieusement merci pour ces discussions autour de notre goût commun pour la fantaisie :) ! Merci **Martin** pour ta franchise et ton rire trop stylé de la bombe, merci **Solenne** pour ta bonne humeur et ta spontanéité rafraichissante.

A ma famille,

Merci à **mes parents**, je suis très contente de partager ce moment avec vous et j'espère vous rendre fiers. Merci pour les valeurs que vous m'avez inculquées et votre soutien. Je n'en serais pas là sans vous aujourd'hui. Merci à ma sœur **Romane**, je suis fière de toi et de tous ce que tu as accompli. Je suis toujours ravie de te retrouver et de redevenir l'ado débile que j'étais avec toi. **Maxime**, je te laisserais 10 secondes pour commenter mes choix vestimentaires aujourd'hui et c'est tout. Toi aussi tu gères la fougère. Vous n'imaginez pas ce que cela représente pour moi que vous soyez tous là aujourd'hui.

Merci **Xavier** d'être à mes côtés depuis plus de 10 ans. On a grandi ensemble et je suis celle que je suis devenue grâce à toi. Merci de partager mes joies et mes peines depuis tous ce temps, merci de partager ma vie.

Et je finirais sur une phrase de **ma grand-mère** pleine de flegme et de sagesse qu'elle m'a dite lorsque j'ai eu ma première année de médecine et que je n'ai compris que bien après : « Médecin, et ben t'as pas choisi un métier facile ma fille ». Un métier pas facile mais tellement riche et qui me rend fière tous les jours.

Table des matières

| | | |
|-------|--|----|
| 1 | RESUME | 21 |
| 2 | INTRODUCTION | 23 |
| 2.1 | Les anomalies du développement : un enjeu diagnostique..... | 23 |
| 2.1.1 | Généralités..... | 23 |
| 2.1.2 | La médecine génomique : un enjeu de santé publique majeur..... | 23 |
| 2.1.3 | Intérêt du diagnostic génétique..... | 24 |
| 2.1.4 | L'arrivée du séquençage haut-débit dans la pratique clinique | 25 |
| 2.1.5 | Le cas particulier de la Déficience Intellectuelle | 29 |
| 2.2 | Le Séquençage à Haut Débit | 36 |
| 2.2.1 | Les différentes technologies de séquençage à Haut Débit | 36 |
| 2.2.2 | Les différentes étapes du séquençage haut-débit : Exemple de la technologie d'Illumina de séquençage par synthèse..... | 37 |
| 2.2.3 | Les différentes stratégies de Séquençage à Haut Débit..... | 43 |
| 2.2.4 | Analyse des CNV sur les données du séquençage de l'exome | 48 |
| 2.3 | Contexte de l'étude : Le projet « FIND » ou « Les données non sollicitées produites par SHD en diagnostic : du besoin des patients aux modalités organisationnelles »..... | 49 |
| 2.4 | Objectif..... | 53 |
| 3 | MATERIELS ET METHODES | 54 |
| 3.1 | Patients..... | 54 |
| 3.2 | Séquençage de l'exome | 55 |
| 3.2.1 | Technique de séquençage..... | 55 |
| 3.2.2 | Analyse bioinformatique | 56 |
| 3.2.3 | Interprétation clinico-biologique..... | 58 |
| 3.3 | Vérification des variants nucléotidiques par séquençage SANGER..... | 67 |
| 3.4 | Vérification des CNV par PCR quantitative..... | 67 |
| 3.5 | Compte Rendu des résultats | 68 |
| 4 | RESULTATS | 69 |
| 4.1 | Patients..... | 69 |
| 4.2 | Diagnostic primaire | 70 |
| 4.3 | Données Non Sollicitées..... | 77 |
| 4.4 | Analyse des CNV | 81 |
| 5 | DISCUSSION | 85 |
| 5.1 | Implication des SNV dans le diagnostic primaire de la cohorte..... | 85 |

| | | |
|-----|---|-----|
| 5.2 | Implication des CNV dans le diagnostic primaire de la cohorte | 87 |
| 5.3 | Recherche des Données Non Sollicitées dans la cohorte | 89 |
| 5.4 | Les Limites du séquençage de l'exome | 92 |
| 5.5 | L'intérêt de la réanalyse des données du séquençage de l'exome..... | 94 |
| 5.6 | L'avenir du Séquençage à Haut Débit | 95 |
| 6 | CONCLUSIONS..... | 97 |
| 7 | REFERENCES | 99 |
| 8 | ANNEXE | 105 |

Tables des illustrations

| | |
|---|----|
| Figure 1 : Evolution du séquençage du génome humain | 26 |
| Figure 2 : Augmentation du nombre de gènes associés à un phénotype au cours des années sur la base de données OMIM..... | 27 |
| Figure 3 : Nombre de gènes liés à la DI isolée ou syndromique en fonction des années et du mode de transmission (1) | 28 |
| Figure 4 : Critères de gravité de la DI selon le DSM-V | 30 |
| Tableau 1 : Etiologie des DI (2) | 32 |
| Figure 5 : Arbre décisionnel pour la prescription d'analyse génétique devant une DI syndromique ou non | 34 |
| Tableau 2 : Résultats des séries d'exomes cliniques chez des patients présentant des troubles du neurodéveloppement | 35 |
| Figure 6 : Les principales technologies de séquençage à haut débit | 37 |
| Figure 7 : Préparation de la librairie : double indexage avec séquence P5 - index 1 et séquence P7 - index2 | 39 |
| Figure 8 : Amplification clonale en « cluster » par la technologie Illumina..... | 39 |
| Figure 9 : Séquençage par synthèse : émission d'une fluorescence par incorporation d'un nucléotide | 41 |
| Figure 10 : Séquençage par synthèse des brins « forward » et « reverse » et des index | 41 |
| Figure 11 : Représentation schématique de la notion de profondeur et de couverture de séquençage. | 43 |
| Tableau 3 : Avantages et inconvénient des différentes stratégies de séquençage..... | 47 |
| Figure 12: Méthode de détection des CNV par comparaison des profondeurs de lecture (34) | 49 |
| Figure 13 : Pipeline bioinformatique utilisé pour l'analyse des données de séquençage de l'exome..... | 57 |
| Figure 14 : Critère de pathogénicité d'un variant selon l'ACMG..... | 59 |
| Figure 15 : Règles de combinaison des critères de l'ACMG pour définir la pathogénicité d'un variant..... | 60 |
| Figure 16 : Capture d'écran des fichiers Excel formatés à partir du fichier VCF..... | 61 |
| Figure 17 : Visualisation d'un fichier BAM avec le logiciel Alamut Visual..... | 62 |
| Figure 18 : Représentation graphique d'une délétion de 417 Kb en 16p11.2 héritée du père . | 63 |

| | |
|---|----|
| Figure 19: Fichier Excel listant les CNV d'un patient | 64 |
| Figure 20 : Workflow du séquençage de l'exome..... | 66 |
| Tableau 4: Caractéristiques cliniques des patients de la cohorte | 70 |
| Tableau 5: Signes cliniques les plus fréquents retrouvés dans la cohorte de patient | 70 |
| Tableau 6: Caractéristiques cliniques et taux diagnostiques des patients ayant eu un DP..... | 72 |
| Tableau 7: Caractéristiques des variants de classe 4 et 5 | 72 |
| Tableau 8 : Phénotype des patients ayant eu un DP et caractéristiques des variants de classe 4 et 5 expliquant le phénotype..... | 76 |
| Tableau 9: Caractéristiques des variants pathogènes et probablement pathogènes des DNS .. | 80 |
| Tableau 10 : Caractéristiques des CNV retenus lors de l'analyse sur les données de séquençage de l'exome..... | 83 |
| Figure 21 : Représentation graphique des CNV retenus | 84 |
| Figure 22 : Algorithme de diagnostic proposé par Srivastava et al. En incorporant le séquençage de l'exome dans l'exploration des DI et des TSA. | 86 |

GLOSSAIRE

AD : Anomalie du Développement

DI : Déficience Intellectuelle

DNS : Données Non Sollicitées

CNV : *Copy Number Variation* ou variation du nombre de copie

SNV : *Single Nucleotide Variation* ou variant nucléotidique

SHD : Séquençage à Haut Débit

NGS: *New Generation Sequencing*

OMIM: *Online Mendelian Inheritance in Man*

QI : Quotient Intellectuel

TSA : Trouble du Spectre Autistique

FISH : *Fluorescent In Situ Hybridation* ou Hybridation fluorescente in situ

DS : Données secondaires

SV : *Structural Variation* ou variant de structure

ACMG: *American College of Medical Genetics*

XHMM : *eXome Hidden Markov Model*

ACP : Analyse en Composantes Principales

HGMD : *Human Gene Medical Database*

VOUS : *Variant of unknown significance* ou variant de signification indéterminé

PIEV : Facteur de susceptibilité

DP : Diagnostic Primaire

1 RESUME

Les anomalies du développement et la déficience intellectuelle (AD/DI) concernent environ 3 % de la population. L'arrivée du séquençage de l'exome a permis d'augmenter le taux diagnostique et de mettre en évidence de nouveaux gènes impliqués dans ces pathologies. L'exome permet de faire le diagnostic primaire mais détecte également des données dites « non sollicitées » (DNS) donnant lieu à des actions préventives et/ou curatives. De plus, l'analyse de variation du nombre de copie (CNV), actuellement étudiée grâce à l'Analyse Chromosomique sur Puce à ADN (ACPA), est également possible via les données de séquençage de l'exome. C'est afin d'évaluer les besoins des patients et les modalités organisationnelles de l'analyse des données non sollicitées détectées par l'exome que le projet FIND regroupant 3 CHU dont les Hospices Civils de Lyon, a vu le jour.

L'objectif de cette thèse est d'analyser les résultats du séquençage de l'exome en trio chez 110 patients porteurs d'AD/DI des Hospices Civils de Lyon inclus dans le projet FIND. Les patients ont au préalable bénéficié d'une ACPA, examen de première intention dans l'exploration d'une AD/DI.

Concernant l'analyse des données primaires, 39 variants pathogènes ont été détectés chez 38 patients dans les gènes suivants : *CDK13*, *BCL11A*, *CYFIP2*, *GRIN2B*, *CAMK2A*, *GNB5*, *PORCN*, *NCKAP1*, *HEXA*, *KANSL1*, *KMT2B*, *KCNQ5*, *EFTUD2*, *PACSI*, *PPP2R5D*, *DYRK1A*, *MED13L*, *ARID1B*, *WDR45*2*, *RAC1*, *GRM1*, *TCF4*, *CDK13*, *MTO1*, *MORC2*, *PCGF2*, *DDX3X*, *STAG1*, *SLTRK6*, *TRRAP*, *ZMYND11*, *PIGC*, *GNB1*, *RAPSN*, *PC*, *DSP*, *HUWE1* et *LRIG2*. Le taux diagnostique est de 34.5%. Concernant les données non sollicitées, 15 variants pathogènes chez 14 patients ont été mis en évidence (12.7% des patients). Ces variants concernaient les gènes actionnables *TNNI3*, *LDLR*, *FBNI*, *PROC* et *SERPINC1*, les gènes de *HFE* et *CFTR* ainsi que dans les gènes de pharmacogénétique *CYP2C9* et *CYP2C19*. Concernant l'analyse des CNV, 6 CNV ont été retenus (analyse sur 84 patients ; résultats provisoires). Tous les CNV ont été détectés par l'ACPA préalablement. 3 CNV initialement classés comme variant de signification indéterminée ont été reclassés en CNV pathologiques à la suite de la nouvelle analyse de la bibliographie récente.

Le taux diagnostique de l'exome retrouvé dans notre cohorte correspond au taux décrit dans une récente méta-analyse de la littérature (36%). L'exploration des DNS permet d'adapter la prise en charge des patients via un conseil génétique approprié et une surveillance

spécifique. L'analyse des CNV est efficace sur les données de l'exome et a permis de retrouver tous les CNV des patients déjà connus. La nouvelle analyse des données de l'ACPA est nécessaire avant de poursuivre les explorations génétiques au vu de l'évolution rapide des connaissances dans le domaine de la génétique. Ceci sera également le cas concernant les données de l'exome. Il est très probable qu'à l'avenir le séquençage de l'exome soit l'examen de première intention pour l'exploration des AD/DI puisqu'il présente un taux diagnostique supérieur à celui de l'ACPA (36% contre 12 à 15%) et qu'il permettra à la fois d'analyser les variants monogéniques et les CNV d'un patient.

Mots clés : Déficience intellectuelle, anomalie du développement, séquençage de l'exome, diagnostique primaire, données non sollicitées, CNV

2 INTRODUCTION

2.1 Les anomalies du développement : un enjeu diagnostique

2.1.1 Généralités

Une maladie rare est définie en Europe par une prévalence inférieure à 5 personnes touchées sur 10 000 (orphanet, Règlement (CE) No 141/2000 du Parlement européen et du Conseil du 16 décembre 1999 concernant les médicaments orphelins, http://ec.europa.eu/health/files/eudralex/vol-1/reg_2000_141/reg_2000_141_fr.pdf) ce qui correspond en France à moins de 30 000 personnes. L'ensemble de ces maladies rares concerne 3 millions de personnes en France. Deux tiers de ces maladies sont graves et invalidants et une personne concernée sur 2 est un enfant. Ces maladies seraient responsables chez eux de 30% de la mortalité infantile et dans 80% des cas, la cause est génétique (<https://fondation-maladiesrares.org/les-maladies-rares/les-maladies-rares-bis/la-definition-des-maladies-rares/>).

Les anomalies du développement (AD) concernent 3% de la population soit environ 1.8 millions de personnes en France. On compte 40 000 nouveaux cas en France par an, toutes étiologies confondues (<http://anddi-rares.org>); ce cadre nosologique constitue donc un enjeu de santé public. Il regroupe les syndromes polymalformatifs et dysmorphiques avec ou sans retard du développement intellectuel secondaire à des anomalies chromosomiques ou géniques. Il existe une hétérogénéité génétique majeure expliquant ces pathologies puisqu'environ 7 000 maladies monogéniques rares distinctes sont répertoriées ainsi qu'un très grand nombre d'anomalies chromosomiques. Les cahiers d'orphanet publient 2 fois par an une liste mise à jour de ces pathologies rares génétiques (<https://www.orpha.net>).

2.1.2 La médecine génomique : un enjeu de santé publique majeur

Ces pathologies rares sont aujourd'hui reconnues par les institutions politiques européennes et nationales avec notamment la mise en place des différents Plan Nationaux

Maladies Rares 1, 2 et 3, le développement de laboratoires de recherche translationnelle en génétique et la constitution de nombreuses filières de soins et de réseaux, dont la filière anDDI-Rares, ayant pour objectif d'améliorer l'accès au diagnostic génétique ainsi que la prise en charge globale de ces pathologies. Plus récemment, le Plan France Médecine Génomique 2025, lancé en 2016 à la demande du premier ministre Manuel Valls définit plusieurs objectifs à atteindre concernant la médecine génomique : placer la France dans le peloton de tête des pays engagés dans la médecine personnalisée de précision, préparer l'intégration de la médecine génomique dans le parcours de soin et mettre en place une filière nationale de médecine génomique capable d'innovation scientifique et technologique. Dans ce plan, le déploiement de 12 plateformes de séquençage à très haut débit couvrant l'ensemble du territoire français était prévu à l'horizon de 2020 ainsi que la mise en place d'un Collecteur analyseur de données (CAD), capable d'exploiter le volume considérable de données générées. En 2017, 2 projets de plateformes pilotes ont été sélectionnées parmi les 10 qui impliquaient l'ensemble des CHU français. Il s'agit de la plateforme de Séquençage Haut-Débit SeQOIA en Ile-de-France regroupant l'Assistance Publique-Hôpitaux de Paris, l'Institut Gustave Roussy et l'institut Curie et la plateforme AURAGEN en région Auvergne et Rhône-Alpes comprenant les 4 CHU de la région, les 2 centres de lutte contre le cancer régionaux et l'institut de cancérologie de la Loire ainsi que l'école des mines de Saint Etienne. Ce projet se concrétise avec le séquençage des premiers génomes de patients dans le domaine de la cancérologie en mai 2019 sur la plateforme AURAGEN. Ces projets montrent une volonté de poser les bases de la médecine génomique et personnalisée de demain comprenant une prise en charge clinique très spécialisée, l'accès à des technologies de séquençage innovantes pour le diagnostic des maladies rares génétiques ainsi que le développement de la recherche afin de mieux comprendre la physiopathologie des anomalies génétiques et d'aboutir à terme à des approches thérapeutiques spécifiques.

2.1.3 Intérêt du diagnostic génétique

Les AD sont difficiles à diagnostiquer devant leur très grand nombre et leur hétérogénéité clinique et génétique. Le diagnostic de certitude arrive souvent des années après les premiers signes cliniques et les patients ont souvent amenés à consulter différents spécialistes. Cette errance diagnostique, également décrite dans certains articles comme une

« odyssee », est difficile à vivre pour le patient sur le plan psychologique et sa prise en charge médicale mais aussi socio-économique s'en retrouve impactée.

Pourtant, poser un diagnostic de certitude est indispensable et permet plusieurs choses :

- Améliorer la prise en charge
 - o Médicale avec la mise en place d'une surveillance adaptée et le dépistage des complications associées
 - o Socio-éducative avec la mise en place d'aides et d'interventions précoces adaptées
- Préciser le pronostic, notamment neurodéveloppemental
- Donner un conseil génétique approprié au patient et à sa famille
- Eviter de nouveaux examens inutiles et potentiellement invasifs
- Nommer la pathologie et ainsi souvent déculpabiliser les familles et le patient
- Soutenir les familles avec le rôle prépondérant des associations de patients qui font le lien entre les familles d'enfant atteint d'une même pathologie
- Proposer aux patients de participer à d'éventuels essais thérapeutiques selon la pathologie
- Améliorer les connaissances sur les causes des anomalies de développement en mettant en évidence de nouveaux gènes et de nouvelles voies de signalisation qui pourraient potentiellement être la cible de nouvelles thérapeutiques

2.1.4 L'arrivée du séquençage haut-débit dans la pratique clinique

Les premières déterminations de séquence d'ADN datent des années 1970 depuis le développement par Frederick Sanger au Royaume Uni de la méthode de séquençage dite « SANGER ». Cette technologie a ensuite été automatisée dans les années 1990 et a permis grâce au « *Human Genome Project* » de séquencer le génome humain. Ce projet a cependant nécessité une collaboration internationale, 10 ans de travail et un investissement d'environ 3 milliards de dollars. Depuis les années 2005, de nouvelles technologies de séquençage dites « haut-débit » (SHD) ou séquençage de nouvelle génération (NGS) ou séquençage massif en parallèle, ont permis d'aboutir à un changement considérable quant à la quantité de données de séquences générées. Ceci a permis d'accéder au séquençage du génome humain entier dans

le cadre de la recherche puis dans le cadre médical avec des durées et des coûts de séquençage en diminution. Aujourd'hui, le séquençage du génome d'un individu prend quelques jours et coûte quelques milliers de dollars. La figure 1 illustre l'historique du séquençage du génome humain.

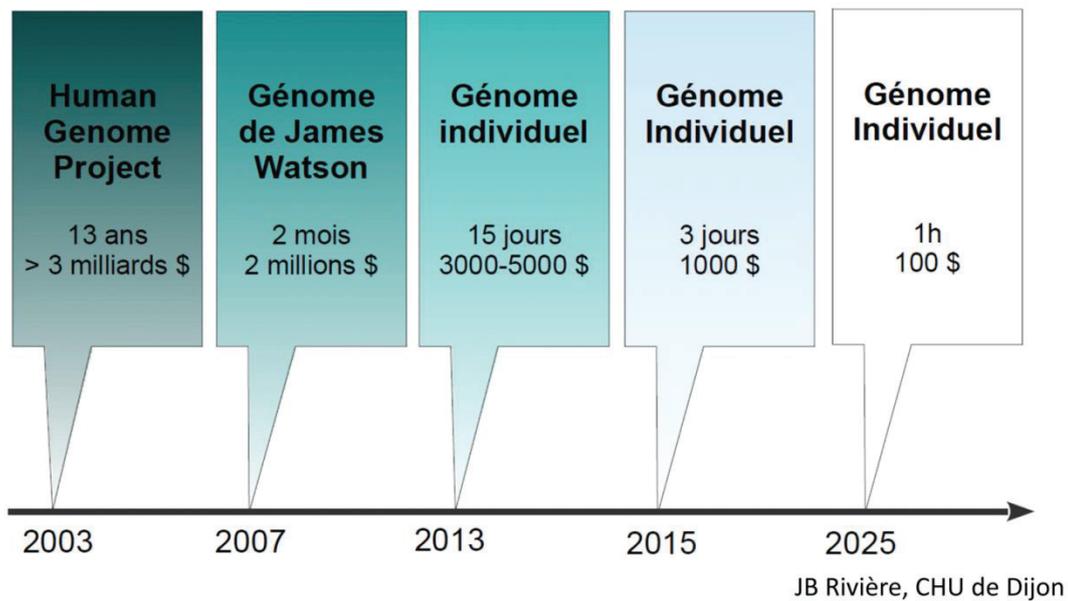


Figure 1 : Evolution du séquençage du génome humain

Concernant les AD, l'hétérogénéité génétique est extrême, tous les gènes n'étant pas encore connus. Les techniques de diagnostic moléculaire des AD ont suivi l'évolution du séquençage de l'ADN. Pour les pathologies monogéniques, jusqu'à il y a quelques années, un séquençage ciblé d'un ou plusieurs gènes pouvait être réalisé devant une orientation diagnostique mais cette approche était très chronophage et le rendement diagnostique dépendait grandement du gène et du phénotype présenté par le patient. L'émergence du séquençage à haut débit a drastiquement modifié le processus diagnostique génétique chez ces patients avec l'accès au séquençage d'un très grand nombre de gènes, à un moindre coût et la possibilité de séquencer plusieurs patients en même temps grâce au multiplexage. Ainsi le séquençage SANGER ciblé d'un ou plusieurs gènes a été remplacé par le séquençage de « panels de gènes » impliqués dans certains types de pathologies comme par exemple le panel DI44 comportant les 44 gènes les plus fréquemment associés à la déficience intellectuelle ou

encore par le séquençage à haut-débit de l'exome c'est à dire le séquençage de l'ensemble des exons (partie codante de l'ADN) du génome.

Depuis l'émergence des nouvelles technologies de séquençage à haut débit datant d'une dizaine d'année, le nombre de gènes découverts associés à une pathologie humaine a augmenté de façon exponentielle. La figure 2 montre cette augmentation. Les gènes sont issus de la base de données OMIM (*Only Mendelian Inheritance in Man*) qui répertorie les gènes impliqués en pathologie humaine et associés à un phénotype.

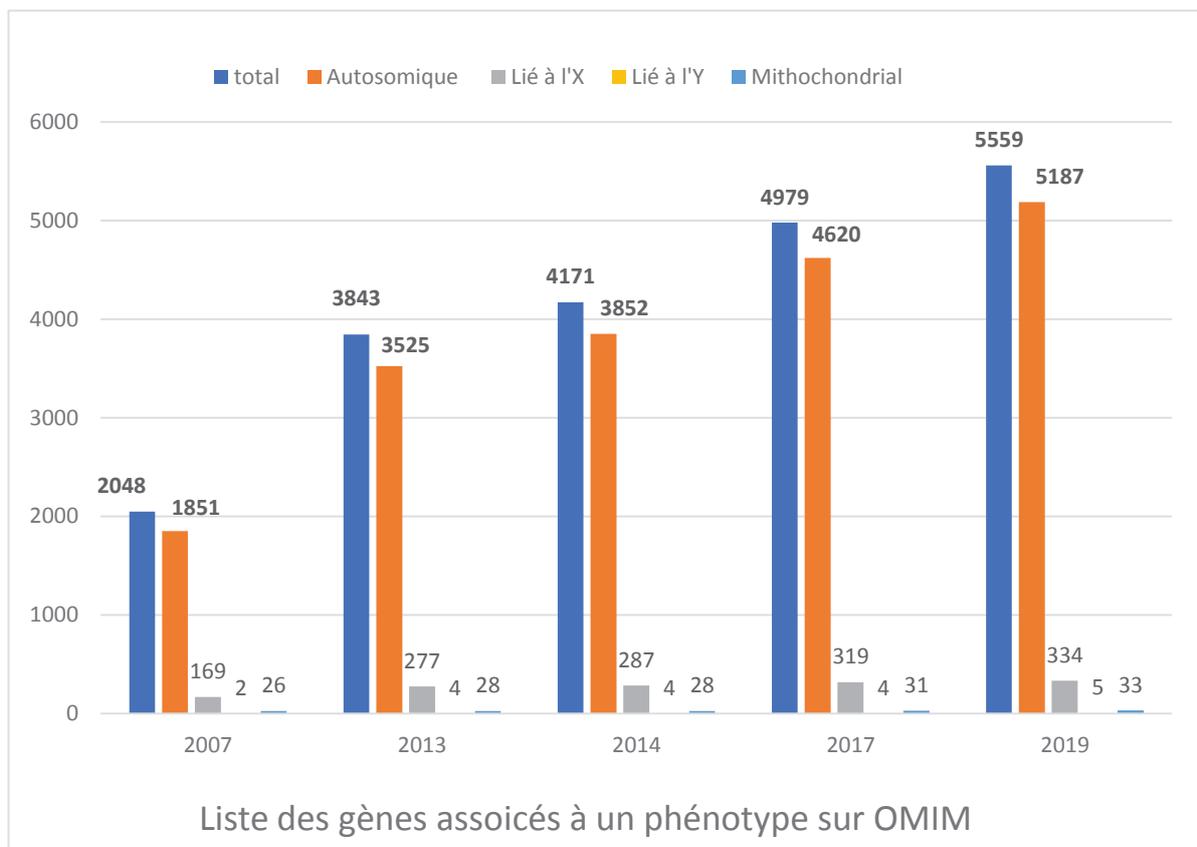


Figure 2 : Augmentation du nombre de gènes associés à un phénotype au cours des années sur la base de données OMIM

Découvrir de nouveaux gènes impliqués en pathologie humaine est nécessaire afin d'avoir une meilleure connaissance des processus biologiques responsables des pathologies, de proposer des axes de recherches pertinents et ainsi à terme de pouvoir proposer aux patients une prise en charge appropriée mais également dans certains cas une thérapeutique adaptée. La découverte de nouveaux gènes reposait principalement sur des études de liaison dans des familles consanguines ou concernant des pathologies liées au chromosome X mais

également grâce à des familles porteuses de pathologie dans des gènes de transmission autosomique dominante (Maladie de Huntington avec le gène *HTT* ou les prédispositions au cancer du sein et de l’ovaire avec les gènes *BRCA1* et *BRCA2*). Cela mettait en évidence des pathologies récessives autosomique ou liées à l’X. Ensuite, l’arrivée de l’Analyse Chromosomique sur Puce à ADN (ACPA) qui permettait une analyse pangénomique a permis de découvrir de nouveaux gènes en mettant en évidence des délétions ou des duplications de ces gènes. Des puces à ADN enrichies sur le chromosome X ont également permis d’associer de nouveaux gènes à la Déficience Intellectuelle (DI) chez le garçon. Aujourd’hui, le séquençage de l’exome ou du génome permet deux principales approches. La première consiste à séquencer l’ADN de patients présentant un phénotype similaire et à filtrer les variants afin de faire émerger un gène ou une voie de signalisation commune. La deuxième repose sur le séquençage de l’ADN d’un patient et de ses parents (séquençage dit en trio) et/ou de la fratrie et de filtrer les variants en fonction des différents modes d’hérédité (*de novo*, autosomique dominant, récessif, lié à l’X, hétérozygote composite) afin de limiter les gènes candidats. Cette dernière approche a augmenté drastiquement la découverte de gènes liés à la DI ces dernières années (Figure 3).

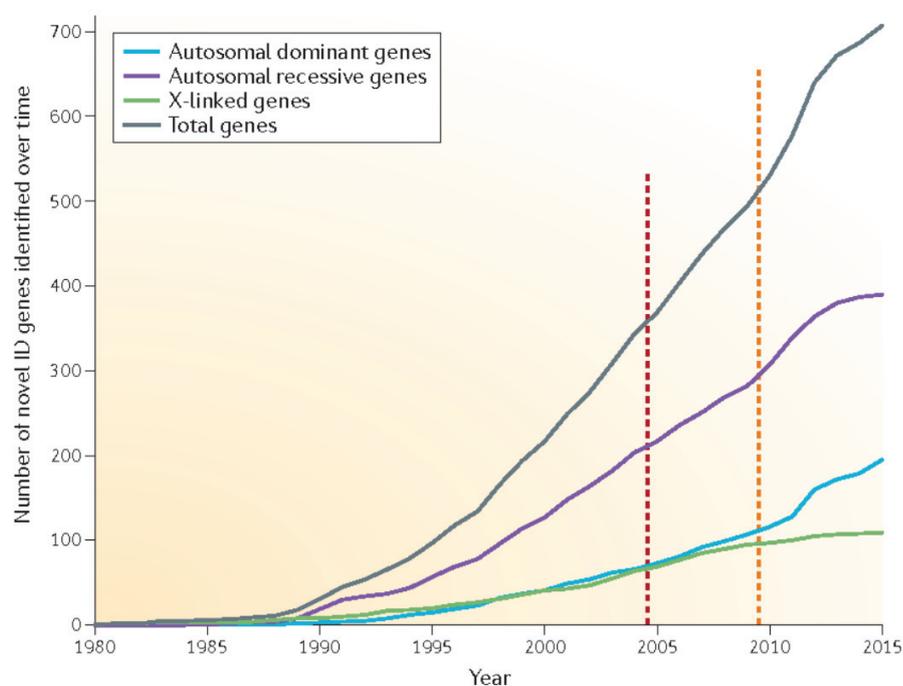


Figure 3 : Nombre de gènes liés à la DI isolée ou syndromique en fonction des années et du mode de transmission (1)

L'accès à l'exome a permis également l'émergence du processus de « *reverse phenotyping* ». En effet, l'identification d'un variant pathologique dans un gène responsable d'une pathologie permet a posteriori de retrouver certains symptômes présentés par le patient et ainsi de conforter le diagnostic moléculaire. De plus, certains variants pathologiques dans des gènes connus pour une pathologie donnée chez un patient présentant un phénotype atypique a permis d'élargir la liste des symptômes décrits dans la pathologie.

2.1.5 Le cas particulier de la Déficience Intellectuelle

2.1.5.1 Définition de la Déficience intellectuelle

Au sein des AD, la Déficience Intellectuelle (DI) tient une part importante puisqu'elle touche 1.5 à 2 % de la population. La DI est définie comme un déficit des compétences intellectuelles avec un quotient intellectuel (QI) <70. Les difficultés de raisonnement logique et d'abstraction retentissent fortement sur les apprentissages scolaires et ensuite sur les capacités d'adaptation dans la vie quotidienne à l'âge adulte. La DI est classée selon la Classification internationale des Maladies – 9 (CIM-9) en fonction de son degré de sévérité déterminé par le QI:

- Retard mental léger : QI 50/55 – 70
- Retard mental moyen : QI 35/40 – 50/55
- Retard mental grave : QI 20/25 – 35/40
- Retard mental profond : QI <20/25

Les classifications DSM-V et de *l'American Association on Intellectual and developmental Disabilities* (AAIDD) complète cette définition par une description typologique de la déficience ainsi que l'intensité des besoins de soutien pour mesurer le degré de DI (Figure 4).

| Gravité | Domaine conceptuel | Domaine social | Domaine pratique |
|----------------|---|---|---|
| Léger | La personne a une manière plus pragmatique de résoudre des problèmes et de trouver des solutions que ses pairs du même âge... | La personne a une compréhension limitée du risque dans les situations sociales ; a un jugement social immature pour son âge... | La personne occupe souvent un emploi exigeant moins d'habiletés conceptuelles... |
| Modéré | D'ordinaire, la personne a des compétences académiques de niveau primaire et une intervention est requise pour toute utilisation de ces compétences dans la vie professionnelle et personnelle... | Les amitiés avec les pairs tout-venant souffrent souvent des limitations vécues par la personne au chapitre des communications et des habiletés sociales... | Présence, chez une minorité importante, de comportements mésadaptés à l'origine de problèmes de fonctionnement social... |
| Grave | La personne a généralement une compréhension limitée du langage écrit ou de concepts faisant appel aux nombres, quantités, au temps et à l'argent... | Le langage parlé est relativement limité sur le plan du vocabulaire et de la grammaire... | La personne a besoin d'aide pour toutes les activités de la vie quotidienne, y compris pour prendre ses repas, s'habiller, se laver et utiliser les toilettes... |
| Profond | La personne peut utiliser quelques objets dans un but précis (prendre soin de soi, se divertir)... Des problèmes de contrôle de la motricité empêchent souvent un usage fonctionnel... | La personne peut comprendre des instructions et des gestes simples... | La personne dépend des autres pour tous les aspects de ses soins physiques quotidiens, pour sa santé et pour sa sécurité, quoiqu'elle puisse participer à certaines de ces activités... |

Figure 4 : Critères de gravité de la DI selon le DSM-V

En France, 10 à 20 personnes sur 1000 sont atteintes de DI légère et 3 à 4 sur 1000 de DI sévère. Il existe un sexe ratio de 1.2 à 1.9 avec une prévalence plus élevée chez les garçons (2).

La DI peut être associée à de nombreux autres troubles neurodéveloppementaux comme des Troubles du Spectre Autistique (TSA), un Trouble du Déficit de l'Attention et Hyperactivité ou des troubles du comportement. Les TSA sont un trouble du neurodéveloppement caractérisés par des difficultés à percevoir les émotions et les intentions d'autrui, à communiquer, à comprendre les codes sociaux et à s'ajuster sur le plan relationnel. Dans environ 2 tiers des cas, les capacités intellectuelles des personnes autistes sont préservées et dans 1 tiers des cas, une DI est associée. En effet, les anomalies congénitales sont 10 fois plus fréquentes chez les patients atteints de DI en lien ou non avec l'étiologie de la DI ; on parle alors de DI syndromique (2).

Des outils psychométriques récents et validés permettent de confirmer le diagnostic de DI. Les capacités de raisonnement sont évaluées à l'aide de plusieurs échelles de Wechsler comme la WPPSI-IV (*Wechsler Prescholl and Primary Scale of Intelligence- IV*) et la WISC-IV (*Wechsler Intelligence Scale for Children-IV*) pour les enfants ou la WAIS-IV (*Wechsler Adult Intelligence Scale-IV*) pour les adultes. Des épreuves non verbales, adaptées pour certains patients présentant une DI ou un TSA, offrent une alternative comme les Matrices Progressives de Raven. Les difficultés adaptatives peuvent être mesurées grâce à l'échelle VINELAND-II et complète l'évaluation du QI (2).

2.1.5.2 Etiologie de la déficience intellectuelle

Concernant la DI syndromique ou non syndromique, il est considéré que la moitié des DI seraient d'origine génétique (2). Les autres causes de DI sont énumérées dans le tableau 1 et comprennent notamment les complications de la prématurité, les causes environnementales (dont la première connue est l'alcoolisation fœtale) et 35 à 40% restent inexplicables. Les causes génétiques considérées séparément sont expliquées par une multitude de causes qui, prises séparément, concernent moins de 1% des cas. Deux pathologies font exception : la trisomie 21 et le syndrome de l'X-fragile qui sont respectivement les causes chromosomiques et monogéniques les plus fréquentes dans la DI. Les causes génétiques de DI syndromique ou non sont les anomalies chromosomiques détectées par caryotype standard, les variations du nombre de copies (CNV = Copy Number Variation) qui correspondent à des pertes ou des gains > à 1kb détectées par Analyse Chromosomique sur Puce à ADN (ACPA) dont le seuil de résolution est d'environ 40 kb, les variants monogéniques et les anomalies génétiques non mendéliennes (maladie liée à l'empreinte parental, maladies mitochondriales). Les modes de transmission de ces pathologies sont variables : *de novo*, autosomique dominant, autosomique récessif, récessif lié à l'X, dominant lié à l'X, mitochondriaux et lié à l'empreinte parentale.

Plusieurs facteurs peuvent influencer le degré de DI comme le contexte socio-économique puisque concernant la DI légère, la prévalence est plus faible dans la catégorie socio-économique élevée. Ce n'est plus le cas lorsque l'on considère la DI sévère où le contexte socio-économique n'influe pas la prévalence (2).

Etiologies des déficiences intellectuelles

| | |
|---|--------|
| Complications de la prématurité | 5% |
| Causes environnementales | 13% |
| Anomalies chromosomiques | 15% |
| Maladies métaboliques | 8% |
| Syndromes reconnaissables | 2% |
| Di liés au chromosome X | 10% |
| Autres maladies monogéniques connues | 10% |
| Di idiopathiques | 35-40% |

Tableau 1 : Etiologie des DI (2)

2.1.5.3 Stratégies d'exploration étiologique d'une DI

Le diagnostic d'une DI syndromique ou non repose en première intention sur une exploration complète clinique avec la recherche des antécédents familiaux, l'histoire de la maladie, l'examen clinique morphologique et neurologique du patient associé des examens d'imagerie médicale, ainsi que des analyses biologiques afin d'orienter le diagnostic. Une IRM cérébrale est demandée devant un retard moteur important, des signes neurologiques, une épilepsie ou une anomalie du périmètre crânien. Dans 30% des cas, l'IRM retrouve une anomalie cérébrale qui apporte une orientation diagnostic dans 0.2 à 3.9%. Concernant le bilan métabolique sans signe d'appel évocateur, les recommandations ne sont pas clairement tranchées devant la rentabilité faible (1%) et le cout élevé de ces examens. Cependant, le caractère potentiellement traitable, le conseil génétique ainsi que le dépistage précoce de ces pathologies sont des arguments en faveur de la réalisation d'un bilan métabolique (2).

En cas d'orientation diagnostique vers un syndrome connu ou un cadre syndromique, une analyse SANGER si le gène est connu ou un « petit » panel de gènes si l'hétérogénéité génétique est limitée (comme par exemple le syndrome de Cornelia de Lange) peuvent être réalisés.

En l'absence d'orientation, une ACPA peut mettre en évidence des remaniements déséquilibrés ou CNV qui permet dans environ 15 à 20 % des cas de déficiences

intellectuelles syndromiques ou non de poser un diagnostic (3). En effet, depuis une dizaine d'année, l'ACPA a remplacé le caryotype standard et/ou l'hybridation en fluorescence in situ (FISH) pour la détection des remaniements déséquilibrés. Cependant, l'ACPA ne permet pas de mettre en évidence des remaniements équilibrés et seul le caryotype standard et/ou la FISH permettent d'identifier ce type de remaniement ou de localiser la position d'une région dupliquée ou insérée. L'ACPA est également limité en termes de résolution puisqu'elle ne peut mettre en évidence que des CNV supérieur à environ 40 kb. Il est donc difficile avec cette méthode de détecter des CNV concernant un ou plusieurs exons. De plus, une analyse d'expansion de triplet du gène *FRAXA* responsable du syndrome de l'X fragile est systématiquement réalisée, quel que soit le sexe, devant la fréquence de cette pathologie au sein des DI et à cause de son hétérogénéité clinique rendant parfois le syndrome difficilement reconnaissable (non réalisé en cas de syndrome polymalformatif). En cas de négativité de ces examens, le séquençage haut-débit offre actuellement 2 alternatives :

- La réalisation d'un panel de gènes impliqués dans la DI comme le panel DI44 ou des panels plus exhaustifs de plus de 500 gènes de DI.
- Le séquençage de l'exome clinique (séquençage des gènes OMIM uniquement) ou l'exome complet avec un rendement diagnostique estimé à 36% selon la méta-analyse la plus récente (4).

L'algorithme du choix des différentes analyses génétique possible devant une AD avec DI est représentée dans la figure 5 (2).

La mise en place des panels DI est une solution transitoire en attendant l'arrivée du séquençage de l'exome pour le diagnostic clinique pour certains laboratoires. Si ces analyses sont négatives, alors le séquençage du génome entier peut être envisagé. Devant les AD sans DI et sans orientation diagnostique, un exome d'emblée est recommandé. En l'absence de diagnostic, une réévaluation diagnostique à intervalle régulier est recommandée avec une consultation en génétique clinique et une relecture des données de séquençage.

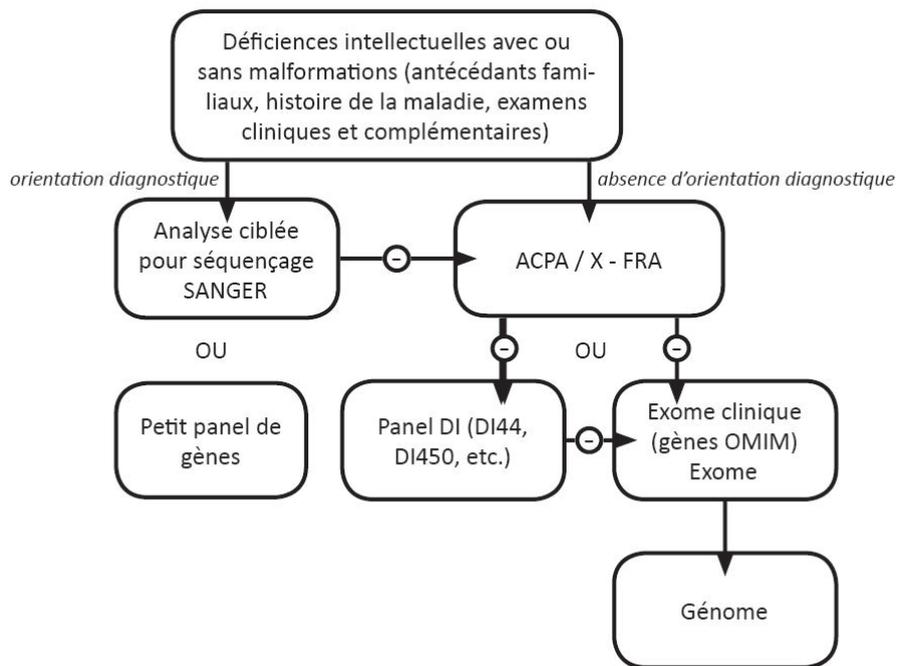


Figure 5 : Arbre décisionnel pour la prescription d'analyse génétique devant une DI syndromique ou non

2.1.5.4 Taux diagnostique du séquençage de l'exome dans les AD avec DI

Plusieurs dizaines de séries d'exomes appliqués au diagnostic concernant des patients présentant des troubles du neurodéveloppement ont été rapportées dans la littérature. Une des premières grandes séries d'exomes cliniques rapportée par Yang et al, en 2014 (5) chez 2000 patients présentant principalement un trouble du neurodéveloppement présentait un taux diagnostique de 25%. Dans cette cohorte, 51% des variants pathogènes étaient *de novo* et 4.6% des patients présentaient une DNS. Une autre grande série d'exome clinique chez 3040 patients rapportée par Retterer et son équipe en 2016 (6) retrouve un taux diagnostique de 28.8% avec 42.5% de mutations *de novo* et 6% de DNS. En France, Thevenon et al, 2016 (7) ont rapporté une série de 43 patients avec un taux diagnostique de 33%. Les principales séries d'exomes cliniques chez des patients présentant des troubles du neurodéveloppement sont répertoriées dans le tableau 2. Cependant, il est difficile de comparer leur taux diagnostique puisque que le phénotype des patients diffère selon les cohortes ainsi que les stratégies de séquençage (exome solo VS exome trio).

| Référence | Nombre de patients | Phénotype | Taux diagnostique | Mutation <i>De Novo</i> | >1 diagnostics | Découvertes secondaires |
|---|--------------------|---|-------------------|-------------------------|----------------|-------------------------|
| De Ligt et al., 2012 (8) | 100 | DI sévère | 16% | 81% | 0 | ND |
| Rauch et al., 2012 (9) | 51 | DI sévère non syndromique | 50% | 96% | 0 | ND |
| Yang et al., 2013 (10) | 250 | Pathologies neurologiques | 25% | 53% | 1,6% | 12% |
| Hamdan et al., 2014 (11) | 41 | DI modérée à sévère non syndromique | 29% | 100% | 0 | ND |
| Srivastava et al., 2014 (12) | 78 | Troubles neurodéveloppementaux | 41% | 59% | 1,3% | ND |
| Lee et al., 2014 (13) | 814 | Variable, 37% troubles du neurodéveloppement | 26% | 50% | ND | 5% |
| Yang et al., 2014 (5) | 2000 | Principalement des troubles neurodéveloppement | 25% | 51% | 4,6% | 4,60% |
| Wright et al., 2015 (14) | 1133 | 87% DI, troubles neurodéveloppementaux | 27% | 65% | ND | ND |
| Retterer et al., 2016 (6) | 3040 | Variable, 36% pathologies neurologiques et DI | 28,80% | 42,50% | 1% | 6% |
| Nolan et Carison, 2016 (15) | 50 | 88% troubles neurodéveloppementaux | 47% | 33% | 0 | 10% |
| Monroe et al., 2016 (16) | 17 | DI/troubles du neurodéveloppement | 29% | ND | 0 | ND |
| Tarailo-Graovac et al., 2016 (17) | 41 | Pathologies métaboliques/DI | 28/41% | 39% | 18% | 44% |
| Stark et al., 2016 (18) | 80 | Anomalies du développement, 74% maladies neuro-métaboliques | 58% | 36% | 1 famille | ND |
| Kuperberg et al., 2016 (19) | 57 | Troubles neurodéveloppementaux, pathologies neurologiques | 49% | 50% | 0 | ND |
| Thevenon et al., 2016 (7) | 43 | Troubles du neurodéveloppement | 33% | 58% | 0 | 1 patient |
| Trujillano et al., 2017 (20) | 1000 | Variable, 77% troubles neurodéveloppementaux et pathologies neurologiques | 31% | 20% | 1% | ND |
| Vissers et al., 2017 (21) | 150 | DI et pathologies neurologiques | 29% | 77% | 0 | ND |
| Evers et al., 2017 (22) | 72 | Troubles du neurodéveloppement, maladies métaboliques, dystonie | 35% | 43% | 1 patient | ND |
| Chérot et al., 2018 (23) | 216 | Troubles du neurodéveloppement | 25,90% | 69% | 0 | ND |
| Cordoba et al., 2017 (24) | 40 | Troubles neurodéveloppementaux, pathologies neurologiques | 40% | 56% | 0 | ? |
| Palmer EE et al., 2018 (25) | 32 | Encéphalopathie épileptique | 50% | 21% | 0 | ? |
| Kim SH et al., 2019 (26) | 108 | Troubles du neurodéveloppement | 38% | ND | 0 | ND |
| Monies et al., 2017 (27) | 1019 | Variable | 43% | 27% | 0,5% | ND |
| Hu X et al., 2018 (28) | 1323 | Variable (25,9% pathologies neuromusculaires) | 28,50% | 40% | ND | ND |
| Snoeijs-Schouwenaars FM et al., 2019 (29) | 100 | Epilepsie et DI | 25% | 86% | 0 | 0 |

Tableau 2 : Résultats des séries d'exomes cliniques chez des patients présentant des troubles du neurodéveloppement

2.2 Le Séquençage à Haut Débit

2.2.1 Les différentes technologies de séquençage à Haut Débit

Les principales technologies de séquençage à Haut Débit (SHD) actuellement sont :

- Le pyroséquençage basé sur l'émission d'un signal photoluminescent secondaire à la libération d'un pyrophosphate lors de l'incorporation de nucléotides (société Roche 454)
- Le séquençage par détection des variations du pH secondaire à la libération d'ions H⁺ lors de l'incorporation de nucléotides (société Thermo-Fisher-Ion torrent)
- Le séquençage par synthèse basé sur l'incorporation de nucléotides modifiés fluorescents (société Illumina)

Concernant ces différentes technologies, le principe général reste le même malgré des spécificités propres à chacune. Chaque fragment d'ADN d'intérêt est sélectionné puis amplifié et ensuite le brin complémentaire est synthétisé. Lors du séquençage, à chaque incorporation d'un nucléotide, un signal est détecté (un signal lumineux dans la technologie d'Illumina, une variation du pH dans la technologie Thermo-Fisher-Ion Torrent). La figure 6 montre les différentes technologies de séquençage haut-débit. La technologie utilisée pour cette étude est la technologie Illumina de séquençage par synthèse qui est actuellement la technologie de NGS dominante sur le marché. Nous expliquerons les différentes étapes de séquençage en prenant comme exemple cette technologie dans le chapitre suivant.

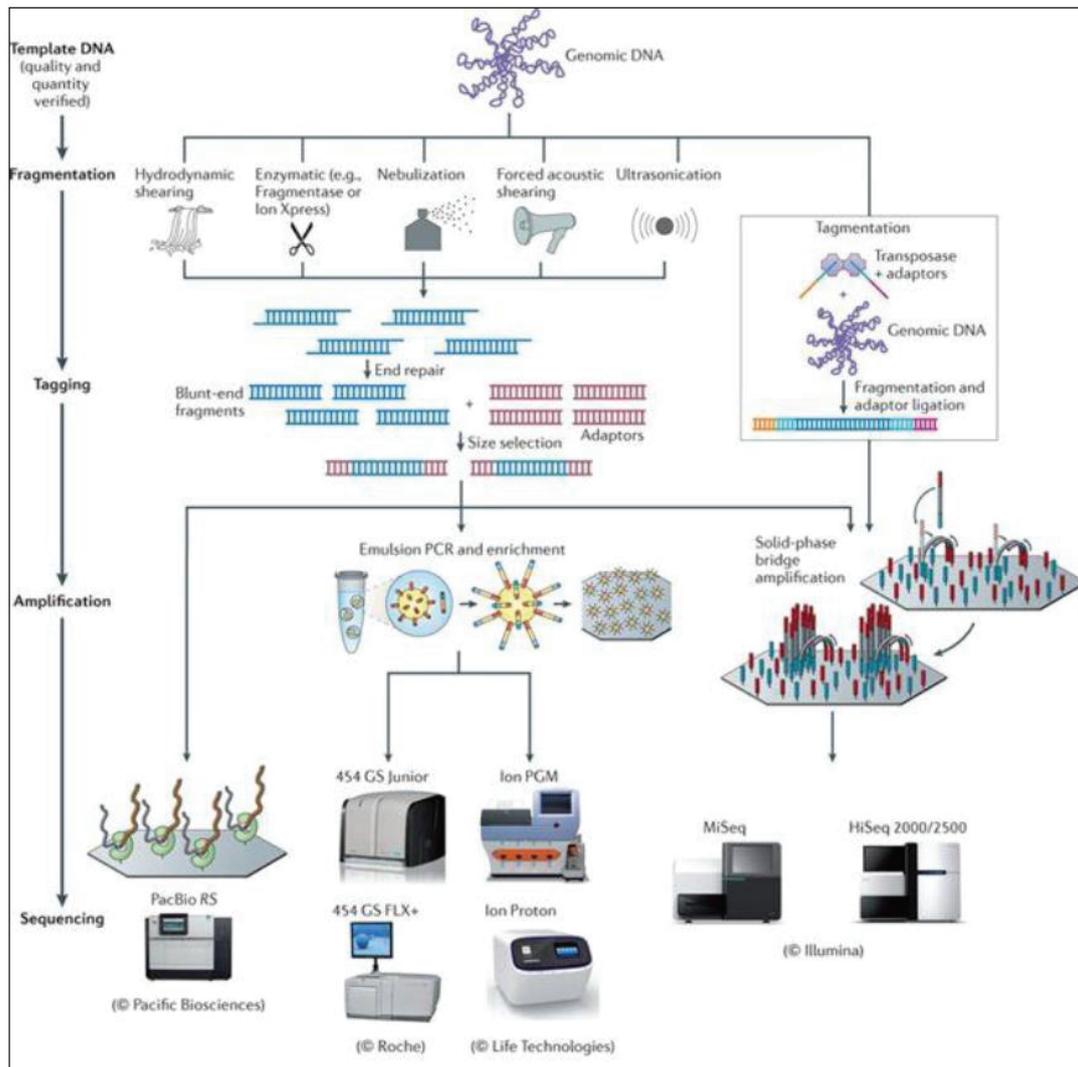


Figure 6 : Les principales technologies de séquençage à haut débit

2.2.2 Les différentes étapes du séquençage haut-débit : Exemple de la technologie d'Illumina de séquençage par synthèse

2.2.2.1 Génération des données de séquençage

Le séquençage haut-débit par synthèse d'Illumina permet de générer massivement des données de séquençage obtenues suites à la détection d'un signal lumineux lors de l'incorporation successive de nucléotides. Le séquençage s'effectue sur l'appareil NextSeq500 d'Illumina et comprend 3 grandes étapes communes aux différentes technologies.

- Préparation des librairies
- Amplification clonale
- Séquençage

2.2.2.1.1 Préparation des librairies

Cette étape comporte elle-même plusieurs phases : La préparation des librairies se fait par une méthode de capture avec la technologie d'Illumina. Une méthode d'enrichissement par PCR existe également.

- Une **étape de fragmentation mécanique par sonication** de l'ADN qui permet d'obtenir des fragments d'ADN de tailles précise en les soumettant à des ultra-sons d'une fréquence spécifique. Une fragmentation par méthode enzymatique existe également.
- Une **étape de multiplexage** des échantillons qui permet de mélanger l'ADN de plusieurs patients par l'ajout d'un « code-barre » ou index de part et d'autre du fragment d'ADN. Ce double indexage ou « Dual Index Library » correspond à 2 séquences spécifiques qui vont être associées à un échantillon. Elles seront séquencées en même temps que les séquences d'intérêt ce qui permettra après l'analyse bioinformatique de réattribuer les séquences à chaque échantillon (Figure 7).
- Une **étape d'enrichissement des régions d'intérêt par capture** qui consiste à sélectionner les régions d'intérêt grâce à des amorces complémentaires aux séquences cibles. Pour le séquençage de l'exome, les régions capturées sont les exons et les jonctions intron-exons de l'ensemble des gènes du génome. Des kits fournis par des sociétés de séquençage sont proposés avec l'ensemble des amorces nécessaires pour la capture de l'exome comme par exemple le kit MedExome de Roche (30) .



Figure 7 : Préparation de la librairie : double indexage avec séquence P5 - index 1 et séquence P7 - index2

2.2.2.1.2 Amplification clonale

Lors de cette étape, les fragments d'ADN sont amplifiés sur un support solide (*Flow Cell*) de manière clonale et isolés en « cluster ». Avant cette étape, lors de la préparation de la librairie, des séquences complémentaires aux adaptateurs fixés sur le support ont été ajoutées aux extrémités 5' (P5) et 3' (P7) des fragments d'ADN à amplifier permettant ainsi leur accrochage au support solide (Figure 7). A chaque cycle, le brin complémentaire est synthétisé, l'ADN double brin nouvellement formé est dénaturé. Chaque brin fixé forme un pont en s'hybridant localement avec l'amorce complémentaire de l'autre extrémité du fragment d'ADN. Le brin complémentaire est ensuite synthétisé. On obtient donc des brins sens et anti-sens du fragment d'intérêt. Cette opération est répétée un grand nombre de fois et permet ainsi d'obtenir des « clusters » de fragments d'ADN clonaux (Figure 8).

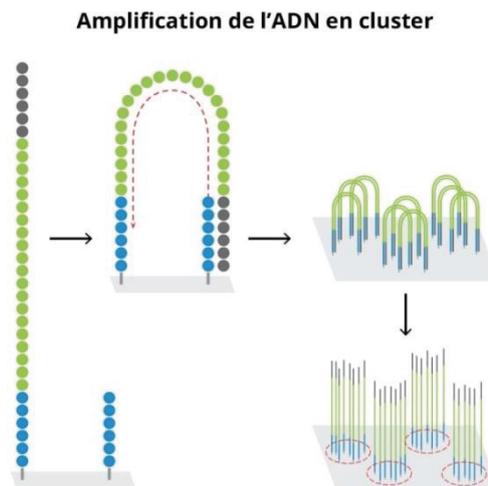


Figure 8 : Amplification clonale en « cluster » par la technologie Illumina

2.2.2.1.3 Séquençage par synthèse

L'étape de séquençage se produit au sein du système NextSeq 500 d'Illumina. Cette étape est basée sur la détection d'un signal luminescent à l'incorporation de chaque nucléotide lors de la synthèse du brin complémentaire. A la fin de l'étape d'amplification clonale, toutes les séquences « *reverse* » sont éliminées du support solide. Les amorces se lient aux séquences « *forward* » (Rd1SP) et la polymérase ajoute un nucléotide associé à un fluorochrome dont la longueur d'onde est spécifique et détectée à chaque incorporation par un système de caméra. Cette succession de cycle dit de « *wash and clean* » permet donc la synthèse du brin complémentaire base par base. Deux longueurs d'ondes sont utilisées : le rouge et le vert. Le nucléotide T est associé au vert, le C au rouge, le A au vert et au rouge et le G n'est associé à aucune fluorescence (Figure 9). Une fois que le brin complémentaire est séquencé, il est éliminé puis l'index 1 est séquencé et éliminé. Il y a ensuite formation du pont avec l'extrémité 3' de l'ADN qui se fixe à l'adaptateur sur le support solide. La polymérase synthétise ensuite le brin complémentaire. L'extrémité 3' de l'ADN est bloquée à la « *Flowcell* », le brin « *forward* » est éliminé puis l'index 2 et ensuite le brin « *reverse* » peuvent être séquencés (Figure 10). Concernant le SHD-E, les fragments d'ADN font 150pb et sont donc lus dans les deux sens. On parle de séquençage « *paired-end* » 2x150pb. Le système NextSeq500 permet de séquencer sur une « *flowcell* » 12 exomes (soit 4 trios) en « *paired-end* » 2x150pb avec une profondeur moyenne de 50x en 29 heures ce qui correspond à environ 800 millions de lectures appariées et un débit de 100 à 120 Gb. (<https://www.illumina.com>)

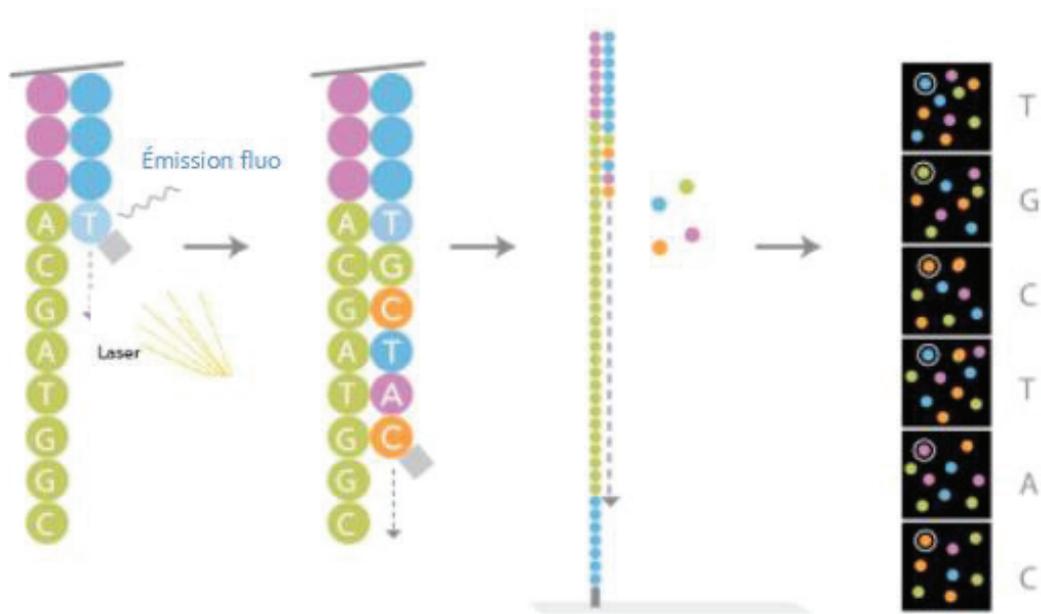


Figure 9 : Séquençage par synthèse : émission d'une fluorescence par incorporation d'un nucléotide

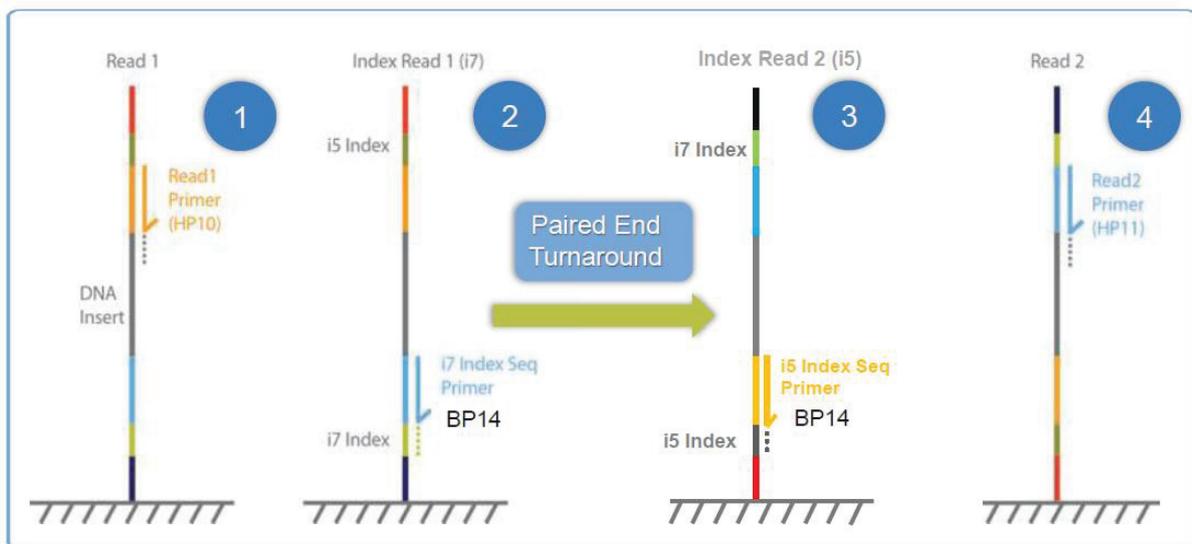


Figure 10 : Séquençage par synthèse des brins « forward » et « reverse » et des index

2.2.2.2 Analyse bio-informatique

Dans un second temps, une analyse bioinformatique des données de séquençage est nécessaire. Elle comprend plusieurs étapes réalisées par des logiciels de traitement automatique des données successifs qui sont regroupé au sein d'un pipeline bioinformatique.

- **L'appel de base ou « Base calling »** : cette étape correspond au séquençage lui-même c'est-à-dire à la traduction du signal lumineux en base jusqu'à l'obtention de la séquence complète de la lecture. Cette étape se fait au sein même du séquenceur.
- **Démultiplaxage** : lors de cette étape, chaque lecture est associée à un patient grâce au séquençage de l'index spécifique.
- **Filtrage et analyse de qualité des séquences** : cette étape permet d'éliminer les lectures de mauvaise qualité.
- **Alignement sur le génome de référence** : lors de cette étape, les différentes lectures obtenues sont alignées sur le génome de référence.
- **Analyse de la profondeur et de la couverture du séquençage**. Ces 2 notions sont très importantes pour appréhender l'efficacité du séquençage. La profondeur de lecture correspond au nombre de fois où une position génomique est lue et est noté par exemple 10x si celle-ci est lue dans 10 lectures différentes. La couverture de lecture correspond au nombre de bases séquencées des régions d'intérêt par rapport au nombre de bases totales des régions d'intérêt. Ce résultat est donné en pourcentage pour une profondeur donnée. Ces 2 notions sont représentées schématiquement sur la figure 11.
- **L'appel de variant ou « variant calling »** : cette étape permet de mettre en évidence les variations de séquence (substitution d'une base, délétion, insertion d'une ou plusieurs bases) par rapport au génome de référence. Concernant le l'exome, environ 30 000 variants sont détectés par patient.
- **Annotation** : cette étape correspond à l'attribution d'un score de pathogénicité du variant allant de 1 à 5 (1=bénin, 2=probablement bénin, 3=signification inconnue, 4=probablement pathogène et 5=pathogène). Les algorithmes d'annotation se basent sur de nombreuses informations comme le gène impliqué, le type de mutation (faux-sens, non-sens, délétion, insertion, variant affectant un site d'épissage), la localisation du variant (exon, intron, région UTR), la fréquence du variant dans la population générale, la présence du variant dans les bases de données de patients, les scores des logiciels de prédiction de pathogénicité *in-silico* qui reflète l'impact potentiel d'un variant sur la structure de la protéine, les données de la littérature, le mode d'hérédité du variant. Beaucoup de ces variants sont hérités et/ou sans conséquences cliniques et/ou biologiques. La description

précise du phénotype du patient est indispensable pour interpréter correctement les variants observés (30).

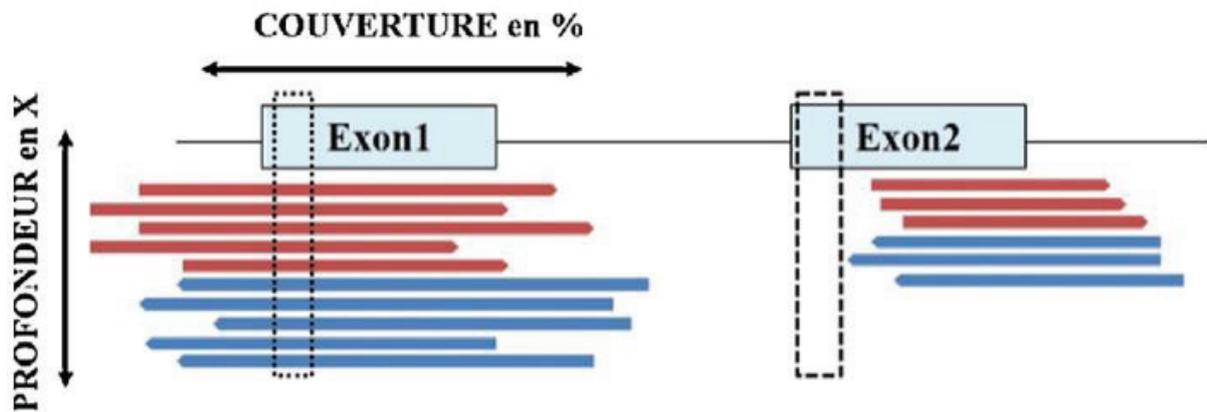


Figure 11 : Représentation schématique de la notion de profondeur et de couverture de séquençage.

Chaque ligne représente une lecture indépendante, en rouge les lectures dans un sens (brin d'ADN forward) et en bleues les lectures dans l'autre sens (brin reverse). La profondeur de lecture (en x) correspond au nombre de lectures indépendantes obtenues pour une position donnée. La couverture (en %) correspond au nombre de bases lues dans la région d'intérêt sur le nombre total de base de la région d'intérêt. L'exon 1 ici est couvert à 100% (pointillés fins) avec une profondeur à 10x. L'exon 2 n'est pas couvert à 100% (pointillés larges non couvert). La partie terminale de l'exon 2 est couvert à 6x seulement.

2.2.3 Les différentes stratégies de Séquençage à Haut Débit

Actuellement, les 3 principales approches pour diagnostiquer les anomalies de développement avec ou sans déficiences intellectuelles sont :

1. Les panels de gènes
2. Le séquençage de l'exome
3. Le séquençage du génome (30)

2.2.3.1 *Les panels de gènes*

L'analyse par « panel de gènes » correspond au séquençage en parallèle d'un groupe de gènes dont les phénotypes associés sont similaires ou chevauchants. Cette approche concerne donc les maladies présentant une forte hétérogénéité génétique c'est à dire pour lesquelles des variants pathogènes ont été retrouvés dans plusieurs gènes différents. Dans ces panels, seules les parties codantes des gènes (exons) et les jonctions intron-exons, qui contiennent les sites d'épissage de l'ARN, sont séquencés. Cette technique offre un haut-débit d'analyse grâce à une méthode de multiplexage qui permet de séquencer plusieurs patients en même temps. Aujourd'hui, entre 800 et 1000 gènes sont connus pour causer des DI syndromiques ou non. Ces gènes peuvent être regroupés au sein d'un panel ce qui permet une analyse génétique plus exhaustive et plus rapide que ne le permettait le séquençage SANGER séquentiel de plusieurs gènes ciblés qui pouvait prendre plusieurs années. Quarante-quatre gènes de DI syndromique ou non sont plus fréquemment retrouvés chez les patients et sont regroupés au sein d'un plus petit panel moins exhaustif mais moins cher et plus rapide appelé « panel DI44 ». L'utilisation de ces panels est déjà bien installée dans les laboratoires de diagnostic génétique et ont montré leur efficacité que ce soit les panels pour la déficience intellectuelle aussi bien que les panels dans le domaine oncologique, les maladies neuromusculaires, les surdités ou encore les maladies dermatologiques d'origine génétique... Cette analyse est une approche ciblée des gènes et prévient donc des découvertes secondaires. De plus, le nombre de paires de bases séquencées est bien moindre que pour le séquençage de l'exome ou le séquençage du génome entier, ainsi une meilleure profondeur de lecture est obtenue et permet donc la détection de variant en mosaïque. Les inconvénients de ces panels sont qu'ils ne permettent pas la découverte de nouveaux gènes, que certains gènes restent mal couverts bien que la couverture reste meilleure que celle d'un exome et que le design du panel doit être mis régulièrement à jour en fonction des nouveaux gènes découverts.

2.2.3.2 *Le séquençage de l'exome*

Le séquençage de l'exome correspond à l'analyse de l'ensemble des exons des gènes du génome entier. Ceci représente environ 1 à 2% du génome et correspond aux régions codant pour des protéines. Les jonctions intron-exons (sites d'épissages) sont également séquencées. Cette stratégie repose sur le fait qu'environ 90% des pathologies génétiques mendéliennes connues aujourd'hui sont dues à des variants pathologiques présents dans la séquence codante.

Plusieurs stratégies de séquençage de l'exome existent : l'exome clinique qui correspond au séquençage des gènes OMIM, donc les gènes sont déjà bien définis en pathologie humaine et l'exome « complet » qui correspond au séquençage des exons des 20 000 gènes du génome. Cette dernière stratégie permet donc de mettre en évidence de nouveaux gènes impliqués dans les AD. Cette approche est plus exhaustive que celle des panels de gènes. Le séquençage de l'exome est à privilégier devant l'absence d'orientation diagnostique et pour des pathologies présentant une forte hétérogénéité clinique et génétique. Ceci est le cas pour les AD avec DI en générale où il existe de très nombreux gènes déjà décrits et d'autres non connus à ce jour. En terme quantitatif, le séquençage de l'exome d'un patient correspond au séquençage d'environ 3 Mb et environ 30 000 variants sont mis en évidence après l'analyse bioinformatique. Ces chiffres montrent bien la complexité de l'analyse comprenant un traitement bioinformatique des données long et une interprétation des variants délicates. Comme pour les panels de gènes, certaines régions de l'exome ne sont pas couvertes. Ceci est inhérent à la méthode de capture lors de la génération des données de séquençage. De plus, il est également important de noter que les expansions de triplets ne sont pas détectables par cette technique. Les données du séquençage de l'exome permettent également de mettre en évidence des anomalies chromosomiques grâce à des logiciels spécifiques qui analysent les profondeurs de lectures des exons.

Une autre problématique soulevée par cette approche est celle des découvertes dites secondaires (DS) c'est-à-dire qui n'ont rien à voir avec le diagnostic initial comme par exemple la mise en évidence de prédispositions à certains cancers ou le statut de porteur hétérozygote de certaines pathologies récessives. La prise en charge de ces DS est un enjeu éthique et de santé publique important avec l'avènement du SHD et sera détaillé dans un autre paragraphe.

2.2.3.3 Le séquençage du génome

Le séquençage du génome correspond au séquençage des régions codantes et non codantes de l'ADN génomique. Cette approche est envisageable lorsque le séquençage de l'exome est négatif mais peut également être choisie à la place du séquençage de l'exome si la plateforme de génétique le permet. Des variants peuvent alors être retrouvés dans des régions non couvertes par l'exome puisqu'il n'y a pas d'étape de capture. D'autres variants peuvent concerner les régions introniques et régulatrices de certains gènes suspectés. La quantité de données engendrées par le séquençage du génome est bien supérieure à celle du séquençage de l'exome avec le séquençage de 3 Gb et l'analyse d'environ 3 millions de variants. Ceci soulève des problèmes quant au stockage des données et l'interprétation complexe des variants. La profondeur de séquençage est bien inférieure que pour les panels ou le séquençage de l'exome devant la quantité de données générées et le débit limité des séquenceurs. Cependant, le séquençage du génome permet de détecter des variants de structure (VS) équilibrés avec la possibilité de mettre en évidence les points de cassures à la base près ce que ne permet pas l'ACPA qui ne met en évidence que les déséquilibres génomiques. À terme, le séquençage du génome devrait remplacer l'ACPA et le séquençage de l'exome puisqu'il permet à la fois l'analyse des variants, des CNV et des VS. Cette approche par séquençage du génome est d'ailleurs celle qui a été choisie lors de la mise en place des 2 plateformes de SHD du plan France médecine génomique. Le tableau 3 résume les principes, les avantages et les inconvénients de chaque approche.

| Type d'analyse génétique | Séquençage SANGER | Panel de gènes | Séquençage de l'exome | Séquençage du génome |
|---------------------------|--|---|---|--|
| Description | Analyse ciblée des exons d'un gène connu | Analyse ciblée des exons d'un ensemble de gène | Analyse ciblée de l'ensemble des exons du génome | Analyse de l'ensemble du génome |
| Indication | Signes cliniques spécifiques orientant vers une pathologie précise | Maladies génétiquement hétérogène | Absence d'orientation diagnostique, signes cliniques non spécifiques, hétérogénéité génétique importante | Idem exome |
| Régions séquencées | Exon du gène cible et jonction intron-exon (sites d'épissage) | Ensemble des exons de gènes causaux bien définis et jonction intron-exon | Séquençage des exons et des jonctions intron-exons des 20 000 gènes du génome dont les 5000 gènes connus impliqués en pathologie humaine | Séquençage de l'ensemble du génome |
| Avantages | Gold standard Pas de donnée secondaire | Analyse de plusieurs gènes et plusieurs patients en même temps Bonne couverture des bases analysées Peu de découvertes secondaires Profondeur élevée (détection de mosaïque+++) Détection des CNV | Interrogation de l'ensemble des gènes Découvertes potentielles de nouveaux gènes | Analyse de l'ensemble des gènes (exon et intron) ainsi que des régions non codantes Meilleure couverture Découvertes de nouveaux gènes Détections des CNV et VS |
| Inconvénients | Analyse séquentielle (exon par exon) Limité à 700-800 pb par séquence | Mise à jour du design du panel régulière en fonction des nouveaux gènes décrits Pas de mise en évidence de nouveaux gènes Mauvaise couverture de certains gènes | Certains exons non couverts Cout élevé, analyse bioinformatique longue Découvertes secondaires fréquentes Détection de CNV possible mais difficile | Faible profondeur de lecture Analyse complexe et interprétation difficile des régions non codantes Découvertes secondaires Cout élevé, analyse bioinformatique longue |

Tableau 3 : Avantages et inconvénient des différentes stratégies de séquençage

2.2.4 Analyse des CNV sur les données du séquençage de l'exome

Les variations du nombre de copie ou *Copy Number Variation* (CNV) sont des variations génétiques incluant des gains (duplication, triplication) ou des pertes (délétions) et sont une source importante de variations à la fois bénignes et pathologiques du génome. Ils représentent environ 4.8 à 9.5% du génome humain (31). Les CNV pathogènes, le plus souvent *de novo*, représente 15 à 20% des causes génétiques dans les troubles du neurodéveloppement et les anomalies congénitales. L'ACPA permet de mettre en évidence ces CNV et est aujourd'hui l'exploration génétique de première intention réalisée dans ce cadre nosologique (2). Les puces à ADN contiennent des centaines de milliers d'oligonucléotides et permettent une résolution d'environ 40kb dans la détection de ces CNV. Cependant, certaines délétions ou duplications intragéniques sont plus petites et ne sont pas détectées par cette technique. Le séquençage de l'exome et du génome permettent à la fois de détecter les variations nucléotidiques mais aussi les CNV notamment de petite taille et les variant de structure ou *Structural Variation* (SV) à l'échelle du génome. Ces petits CNV s'avèrent très fréquents ce qui a amené à redéfinir la taille minimale d'un CNV à des délétions et duplications à 50pb (32). L'implémentation de l'analyse des CNV au séquençage de l'exome permettrait à l'avenir de diminuer le nombre d'explorations génétiques et d'analyser en une fois les variations nucléotidiques et les CNV d'un patient en une seule technique.

La détection des CNV sur les données de séquençage de l'exome nécessite des approches et des outils bioinformatiques particuliers différents de ceux utilisés pour la détection des variants ponctuels. La méthode privilégiée pour l'analyse de ces CNV sur les données de l'exome est l'analyse de profondeur des lectures (Figure 12). Cette méthode se base sur le fait que si le séquençage est uniforme, le nombre de lectures alignées sur une région du génome est proportionnelle au nombre de fois où cette région est représentée. Si une région est délétée ou dupliquée chez un patient, il y aura donc deux fois moins ou deux fois plus de lectures alignées sur cette région. Cependant, le séquençage n'est pas uniforme sur l'ensemble du génome et il est donc nécessaire de normaliser les données de séquençage contre une banque de données de centaines d'échantillons afin de compenser la variabilité de profondeur de séquençage (33). D'autres méthodes d'analyse des CNV sur les données de séquençage haut débit existe : analyse de l'alignement des paires de *reads*, analyse des *reads* tronquées, méthode de l'assemblage *de novo* et des combinaisons de ces différentes méthodes sont possibles (33).

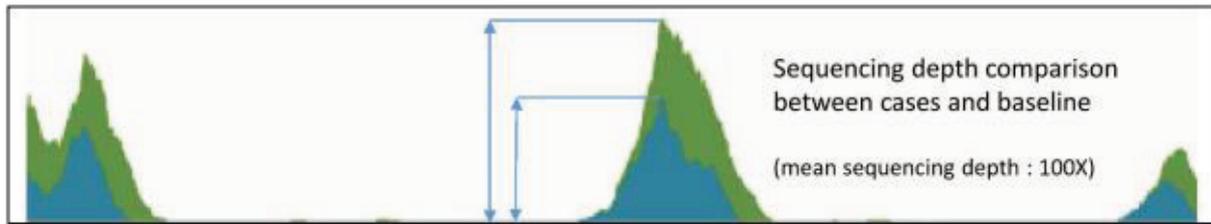


Figure 12 : Méthode de détection des CNV par comparaison des profondeurs de lecture (34)

2.3 Contexte de l'étude : Le projet « FIND » ou « Les données non sollicitées produites par SHD en diagnostic : du besoin des patients aux modalités organisationnelles »

La médecine génomique, qui repose sur notre capacité à séquencer le génome d'un individu, est une branche médicale émergente qui consiste à utiliser l'information génétique d'un individu dans le cadre du soin clinique concernant le diagnostic et la thérapeutique. Devant l'émergence de cette médecine génomique et la volonté politique ainsi que celle de l'opinion publique d'accéder au SHD s'inscrivant dans un parcours de soin personnalisé, le projet « FIND » ou « Les données non sollicitées produites par SHD en diagnostic : du besoin des patients aux modalités organisationnelles » a été mis en place. Ce projet aborde les nouveaux enjeux scientifiques, économiques et organisationnels auxquels le système de soins devra répondre.

En effet, le SHD pangénomique permet d'aboutir à un diagnostic primaire dans environ 36% des cas (4) mais peut aussi conduire à la découverte de variants non en lien avec la pathologie du patient et qui peuvent être pertinents en terme de prévention pour les patient et/ou leur famille. Ces données génétiques dites « non sollicitées » peuvent prédisposer à certaines pathologies à révélation tardive pour lesquelles un traitement ou une prévention est disponible. Ces DNS peuvent être également importantes pour le conseil génétique et donner des informations sur le risque de transmettre une pathologie. De plus, certains variants peuvent prédire la réaction d'une personne face à un traitement pharmaceutique et ainsi prédire une toxicité ou une absence d'efficacité. Cette dernière information est d'autant plus importante que certains patients sont hautement médicalisés.

La terminologie définissant ces données non sollicitées est riche. La littérature nord-américaine parle d'«*incidental findings*» ou «données incidentes» (35). En Europe, on parle de «résultats secondaires» par position au résultat primaire expliquant la maladie, en faisant remarquer que le terme «incident» ne décrit pas le caractère actif de la recherche de ces données (36). D'autres termes comme «données secondaires», «données accessoires» ou encore «résultats non liés» ont été évoqués dans la littérature ou lors de rassemblement d'experts mais ne permettent pas de saisir le contexte et la nature de ces données. Le choix du comité d'éthique de l'INSERM s'est arrêté sur «Données Non Sollicitées» ou DNS qui a donc été utilisé pour l'élaboration du projet FIND.

L'ACMG a été le premier à définir une liste minimale de 56 gènes considérés comme médicalement actionnables pour lesquels une recherche active de variants pathogènes ou probablement pathogènes devraient être réalisée lors d'un séquençage de l'exome ou du génome. Les gènes dit «actionnables» sont définis comme des gènes responsables d'un risque de développer une pathologie génétique future, accessible à une action préventive ou un traitement efficace surtout si mis en place précocement, améliorant ainsi le pronostic du patient en terme de mortalité et de morbidité (35). Ces gènes concernent surtout des pathologies cardiaques et des prédispositions aux cancers. Certains articles proposent une liste élargie de ces gènes (37,38) pour laquelle on retrouve des variants pathogènes ou probablement pathogènes chez environ 2 % des patients.

Cette recherche active de variants pathogènes dans des gènes actionnables demande évidemment un temps de travail supplémentaire aussi bien lors de la consultation clinique où une information et un consentement éclairé du patient sont indispensables que lors de l'analyse biologique ou l'interprétation et la validation du résultat par une autre technique sont nécessaires.

Ces DNS sont divisés en trois catégories selon leur implication clinique. Il s'agit de variants pathogènes ou probablement pathogènes dans :

- 1- 122 gènes pouvant être responsable d'une prédisposition génétique à une maladie à révélation tardive accessible à une action préventive ou à un traitement. Ce groupe comprend les 56 gènes définis par l'ACGM ainsi que les 58 gènes supplémentaires évoqués par Dorschner et al., 2013. 4 gènes responsables de prédispositions à des

coagulopathies (*F2 (RS1799963)*, *F8*, *F9*, *VWF*) ont été rajoutés car leur implication en termes de prise en charge et prévention semble importante. Ce groupe de gènes correspond donc à des prédispositions à des maladies pouvant apparaître dans l'enfance ou à l'âge adulte. La majorité des gènes répondent à un mode de transmission autosomique dominant mais également autosomique récessif ou lié à l'X. Dans ce groupe, le patient a un risque significatif de développer une pathologie au cours de sa vie pour laquelle une prévention ou un traitement précoce existent.

- 2- 114 gènes ayant un impact sur le conseil génétique, c'est-à-dire des gènes responsables de pathologies autosomiques récessives ou récessives lié à l'X pour lesquels des informations concernant le statut hétérozygote pourrait être importantes. Cette liste comprend 4 gènes de pathologies récessives dont la prévalence de porteurs hétérozygotes en France est élevée (*CFTR*, *HFE*, *SMN1* et *CYP21A2*) ou seulement certains variants connus pathogènes sont recherchés et 110 gènes de pathologies récessives liées au chromosome X chez des femmes conductrices (39). Dans ce groupe, le patient n'a pas de risque de développer la maladie mais un risque de la transmettre à sa descendance si son conjoint est également porteur d'un variant pathogène ou si la mère est porteuse d'un variant sur le chromosome X qu'elle pourrait transmettre à son fils. Un conseil génétique est donc indispensable pour le patient et sa famille afin d'informer des risques et de dépister d'autres membres de la famille.

 - 3- 2 gènes comprenant des variants pharmacogénétique d'intérêt clinique conduisant à des ajustements ou des précautions thérapeutiques (*CYP2C19* et *CYP2C9*). Certains variants homozygotes dans ces gènes confèrent un statut de « métaboliseur lent » avec un risque de toxicité ou d'absence d'effet de certains médicaments notamment des traitements psychotropes ou anti-épileptiques. Dans ce groupe de gènes, le patient n'a pas de risque de développer une maladie ni d'en transmettre une à sa descendance mais ceci permettra une adaptation thérapeutique lors de la prescription de certains médicaments.
- L'ensemble des gènes est répertorié en annexe.

La recherche de ces DNS reste un sujet très débattu en France et à l'étranger. Ceci soulève des questions éthiques autant au niveau individuel que sociétal. Récemment, l'avis 130 du Comité Consultatif National d'Ethique datant du 29 mai 2019 a donné un avis favorable pour la recherche des DS concernant les gènes actionnables pour lesquels une stratégie préventive ou thérapeutique est possible tout en rappelant la nécessité d'informer les patients et leur famille des conséquences de la recherche de ces DNS au moment de la rédaction initiale du consentement au préalable. La Société Française de Médecine Prédictive et Personnalisée a publié récemment des recommandations concernant les données incidentes touchant des gènes de prédispositions au cancer (40). Pour l'instant, il n'existe pas de recommandations européennes proposant une liste de gènes minimale. Chaque laboratoire peut faire le choix d'une stratégie locale consignée dans des procédures. Cependant, le cadre légal français (article 16-10 du code civil et R.1131-5 du code de santé publique) ne permet pas actuellement la recherche de ces DNS. Le cadre de recherche à travers le protocole FIND permet d'encadrer l'analyse des DNS en attendant une évolution législative à ce sujet.

Les objectifs du projet « FIND » sont :

- D'évaluer la perception initiale par les patients/parents (avant réalisation du séquençage de l'exome) des DNS des 3 catégories ainsi que l'expérience du vécu des résultats chez des patients porteur d'une AD après la réalisation d'un séquençage de l'exome.
- D'évaluer le coût et la faisabilité de la mise en place de cette démarche au sein des établissements hospitaliers.
- D'évaluer les conséquences des DNS en termes de vécu psychologique des familles et en termes de recours au soin.

Cette étude se déroule dans 3 centres de références :

- Le CHU de DIJON
- L'Assistance Publique- Hôpitaux de Paris, Hôpitaux Universitaires Pitié Salpêtrière
- Les Hospices Civils de Lyon

Nous nous consacrerons ici uniquement sur les données génétiques primaires (expliquant le phénotype du patient) et les DNS des patients inclus aux Hospices Civils de Lyon.

2.4 Objectif

Le travail de thèse s'inscrit dans le cadre du protocole FIND intitulé « Les données non sollicitées produites par séquençage haut débit en diagnostic : du besoin des patients aux modalités organisationnelles » qui s'intéressent donc à évaluer la perception initiale par les familles des DNS et l'expérience du vécu des résultats chez des patients porteur d'une AD/DI et bénéficiant d'un séquençage de l'exome. Ce projet regroupe 3 centres hospitalo-universitaires : le CHU de Dijon, l'APHP et les Hospices Civils de Lyon. Il a permis d'inclure 330 patients, dont 110, à Lyon. L'objectif de cette thèse sera d'analyser les résultats du séquençage de l'exome chez les 110 patients porteurs d'AD/DI des Hospices Civils de Lyon inclut dans ce projet. L'analyse des résultats portera d'une part sur les données primaires expliquant le diagnostic des patients à la fois par l'analyse des variants nucléotidiques (SNV) et par celle des variations du nombre de copies (CNV), d'autre part sur l'analyse des DNS portant sur des gènes actionnables pour lesquels une action préventive et/ou curative est possible à partir des données du séquençage de l'exome.

Auparavant, l'exploration des DI au sein du laboratoire de génétique passait par la réalisation d'un panel de gènes d'environ 500 gènes impliqués dans la DI si l'ACPA préalablement réalisée était négative. Le projet FIND a été l'occasion d'opter pour une approche par le séquençage de l'exome pour l'exploration des AD/DI au sein du laboratoire de génétique du CHU de Lyon.

3 MATERIELS ET METHODES

3.1 Patients

110 patients présentant une anomalie de développement ont été inclus dans le projet « FIND » au sein du service de génétique médicale des Hospices Civils de Lyon de novembre 2017 à avril 2019.

Les critères d'inclusions étaient :

- Adultes ou parents d'enfants ou d'adultes sous tutelle ou curatelle atteints d'Anomalie du Développement qui bénéficieront pour la première fois d'un séquençage de l'exome dans un contexte diagnostique
- Acceptation de participation à l'étude
- Souhait de rechercher au moins un groupe de DNS
- Capacité à s'exprimer en langue française

Les critères de non inclusion étaient :

- Patient non affilié à la sécurité sociale
- Pas de consentement du patient ou de son représentant légal

Les patients ont bénéficié au préalable de la réalisation d'une ACPA qui n'a pas permis de mettre en évidence de déséquilibre chromosomique ainsi que la recherche d'une expansion de triplet du gène FRAXA responsable du syndrome de l'X-fragile devant une DI sans syndrome polymalformatif. Certains patients ont bénéficié d'explorations supplémentaires avant le séquençage de l'exome comme des examens biologiques à la recherche de pathologie métabolique, la recherche ciblée de certains gènes en séquençage SANGER si une hypothèse diagnostique était évoquée ou la réalisation d'un panel de gène DI450 ou d'un panel de gène impliqué dans les épilepsies d'origine génétique en fonction du tableau clinique du patient. Ces explorations étaient toutes négatives.

Chaque patient a bénéficié d'une consultation de génétique auprès d'un généticien qui a recueilli le phénotype du patient et a proposé l'exome. Une deuxième consultation auprès

d'une conseillère en génétique permettait d'informer plus spécifiquement les patients sur la recherche des DNS. Lors de cette consultation, le consentement écrit, libre et éclairé du patient ou des parents ou des tuteurs était recueilli et des auto-questionnaires relatifs aux objectifs du projet « FIND » sur l'appréciation des DNS par les patients étaient donnés aux patients. Un prélèvement veineux périphérique du patient et de ses parents ont été effectués dans le centre de prélèvement rattaché au laboratoire de cytogénétique où les explorations ont été réalisées pour la réalisation du séquençage de l'exome. Dans le cadre de ce protocole, une consultation de rendu de résultat ainsi que des consultations post-résultat à 6 mois puis à 12 mois étaient prévues dans le but de recueillir le ressenti des patients face à l'exome et aux DNS à l'aide d'auto-questionnaires et d'entretiens auprès de psychologues.

3.2 Séquençage de l'exome

3.2.1 Technique de séquençage

Une stratégie de séquençage en trio, c'est-à-dire du patient et de ses parents pour faciliter l'analyse bioinformatique en fonction des modes de transmission, a été choisie. L'ADN a été extrait à partir du sang veineux périphérique sur un extracteur automatisé Hamilton grâce au kit Nucleospin Blood (Macherey-Nagel) sur la plateforme Biogenet du Centre de Biologie de Pathologie Est des Hospices Civils de Lyon. Après uniformisation des concentrations d'ADN à 20 ng/μL (dosage sur l'automate Qubit, (Thermo Fisher)), l'ADN a été fragmenté par sonication sur l'automate Covaris (PerkinElmer) en fragments de 180 à 220 paires de bases. Une électrophorèse sur le système Labchip GX Touch (PerkinElmer) a permis de contrôler la bonne fragmentation des échantillons. La préparation des bibliothèques s'est déroulée sur l'automate Siclone (PerkinElmer) avec le kit réactif HyperPrep Kapa HTP (Roche) (Étapes de End repair, adenylation et ajout des adaptateurs, size selection, courte amplification pré-capture). Elle a été suivie d'une purification, sur l'automate Zéphyr (PerkinElmer). Ensuite, les bibliothèques ont été regroupées en pools de 12 patients ou 4 trios. L'enrichissement des fragments s'est fait grâce au kit Hypercap Target Enrichment (Roche). Une étape de 16 à 20 heures d'hybridation est suivie d'une purification avec des billes Capture beads, d'une amplification par PCR et de nouvelles étapes de purification avant élution des fragments amplifiés d'intérêt. La concentration en ADN et la taille moyenne des

fragments ont été de nouveau vérifiés sur une puce Bioanalyser Agilent High Sensitivity DNA (Agilent Technologies). Le séquençage a été lancé après chargement de la «*flowcell*» et des consommables nécessaires pour le fonctionnement du NextSeq500 (Illumina). Les fragments dénaturés se sont fixés de façon aléatoire sur la «*flowcell*». L'amplification s'est faite en bridge, c'est-à-dire que des fragments double-brin ont été créés et dénaturés de nombreuses fois de façon à créer un cluster dense de fragments identiques. La réaction de séquençage a utilisé des terminateurs marqués détectés par laser. Ce système utilise la technologie *paired-end*, c'est-à-dire que les fragments ont été séquencés dans les deux sens. Chaque fragment compte 150 pb.

3.2.2 Analyse bioinformatique

3.2.2.1 Analyse des variants nucléotidiques

Les analyses bio-informatiques ont été réalisées par les bio-informaticiens de la cellule bioinformatique des HCL et suivent les recommandations du Broad Institute. Les données brutes à la sortie du séquençage sont sous forme d'un fichier bcl. Le programme fourni par Illumina bcl2fastq a été utilisé pour l'étape de démultiplexage. La qualité des lectures ont été analysé avec le logiciel FASTQC (41). Pour chaque échantillon, les fragments d'intérêt ont été «*découpée*» avec le programme Trimmomatic v 0.33 (42). Ensuite, les lectures ont été alignées sur le génome de référence GRCH37 (hg19) grâce au logiciel BWA-MEM v 0.7.12 (43), produisant pour chaque échantillon un fichier BAM qui a été indexé et trié avec le programme samtools v 1.3.1 (44). Ces fichiers ont ensuite été prétraités :

- Marquage des duplicates à l'aide du logiciel PicardTools MarkDuplicates v 1.138 (Picard Toolkit, 2018)
- Réalignement des indel en utilisant GATK IndelRealigner (45)
- Recalibration des nucléotides en utilisant GATK BaseRecalibrator (45)

L'appel de variants a été réalisé à l'aide du programme GATK HaplotypeCaller (45) produisant un fichier VCF génomique pour chaque échantillon. Le génotypage a été fait grâce au programme GATK GenotypeGVCFs (45), rassemblant tous les échantillons en un unique

VCF. La normalisation des variants et l'annotation ont respectivement été pris en charge avec les outils GATK LeftAlignAndTrimVariants (45) et snpEff/SnpSift toolbox (46). Les fichiers VCF annotés étaient ensuite formatés pour être exploités sur le logiciel tableur Microsoft Excel. Genofilter est un programme « maison » développé par le Docteur Thomas Simonet qui permet ensuite d'analyser les variants en fonction des hypothèses de transmission.

La figure 13 décrit le pipeline bioinformatique utilisé pour l'analyse des données de séquençage de l'exome.

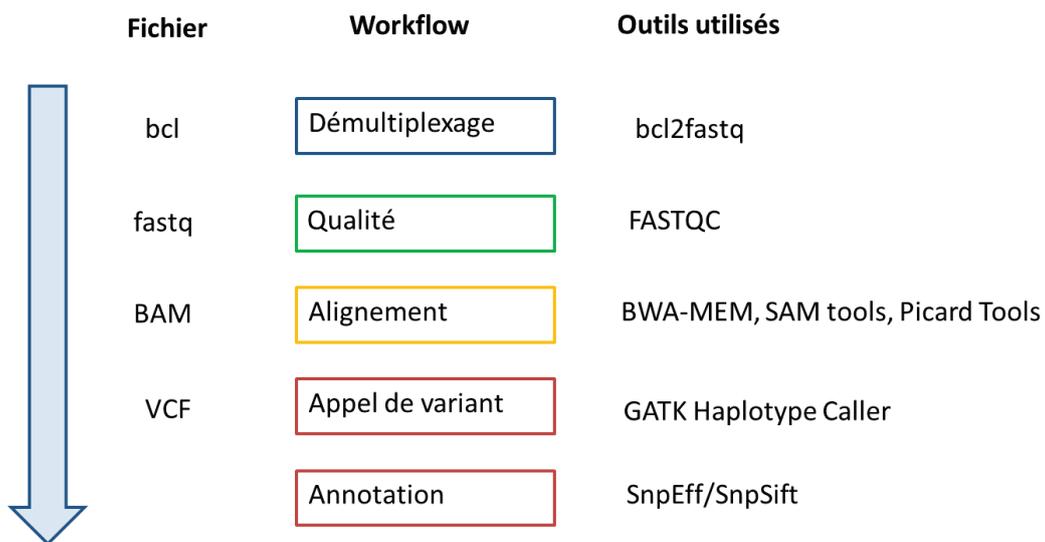


Figure 13 : Pipeline bioinformatique utilisé pour l'analyse des données de séquençage de l'exome

3.2.2.2 Analyse des CNV

Concernant l'analyse des CNV, elle a été réalisée par l'équipe de bioinformatique du CHU de Dijon. Le logicielXHMM (eXome Hidden Markov Model) a été utilisé. Ce logiciel permet la comparaison des profondeurs de lecture entre le patient et des témoins au niveau des sondes de la capture pour déterminer la présence d'un CNV sur une région donnée. La profondeur a été calculée par le logiciel GATK. Le logiciel crée une ligne de base en moyennant les profondeurs des témoins qui sert de base de comparaison pour la détection des CNV chez le patient. Cette ligne de base permet de normaliser les profondeurs de chaque sonde via une analyse de composante principale (ACP). Le logiciel XHMM détecte ensuite les CNV grâce à une méthode basée sur la chaîne de Markov cachée. Les résultats sont des

déviations par rapport à un z-score. Cette analyse se fait à partir des fichiers BAM. Le CHU de Dijon possède une base de données d'environ 900 exomes qui a permis de réaliser cette analyse ce qui permet de normaliser les profondeurs de séquençage.

3.2.3 Interprétation clinico-biologique

3.2.3.1 Interprétations des SNV

Les données métriques de couverture et de profondeur ont été vérifiées avant l'analyse de chaque exome. Une couverture minimum de 95% des gènes OMIM à 20X et une profondeur moyenne de 100X étaient attendues. De plus, la filiation était vérifiée par la vérification de la présence d'un nombre suffisant de variants rares hérités des parents. Les variants présentant un biais de brin (variant présent uniquement sur les lectures du même sens), une fréquence >1% dans les bases de données publics comprenant Exac, dbSNP, Exome Sequencing Project (ESP) et 1000 Genome Project, présent en mosaïque faible (< 1%) ou présentant une faible profondeur (< 10X) ont été filtrés. Dans un premier temps, la recherche de variants responsables du phénotype a été effectuée (recherche des données primaires). Une analyse des variants prenant en compte les modes de transmission *de novo*, autosomique récessif homozygote et hétérozygote composite, lié à l'X a été effectuée. Les variants se trouvant dans des régions codantes (faux-sens, non-sens, insertion, délétion ou affectant les sites d'épissage) ont été conservés ainsi que les variants présentant des scores de pathogénicité positifs avec un ou plusieurs logiciels de prédictions in-silico (SIFT, Polyphen2, Mutation Taster). Un filtre sur l'occurrence du variant dans la série permettait également de sélectionner les variants avec un seuil à 3 sauf pour la recherche de variants hétérozygotes composites où on ne filtrait pas sur ce critère afin de garder des variants relativement fréquents. Les variants d'intérêt révélés au décours de cette analyse ont ensuite fait l'objet d'une bibliographie extensive (Pubmed, OMIM, HGMD, Genecards, clinVar...). Les variants étaient ensuite classés selon les critères de l'*American College of Medical Genetics and Genomics* (ACMG) allant de « *benign* » à « *pathogenic* » en passant par « *of unknown significance* ». Les figures 14 et 15 décrivent les critères de pathogénicité de l'ACMG et les combinaisons de ces critères pour définir la pathogénicité d'un variant (47).

Si un gène n'était pas encore décrit en pathologie humaine mais que la fonction de ce gène et le variant était en faveur d'un effet délétère, le gène pouvait être déposé sur la plateforme *Genematcher* qui met en lien différents laboratoires dans le monde et permet ainsi de colliger des cas cliniques et de décrire de nouveau gène responsable de pathologies humaines.

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

Figure 14 : Critère de pathogénicité d'un variant selon l'ACMG

| |
|--|
| <p>Pathogenic</p> <ol style="list-style-type: none"> 1 Very Strong (PVS1) AND ≥1 Strong (PS1–PS4) OR ≥2 Moderate (PM1–PM6) OR 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) OR ≥2 Supporting (PP1–PP5) ≥2 Strong (PS1–PS4) OR 1 Strong (PS1–PS4) AND ≥3 Moderate (PM1–PM6) OR 2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5) OR 1 Moderate (PM1–PM6) AND ≥4 Supporting (PP1–PP5) <p>Likely Pathogenic</p> <p>1 Very Strong (PVS1) AND 1 Moderate (PM1–PM6) OR 1 Strong (PS1–PS4) AND 1–2 Moderate (PM1–PM6) OR 1 Strong (PS1–PS4) AND ≥2 Supporting (PP1–PP5) OR ≥3 Moderate (PM1–PM6) OR 2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5) OR 1 Moderate (PM1–PM6) AND ≥4 Supporting (PP1–PP5)</p> <p>Benign</p> <p>1 Stand-Alone (BA1) OR ≥2 Strong (BS1–BS4)</p> <p>Likely Benign</p> <p>1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) OR ≥2 Supporting (BP1–BP7)</p> |
|--|

Figure 15 : Règles de combinaison des critères de l'ACMG pour définir la pathogénicité d'un variant

Dans un second temps, une analyse des DNS a été effectuée grâce à des filtres sur les gènes présents dans les groupes 1 (gènes actionnables) et 2 (gènes importants pour le conseil génétique) des DNS. Avant d'effectuer les filtres sur la fréquence des variants dans les bases de données, les variants pathologiques p.Phe508* du gène *CFTR* et p.Cys282Tyr du gène *HFE* étaient recherchés activement à cause leur fréquence supérieur à 1% dans la population général. Pour ces 2 gènes, une liste définie de variants pathogène a été préétabli. Les 2 gènes du groupe 3 (pharmacogénétique) ont nécessité une analyse bioinformatique supplémentaire.

La description précise du phénotype du patient ainsi que l'accès aux examens complémentaires et aux photographies afin d'apprécier les particularités morphologiques sont primordiales pour l'interprétation des variants. Le logiciel visualisateur Alamut Visual (integrative biosoftware, Sophia Genetics), intégrant des informations génétiques notamment les bases de données et les logiciels de prédictions *in silico*, permettait de visualiser les variants et la profondeur locale grâce à l'importation des BAM. La figure 16 représente une capture d'écran du fichier du VCF formaté exploité sur le logiciel Excel. La figure 17 représente une capture d'écran du logiciel visualisateur Alamut Visual avec le fichier BAM importé d'un patient.

| E75 RGPD3 | | | | | | | | | | | | | | | | |
|-----------|----------|------------|-------------------------|---------|-------------------|-----------|---------------------------|------------------|-----------|-----------|-------------|---------|----------|----------|--------------|----|
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | | |
| | Patients | Occurrence | Occurrences_par_variant | Variant | Gene | HGVS_Nucl | HGVS_Proteo | Effets | Transcrit | Tissus_Ex | Variant_T | Locus_M | Genotype | Genotype | Prof_Alle | |
| 5 | 13A3395 | 2 | | 2 | Chr1:168855 NADK | A | p.Ala259Thr... | intron_variant | 994.1 | NM_001198 | NA | 0 | 0 | C/T | Heterozygote | 15 |
| 6 | 13A3395 | 2 | | 2 | Chr1:211755 PRKCZ | c.92C>T | p.Ala31Valp.Ala31Val..... | missense_variant | 671.1 | NM_001282 | NANAPlacent | 0 | 0 | G/A | Heterozygote | 13 |

| E75 RGPD3 | | | | | | | | | | | | | | | | | | | | | |
|-----------|-----------|-----------|-------------------|----------|------|----------|----------|----------|-------|-----------------------|---------------|----------|----------|-------------|----------|----------|-----------|-------------|-----------|---------|------|
| O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG | AH | AI | |
| | Prof_Alle | Prof_Alle | Novosaicme | Prof_Gen | SIFT | Polyphen | Polyphen | Mutation | dbSNP | 1000G_Freq | 1000G_Afrique | 1000G_Eu | ESP_Freq | GME_AF | ExAC_Occ | ExAC_Hon | ExAC_Freq | ExAC_pLI | ClinVar_S | OMIM_PH | OMIM |
| 5 | 15 | 17 | 0.46875 | | 100 | Benin | Benin | Benin | Benin | rs19993181:3.99361E-4 | 0.0000 | 0.0020 | 0 | 0.00125313: | 67 | 0 | 1.544e-03 | 0.00126768: | NA | NA | NA |
| 6 | 13 | 16 | 0.448275862068966 | | 100 | NA | NA | NA | NA | rs55252603:1.99681E-4 | 0.0000 | 0.0000 | 0 | NA | 8 | 0 | 4.599e-04 | NANA0.2669 | NA | NA | NA |

| E75 RGPD3 | | | | | | | | | | | | | | | | | | | | | | |
|-----------|-----------|-------------|-----------|---------|--------|-----------|-----------|-----------|----------|----------|--------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|-------|---|
| AE | AF | AG | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR | AS | AT | AU | AV | AW | AX | AY | AZ | |
| | ExAC_Freq | ExAC_pLI | ClinVar_S | OMIM_PH | OMIM_H | Biais_Bri | Effets_Ec | Mosirique | Profonde | Variants | HGMD_H | HGMD_Ph | HGMD_Bi | HGMD_Ph | Classe1 | Classe2 | Classe3 | Autosom | De_Novo | Heterozyg | Lie_X | |
| 5 | 1.544e-03 | 0.00126768: | NA | NA | NA | NA | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 4.599e-04 | NANA0.2669 | NA | NA | NA | NA | 0 | 0 | 0 | 0 | NA | NA | NA | NA | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

Figure 16 : Capture d'écran des fichiers Excel formatés à partir du fichier VCF

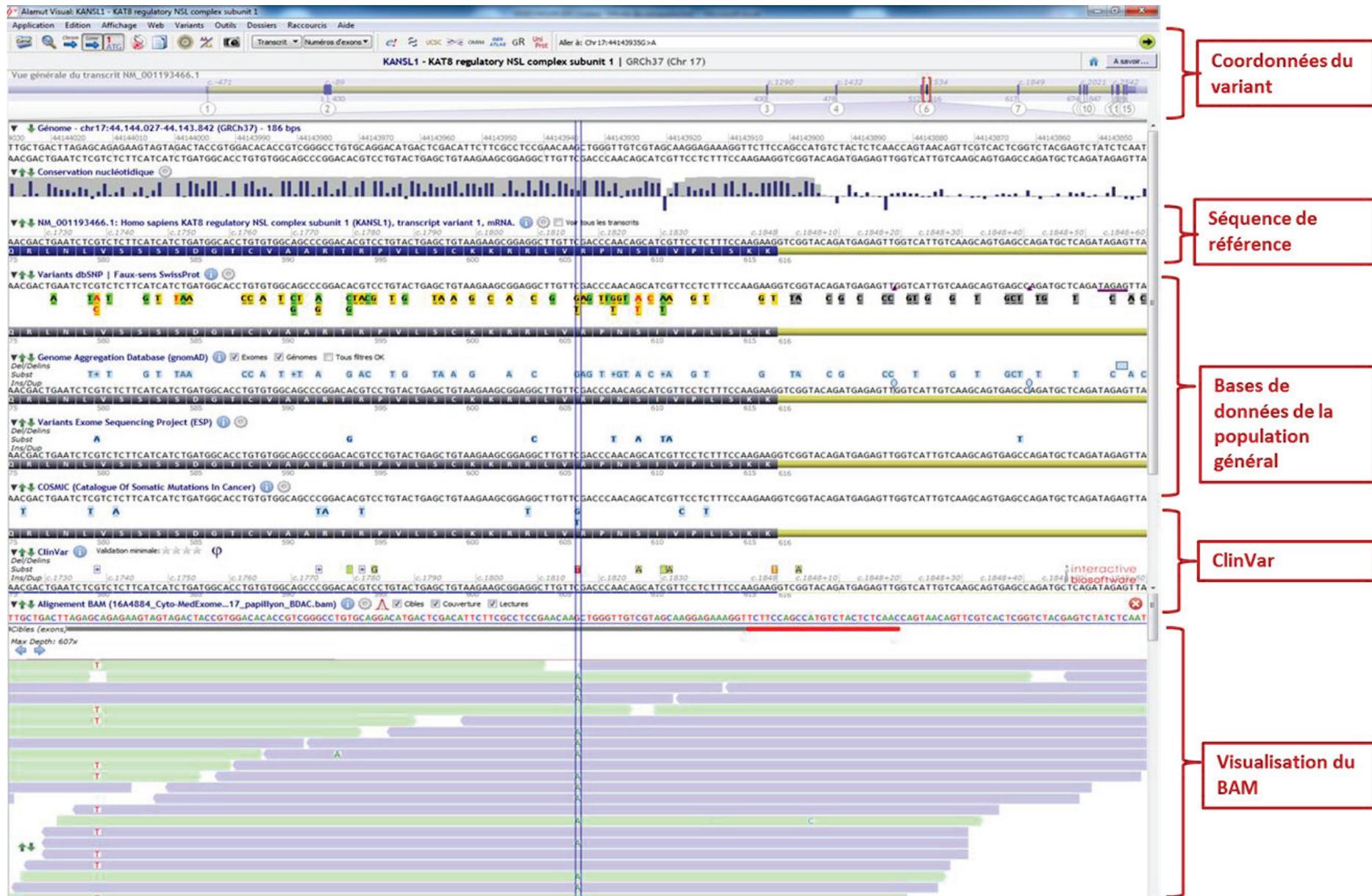
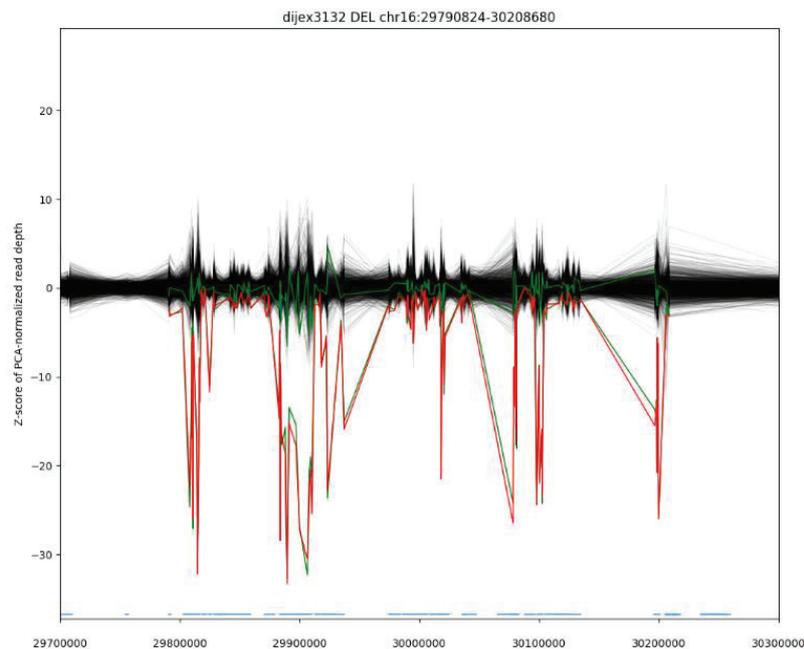


Figure 17 : Visualisation d'un fichier BAM avec le logiciel Alamut Visual

3.2.3.2 Interprétation des CNV

Concernant l'interprétation des CNV, le CHU de Dijon nous a transféré un fichier Excel par échantillon comportant la liste des CNV détectés ainsi qu'une représentation graphique pour chaque CNV. Le graphique du CNV représentait en abscisse les positions génomiques des sondes de capture (exons) et en ordonnée le z-score de la profondeur normalisée par ACP. La figure X montre un exemple de la représentation graphique d'un CNV. Le fichier Excel nécessitait une étape de filtrage sur l'occurrence (inférieur ou égale à 3 occurrences) au sein de notre série de patient ainsi qu'au sein de la base de données de Dijon pour éliminer les CNV fréquents. Un score indicateur de probabilité de la présence du CNV aidait également à trier les CNV (score en pourcentage ; 99% Présence du CNV très probable). Les patients ayant déjà bénéficié d'une ACPA avant le séquençage de l'exome les CNV qualifiés de VOUS (*Variant of Unknown Significance*) détectés par cette technique ont pu être mis en évidence également par cette analyse de CNV sur les données de l'exome. La figure 18 et 19 montre respectivement la représentation graphique d'une délétion et le fichier



Excel listant les CNV d'un patient.

Figure 18 : Représentation graphique d'une délétion de 417 Kb en 16p11.2 héritée du père

La ligne rouge représente le patient et les 2 lignes vertes, les parents. L'axe des abscisses représente les coordonnées génomiques et l'ordonnée une déviation en z-score.

Coordonnées, genotype et taille du CNV **Score de précision** **Occurrence dans la série** **Parents**

| #CHR | START | END | CNV | INTERVAL | KB | NUM_TARG | Q_NON_DIP | MEAN_RD | BATCH_COU | BATCH_FREQ | XHMM_COU | XHMM_FREQ | dijex3133 | dijex3134 | CNVLength | CNVInterval | CytoBand | GeneList | Accession | Overl |
|------|-----------|-----------|-----|--------------|--------|----------|-----------|---------|-----------|-------------|----------|-------------|--------------|-----------|-----------|--------------|----------|------------|------------|---------|
| chr1 | 1654025 | 1666276 | DUP | chr1:1654025 | 12.25 | 5 | 92 | 4.27 | 6 | 0.006036217 | 29 | 0.03399765 | . | . | 12251 | chr1:1654025 | 1p36.33 | CDK11B,CDK | NM_033490, | partia |
| chr1 | 16892900 | 17083890 | DUP | chr1:1689290 | 190.99 | 5 | 99 | 8.43 | 15 | 0.015090543 | 10 | 0.011806375 | . | . | 190990 | chr1:1689290 | 1p36.13 | NBPF1,CROC | NM_017940, | partia |
| chr1 | 25599037 | 25655417 | DEL | chr1:2559903 | 56.38 | 3 | 35 | -4.70 | 36 | 0.036217303 | 60 | 0.061162079 | . | . | 56380 | chr1:2559903 | 1p36.11 | RSRP1,RHD | NM_0013217 | intra,i |
| chr1 | 45288913 | 45297471 | DEL | chr1:4528891 | 8.56 | 16 | 99 | -2.47 | 1 | 0.001006036 | 1 | 0.001029866 | . | . | 8558 | chr1:4528891 | 1p34.1 | PTCH2 | NM_0011662 | intra,i |
| chr1 | 148022925 | 148594642 | DUP | chr1:1480225 | 571.72 | 10 | 99 | 3.87 | 31 | 0.031187122 | 2 | 0.002339181 | DUP;KB=578.. | . | 571717 | chr1:1480225 | 1q21.2 | NBPF19,LOC | NM_0013513 | partia |

Gène présent dans le CNV **Bases de données**

| dijex3133 | dijex3134 | CNVLength | CNVInterval | CytoBand | GeneList | Accession | Overlap | DistanceToC | ClosestGene | OMIM | ClinGenDose | CNVmap(No | CNVmap | DGVStructur | DGVSupport | ISCABenign | ISCAPathoge | DevDelayCoi | DevDelayCass |
|-----------|--------------|-----------|--------------|----------|------------|------------|-----------------|-------------|-------------|--------------|--------------|-------------|--------------|-------------|------------|------------|-------------|-------------|--------------|
| . | . | 12251 | chr1:1654025 | 1p36.33 | CDK11B,CDK | NM_033490, | partial,parti | 0. | . | ./././././ | ./././././ | chr1:521413 | . | 0 | 0 | 0 | 0 | 0 | 0 |
| . | DUP;KB=190. | 190990 | chr1:1689290 | 1p36.13 | NBPF1,CROC | NM_017940, | partial,full,ft | 0. | . | ./././././ | ./././././ | chr1:168317 | . | 8 | 141 | 0 | 0 | 125 | 0 |
| . | DEL;KB=56.3 | 56380 | chr1:2559903 | 1p36.11 | RSRP1,RHD | NM_0013217 | intra,intra,in | 0. | . | ./[Rh-negati | ./ | chr1:255850 | chr1:2558501 | 22 | 2365 | 0 | 0 | 1457 | 0 |
| . | . | 8558 | chr1:4528891 | 1p34.1 | PTCH2 | NM_0011662 | intra,intra | 0. | . | Basal_cell | ci | . | . | 0 | 0 | 0 | 0 | 0 | 0 |
| . | DUP;KB=578.. | 571717 | chr1:1480225 | 1q21.2 | NBPF19,LOC | NM_0013513 | partial,parti | 0. | . | ././././././ | ././././././ | chr1:148594 | . | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 19 : Fichier Excel listant les CNV d'un patient

L'interprétation des CNV à partir des données de l'exome repose sur plusieurs critères :

- Bases de données de populations contrôle (DGV, 1000 Genomes Project, ExaC, gnomAD)
- Bases de données de patient (OMIM, UCSC, Ensembl, Clingen, ClinVar, LOVD, HGMD, DECIPHER)
- Données de ségrégation familiale : un CNV *de novo* est plutôt en faveur d'un caractère pathogène alors qu'un CNV hérité est plutôt en faveur d'un caractère bénin même s'il existe des exceptions (CNV pathogène familiale, pénétrance, expressivité variable, délétion héritée révélant une maladie récessive, transmission liée à l'X, région soumise à empreinte, parent porteur du CNV à l'état de mosaïque)
- Taille du CNV : Généralement, un CNV pathogène est significativement plus grand qu'un CNV bénin. En effet, 90 à 95% des CNV bénins ont une taille inférieure à 500kb et seulement 1 à 2 % des CNV identifiés chez des individus sains ont une taille supérieure à 1kb (48,49). Cependant, certains CNV de grande taille peuvent être bénins et des CNV de très petite taille peuvent être pathologiques s'ils concernent quelques exons d'un gène pathologique.
- Contenu génique : un CNV dans une région dont la densité génique est haute est en faveur de son caractère pathogène. De plus, si le CNV concerne un gène décrit en pathologie humaine et qu'il existe une corrélation avec le phénotype, ceci est également en faveur de son caractère pathogène. Il faut cependant prendre en compte les conséquences fonctionnelles d'un CNV impliquant un gène donné (Haploinsuffisance, pathologie dominante, récessive, lié à l'x, perte ou gain...)
- Nature du CNV : délétion hétérozygote, homozygote, duplication, triplification...

La classification de ces CNV a récemment été remise à jour en juin 2019 au sein du réseau Achro-puce et comporte ainsi 6 classes de variant :

- Classe 6 : PIEV : CNV de susceptibilité aux troubles neurodéveloppementaux (TND), à pénétrance Incomplète et/ou expressivité variable (classe récemment ajoutée)
- Classe 5 : CNV pathogène
- Classe 4 : CNV probablement pathogène

- Classe 3: VOUS (*Variant Of Unknown Significance*)
- Classe 2: CNV probablement bénin
- Classe 1 : CNV Bénin

La figure 20 représente les différentes étapes de séquençage de l'exome.

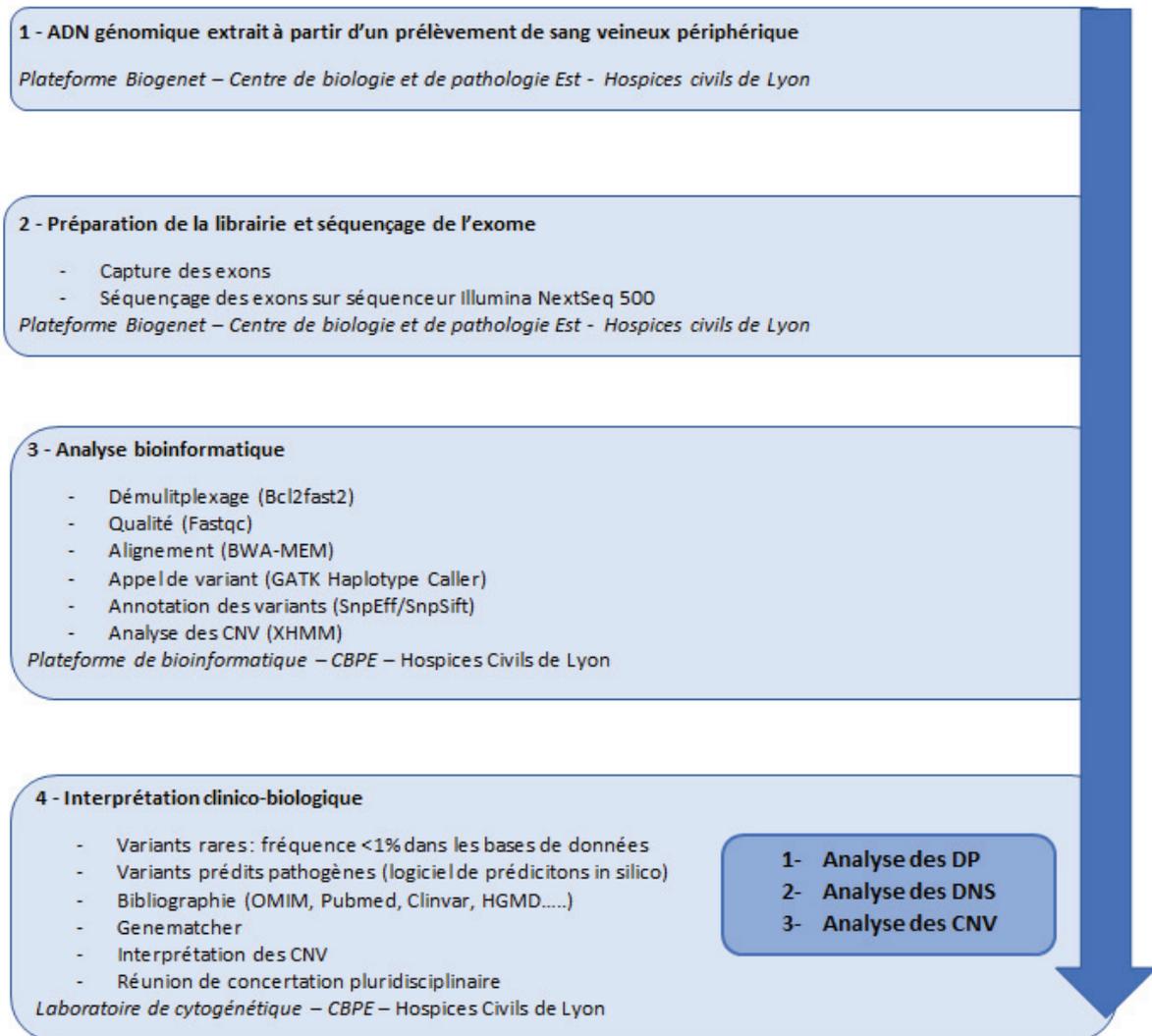


Figure 20 : Workflow du séquençage de l'exome

3.3 Vérification des variants nucléotidiques par séquençage

SANGER

Les diagnostics primaires (DP) n'ont pas été vérifiés par séquençage SANGER. Concernant les DNS, les variants de catégories de pathogénicité 4 et 5 des groupes de DNS 1 (gènes actionnables), 2 (gène intéressant pour le conseil génétique) et 3 (pharmacogénétiques) ont tous été vérifiés soit au sein des laboratoires référents du gène responsable de la pathologie en question au sein des Hospices Civils de Lyon ou d'autres CHU si nécessaire (CHU de Dijon pour les variants de pharmacogénétique).

3.4 Vérification des CNV par PCR quantitative

Les CNV de nature indéterminée, probablement pathogène ou pathogène ont été vérifiés par une technique de PCR quantitative (qPCR) chez le patient et ses apparentés. La technique choisie utilise le SYBR Green, intercalant de l'ADN qui a la propriété d'être fluorescent lorsqu'il est lié à l'ADN double brin. La PCR quantitative se base sur le calcul du Ct ou cycle seuil, qui est le moment où l'amplification devient mesurable par la mesure de la fluorescence. Celle-ci est effectuée en continu pendant toute la période d'amplification exponentielle, selon une méthode en temps réel. Le nombre de cycles nécessaires pour atteindre ce Ct est alors directement proportionnel à la quantité d'ADN de départ. Cette technique nécessite d'analyser en parallèle du gène d'intérêt un gène de référence, *ADORA2B* dans le laboratoire de cytogénétique. Une gamme de dilution pour le gène d'intérêt et une gamme de dilution pour le gène de référence de 5 points en triplicat avec des concentrations décroissantes d'ADN ont été réalisées. Chaque gène est également analysé en triplicat pour chaque membre de la famille étudié. La réaction se fait sur plaque 96 puits et utilise le SYBR Greenmaster mix (Qiagen). La PCR est lancée sur l'automate LightCycler®480. L'observation de la courbe de fusion permet de vérifier la spécificité des amorces utilisées. Le calcul du nombre de copies se base sur l'analyse des ratios entre le gène d'intérêt et *ADORA2B*.

3.5 Compte Rendu des résultats

Les résultats des DP et des DNS ont été rendus sur 2 comptes rendus distincts. Concernant les DP, un compte rendu pour le patient ainsi qu'un compte rendu pour chacun de ses parents ont été réalisés. Seul les variants de classe 3, 4 et 5 étaient rendus. Sur le compte rendu figurait les informations relatives à l'identité du patient et au prélèvement, l'indication clinique, la technique utilisée pour le séquençage de l'exome et les informations sur le variant. Ces informations comportaient le nom du gène, les coordonnées génomiques du variant ainsi que sa nomenclature en ADNc et protéique, son mode de transmission et son génotype. L'interprétation apportait des informations sur les conséquences du variant sur la séquence protéique, la présence du variants dans les bases de données ainsi que dans la littérature et concluait sur son caractère inconnu ou pathogène et la nécessité d'un conseil génétique. Les variants bénins et probablement bénins n'étaient pas rapporté sur le compte rendu. La profondeur moyenne et la couverture supérieure ou égale à 10 X précisant la qualité du séquençage de l'exome du patient.

Concernant les DNS, seul les variants de catégorie 4 et 5 c'est-à-dire probablement pathogène et pathogène sur le compte rendu étaient rapportés sur le compte rendu. Ces variants devaient avoir déjà été rapportés au moins une fois dans HGMD (Human Gene Data Base) (base de données qui répertorie les variants décrits en pathologie humaine dans la littérature) ou sur la base de données ClinVar qui répertorie les variants génétiques responsables de pathologies humaines). Le caractère transmis ou non du variant n'était pas précisé. Sur ce compte rendu figurent les mêmes informations que sur le compte rendu des DP à l'exception du mode de transmission. La nécessité d'avoir recours à une consultation spécialisée était précisée. Concernant les données de pharmacologie, la liste des médicaments à utiliser avec précaution était précisée. Un deuxième compte rendu était transmis par le laboratoire ayant réalisé la confirmation du variant.

4 RESULTATS

4.1 Patients

Lors des analyses statistiques, la cohorte comportait 110 patients. Tous les patients ont préalablement bénéficié d'une étude pangénomique par ACPA qui n'a pas retrouvé de déséquilibre chromosomique pouvant expliquer le phénotype. Certains patients ont également eu une recherche du syndrome de l'X fragile ou une ou plusieurs analyses ciblées de gènes selon les orientations diagnostiques. De plus, un panel DI a été réalisé chez 9 patients lorsque celui-ci était encore proposé au sein du laboratoire de cytogénétique ainsi que 14 panels épilepsie. Chez tous ces patients, les analyses de panel n'ont pas permis de mettre en évidence de variant pouvant expliquer le phénotype du patient. Le panel épilepsie est toujours en activité et régulièrement mis à jour.

Parmi les patients, 62 étaient des hommes (56.4 %) et 48 des femmes (43.6%). L'âge moyen des patients était de 11 ans allant de 0 ans (2 exomes réalisés sur des fœtus) à 36 ans pour le plus âgé. Les familles consanguines représentaient 11.8% des familles (13/110). Concernant la présentation clinique des patients : 5 cadres ont été définis : les patients présentant une DI ou un TSA isolé soit 17,3% (19/110) des patients, les patients présentant une DI syndromique c'est-à-dire associée à au moins une malformation congénitale soit 62.7% (69/110) des patients, les patients présentant une encéphalopathie épileptique au premier plan soit 10% (11/110) des patients, les patients présentant une atteinte neurologique au premier plan (dystonie, neurodégénérescence, neuropathie) soit 6.4% (7/110) des patients et les patients présentant une anomalie du développement sans DI soit 3.6% (4/110) des patients. Les caractéristiques des patients sont résumées dans le tableau 4. Plus précisément, parmi les signes cliniques les plus retrouvés en dehors de la DI, 27,3% (30/110) des patients présentaient une épilepsie, 20.9% (23/110) des troubles du comportement dont des TSA, 10.9% (12/110) un examens neurologique anormal, 16.4% (18/110) une imagerie cérébrale anormale, 19,1% (21/110) une microcéphalie ou une macrocéphalie, 16,4% (18/110) des troubles staturo-pondéraux, 27.3% (30/110) une dysmorphie et 39.1% (43/110) une ou plusieurs malformations congénitales. Les signes cliniques sont répertoriés dans le tableau 5.

| Caractéristiques des patients | |
|--------------------------------------|------------------------|
| Sexe ratio H/F | 62/48 |
| Age moyen au diagnostique | 11 ans (de 0 à 36 ans) |
| Consanguinité | 11,8% (13/110) |
| Présentation clinique | |
| DI/TSA isolé | 17,3% (19/110) |
| DI syndromique | 62,7% (69/110) |
| Encéphalopathie épileptique | 10% (11/110) |
| Maladie neurologique | 6,4% (7/110) |
| Anomalie de développement sans DI | 3,6% (4/110) |

Tableau 4 : Caractéristiques cliniques des patients de la cohorte

| Signes cliniques | |
|----------------------------------|----------------|
| Epilepsie | 27,3% (30/110) |
| Trouble du comportement dont TSA | 20,9% (23/110) |
| Signes neurologiques | 10,9% (12/110) |
| Imagerie cérébrale anormale | 16,4% (18/110) |
| Micro/macrocéphalie | 19,1% (21/110) |
| Trouble staturo-pondéraux | 16,4% (18/110) |
| Dysmorphie | 27,3% (30/110) |
| Malformation congénitale | 39,1% (43/110) |

Tableau 5 : Signes cliniques les plus fréquents retrouvés dans la cohorte de patient

4.2 Diagnostic primaire

La profondeur moyenne de séquençage dans la cohorte des 110 patients était de 81.92X et la couverture moyenne à plus de 10X était de 95.62%.

Parmi les 110 patients analysés, 40 variants pathogènes ou probablement pathogènes (classe 4 et 5 du classement de l'ACMG) expliquant le phénotype ont été retrouvés chez 39 patients de la cohorte. Le taux diagnostique au sein de la cohorte est donc de 35.5%. Parmi les 9 patients ayant pu bénéficier d'un panel DI au préalable, 5 ont présenté un variant pathogène ou probablement pathogène lors de l'analyse de l'exome. Parmi les 13 patients ayant pu bénéficier d'un panel épilepsie au préalable, 3 ont présenté un variant pathogène ou probablement pathogène lors de l'analyse de l'exome.

Vingt-cinq variants (62.5%) étaient des variants responsables de pathologies autosomiques dominantes, 9 variants (22.5%) responsables de pathologies autosomiques

récessives, 5 variants (12.5%) responsables de pathologies lié à l’X, récessives lié à l’X dans un cas et dominantes lié à l’X dans 5 cas. 9 variants (22.5%) ont été retrouvés à l’état homozygote, 2 sur le chromosome X à l’état hémizygote (5%) et 29 à l’état hétérozygote (72.5%).

Concernant le type de mutation, on retrouve 22 variants faux-sens (55%), 5 variants responsables d’un décalage de cadre de lecture (12.5%), 10 variants non-sens (25%), un variant (2.5%) responsable de la perte du codon d’initiation et 2 variants affectant l’épissage (5%). Parmi les 39 patients, 11 provenaient de familles consanguines. Un fœtus provenant de parents cousins germains présentait 2 variants à l’état homozygote dans 2 gènes responsable de pathologies récessives (*RAPSN* et *PC*) expliquant le phénotype complexe du fœtus.

Parmi les patients présentant un DP, 5 présentaient également un DS : 2 patients sur des gènes du groupe 1 (gène actionnable) (*FBNI* responsable du syndrome de Marfan et *PROC* responsable d’une coagulopathie), 2 patients sur les gènes du groupe 3 (pharmacogénétique : *CYP2C9* et *CYP2C19*) et un patient était porteur hétérozygote du variant p.Cys282Tyr dans le gène *HFE* responsable de l’hémochromatose de type 1.

Parmi les 40 variants pathogènes décrits, 16 (40%) avait déjà été rapportés soit dans la littérature soit dans la base de données ClinVar. Ces variants concernent les gènes suivants : *GNB5*, *HEXA*, *KANSL1*, *PACSI*, *PPP2R5D*, *DYRK1A*, *WDR45*, *RAC1*, *CKD13*, *MTO1*, *MORC2*, *PCGF2*, *DDX3X*, *ZMYND11*, *GNB1* et *RAPSN*.

Parmi les présentations cliniques des patients ayant un DP, 2/39 présentaient une DI/TSA isolée (5.2%), 27/39 présentaient une DI syndromique (69.2%), 5/39 une encéphalopathie épileptique (12.8%), 3/39 une pathologie neurologique au premier plan (7.7%) et 3/39 une ou plusieurs anomalies de développement sans DI (7.7%). Au sein des patients ayant eu un DP et présentant une DI isolée ou un TSA isolé, le taux diagnostique est de 10.5%. Parmi les DI syndromique, le taux diagnostique est de 39.1%, parmi les encéphalopathies épileptiques, le taux diagnostique est de 45.5% et respectivement de 42.8% et 75% dans les groupes où la pathologie neurologie est au premier plan et dans le groupe des AD sans DI.

Parmi les DP, 5 patients sur 38 avaient bénéficiés d’un panel DI ou épilepsie avant la réalisation de l’exome. Les gènes responsables de la pathologie retrouvées grâce au séquençage de l’exome n’étaient pas présents sur les panels au moment de leur réalisation. 14 gènes sur les 39 responsables des pathologies des patients sont sur les panels DI ou épilepsie.

Pour ces patients, aucun panel n'avait été réalisé au préalable puisque la porte d'entrée n'était pas l'épilepsie et que le panel n'était plus réalisé. Les caractéristiques des variants et de la présentation clinique des patients ayant eu un DP sont résumés dans les tableaux 6 et 7. De plus, le tableau 8 répertorie les 40 variants pathogènes ou probablement pathogènes retrouvés dans la cohorte des 110 patients.

| Caractéristiques cliniques des patients ayant eu un DP | Nombre de patients | Pourcentage | Taux diagnostique |
|---|---------------------------|--------------------|--------------------------|
| DI/TSA isolée | 2/39 | 5,2 | 2/19 (10,5%) |
| DI syndromique | 27/39 | 69,2 | 27/69 (39,1%) |
| Encéphalopathie épileptique | 5/39 | 12,8 | 5/11 (45,5%) |
| Maladie neurologique au premier plan | 3/39 | 7,7 | 3/7 (42,8%) |
| AD sans DI | 3/39 | 7,7 | 3/4 (75%) |

Tableau 6 : Caractéristiques cliniques et taux diagnostiques des patients ayant eu un DP

| Caractéristiques des variants de classe 4 et 5 | Nombre | % |
|---|---------------|----------|
| Type de mutation | | |
| Faux-sens | 22/40 | 55 |
| Décalage du cadre de lecture | 5/40 | 12,5 |
| Non-sens | 10/40 | 25 |
| Perte du codon d'initiation | 1/40 | 2,5 |
| Variant affectant l'épissage | 2/40 | 5 |
| Zygotie | | |
| Homozygote | 9/40 | 22,5 |
| Hémizygotie | 2/40 | 5 |
| Hétérozygote | 29/40 | 72,5 |
| Mode de transmission | | |
| AR, hérité des parents | 9/40 | 22,5 |
| DLX, <i>de novo</i> | 5/40 | 12,5 |
| RLX, hérité de la mère | 1/40 | 2,5 |
| AD, <i>de novo</i> | 25/40 | 62,5 |
| DS | 5/3 | 12,5 |
| Consanguinité | 11/39 | 28,2 |

Tableau 7 : Caractéristiques des variants de classe 4 et 5

| Patient | Clinique | Gène | Nomenclature HGVS | Conséquence protéique | Type de mutation | Zygotie | Mode de transmission | DNS |
|---------|---|----------------|--------------------------|-----------------------|------------------------------|--------------|------------------------|--------------|
| 1 | DI légère, dysmorphie | CDK13 | NM_003718.4:c.2140G>T | p.(Gly714Cys) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 2 | DI, microcéphalie, dysmorphie | BCL11A | NM_022893.3:c.1633del | p.(Asp545Thrfs*53) | Décalage du cadre de lecture | Hétérozygote | AD, <i>de novo</i> | Non |
| 3 | Encéphalopathie épileptique | CYFIP2 | NM_001291722.1:c.322T>C | p.(Tyr108His) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 4 | DI, surpoids, épilepsie | GRIN2B | NM_000834.3:c.2236T>A | p.(Cys746Ser) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 5 | DI, TSA | CAMK2A | NM_015981.3:c.49G>A | p.(Glu17Lys) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 6 | DI, bradycardie dysautonomique | GNB5 | NM_006578.3:c.242C>G | p.(Ser81Trp) | Faux-sens | Homozygote | AR, hérité des parents | Non |
| 7 | Microphthalmie colobomateuse, anomalie claviculaire | PORCN | NM_203475.2:c.626G>C | p.(Cys209Ser) | Faux-sens | Hémizygotie | RLX, hérité de la mère | Non |
| 8 | DI, trouble du comportement | NCKAPI | NM_205842.2:c.796A>T | p.(Lys266*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 9 | Régression cognitive et motrice | HEXA | NM_000520.4 : c.533G>A | p.(Arg178His) | Faux-sens | Homozygote | AR, hérité des parents | Non |
| 10 | RPM, dysmorphie | KANSL1 | NM_001193466.1:c.1816C>T | p.(Arg606*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 11 | Dystonie généralisée à début précoce | KMT2B | NM_014727.2 : c.424C>T | p.(Arg142*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 12 | Encéphalopathie épileptique, leucémie lymphoblastique | KCNQ5 | NM_001160133.1:c.1039G>A | p.(Gly347Ser) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Oui (CYP2C9) |
| 13 | DI, microcéphalie, dysostose mandibulo-faciale | EFTUD2 | NM_004247.3:c.342T>G | p.(Tyr114*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 14 | DI, dysmorphie, dilatation pyélocalicielle, | PACSI | NM_018026.3:c.607C>T | p.(Arg203Trp) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 15 | DI, macrocéphalie | PPP2R5D | NM_006245.3:c.592G>A | p.(Glu198Lys) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 16 | DI, TSA, microcéphalie, dysmorphie | DYRK1A | NM_001396.4:c.787C>T | p.(Arg263*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 17 | DI, TSA, retard staturo-pondéral, metatarsus varus | MED13L | NM_015335.4:c.1135C>T | p.(Arg379*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |

| | | | | | | | | |
|----|---|---------------|-------------------------------|---------------------|------------------------------|--------------|------------------------|------------|
| 18 | DI, épilepsie, agénésie du CC, dysmorphie, tâches hypopigmentées | ARID1B | NM_017519.2:c.3693_3694del | p.(Ser1231Argfs*32) | Décalage du cadre de lecture | Hétérozygote | AD, <i>de novo</i> | Non |
| 19 | Encéphalopathie épileptique | WDR45 | NM_007075.3:c.81_82del | p.(Gly28Cysfs*42) | Décalage du cadre de lecture | Hétérozygote | DLX, <i>de novo</i> | Non |
| 20 | DI, épilepsie | WDR45 | NM_007075.3:c.19C>T | p.(Arg7*) | Non-sens | Hémizygote | DLX, <i>de novo</i> | Non |
| 21 | Syndrome de Kabuki RPM, ataxie cérébelleuse, | RAC1 | NM_006908.4:c.116A>G | p.(Asn39Ser) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 22 | hypotonie, nystagmus | GRM1 | NM_001278064.1:c.889C>T | p.(Arg297*) | Non-sens | Homozygote | AR, hérité des parents | Non |
| 23 | RPM sévère, TSA, trigonocéphalie, CC court, épilepsie, dysmorphie, anomalie des extrémités | TCF4 | NM_001243226.2:c.1765_1768dup | p.(Asn590Thrfs*26) | Décalage du cadre de lecture | Hétérozygote | AD, <i>de novo</i> | Non |
| 24 | RPM, TSA, sténose artérielle pulmonaire, dilatation pyélocalicielle | CDK13 | NM_003718.4:c.2141G>A | p.(Gly714Asp) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 25 | DI modérée, anomalie imagerie cérébrale, RCIU, dysmorphie, hypermétropie, xérose cutanée, lanugo | MTO1 | NM_001123226.1:c.1402G>A | p.(Ala468Thr) | Faux-sens | Homozygote | AR, hérité des parents | Oui (FBN1) |
| 26 | DI légère, neuropathie axonale sensitivo-motrice, microcéphalie, petite taille, anomalie des extrémités, scoliose | MORC2 | NM_014941.3:c.74C>T | p.(Ser25Leu) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 27 | RPM, macrocéphalie | PCGF2 | NM_007144.2:c.194C>T | p.(Pro65Leu) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 28 | DI légère, TSA, dysmorphie, anomalies des extrémités | DDX3X | NM_001356.4:c.865-2A>G | p.? | Variant affectant l'épissage | Hétérozygote | DLX, <i>de novo</i> | Non |

| | | | | | | | | |
|----|--|----------------|------------------------------------|----------------------|------------------------------|--------------|------------------------|---------------|
| 29 | RPM, laryngomalacie, dysmorphie, hygroma, hypoplasie rénale bilatéral, RGO | STAG1 | NM_005862.2:c.1102G>A | p.(Glu368Lys) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 30 | Surdité profonde de perception bilatérale, myopie forte | SLTRK6 | NM_032229.2:c.3G>C | p.? | Perte du codon d'initiation | Homozygote | AR, hérité des parents | Non |
| 31 | RPM, hypoplasie du vermis cérébelleux, retard pubertaire, FLP, dent néonatale, dysmorphie, anomalie des extrémités, hypospadias, surdité | TRRAP | NM_001244580.1:c.3103C>G | p.(Arg1035Gly) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Oui (CYP2C19) |
| 32 | RPM, ataxie, retard de croissance, cardiopathie, surdité, dysmorphie, anomalie des extrémités | ZMYND11 | NM_006624.5:c.1798C>T | p.(Arg600Trp) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 33 | DI, macrocéphalie, épilepsie, SMG, myopie | PIGC | NM_002642.3:c.12_13insTTGTGACTAACA | p.(Gln4_Pro5insLeu*) | Décalage du cadre de lecture | Homozygote | AR, hérité des parents | Oui (PROC) |
| 34 | RPM, dégradation progressive, paraplégie spastique, ataxie cérébelleuse | GNBI | NM_001282539.1 : c.239T>C | p.(Ile80Thr) | Faux-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 35 | Arthrogrypose, épiphyse ponctuée, anomalie cérébrale | RAPSN | NM_032645.4 :c.848T>C | p.(Leu283Pro) | Faux-sens | Homozygote | AR, hérité des parents | Non |
| 36 | Arthrogrypose, épiphyses ponctuées, anomalie cérébrale | PC | NM_000920.3 :c.1408C>T | p.(Gln470*) | Non-sens | Homozygote | AR, hérité des parents | Non |
| 37 | RPM, microcéphalie, dysmorphie, RCIU, retard staturo-pondéral | DSP | NM_004415.2 :c.1783C>T | p.(Gln595*) | Non-sens | Hétérozygote | AD, <i>de novo</i> | Non |
| 38 | RPM, CIA, dysmorphie, hyperlaxité, strabisme | HUWE1 | NM_031407.6 :c.12408C>G | p.(Ile4136Met) | Faux-sens | Hétérozygote | DLX, <i>de novo</i> | Oui (HFE) |

| | | | | | | | | |
|----|--|--------------|-----------------------------|-------------|---------------------------------|--------------|------------------------|-----|
| 39 | Syndrome uro-facial Encéphalopathie | LRIG2 | NM_001312686.1 : c.495-2A>C | p.? | Variant affectant l'épissage | Homozygote | AR, hérité des parents | Non |
| 40 | épileptique | GRIA3 | NM_000828.4 :c.1961C>T | p.Ala654Val | Faux-sens | Hétérozygote | DLX, <i>de novo</i> | Non |

Tableau 8 : Phénotype des patients ayant eu un DP et caractéristiques des variants de classe 4 et 5 expliquant le phénotype

4.3 Données Non Sollicitées

Concernant les données secondaires, 106/110 (96.4%) patients ont accepté la recherche des DNS des gènes du groupe 1 (gènes actionnables), 109/110 (98.2%) patients ont accepté la recherche des DS des gènes du groupe 2 (gènes importants pour le conseil génétique) et 104/110 (94.5%) patients ont accepté la recherche des DS des gènes du groupe 3 (gènes pharmacogénétiques).

Au total, 15 DNS de classe 4 ou 5 de la classification de l'ACMG ont été retrouvés chez 14 patients ce qui représente un taux de diagnostic de 12.7%. Ces DNS sont répertoriées dans le tableau 9. Cinq patients sur 106 (4.7%) ont présenté des variants pathogènes ou probablement pathogène dans les gènes du groupe 1 (gènes *TNNI3*, *LDLR*, *FBNI*, *PROC* et *SERPINC1*). Six patients sur 109 (4.6%) ont présenté des variants pathogènes ou probablement pathogènes dans les gènes du groupe 2 dont 5 dans le gène *HFE* et un patient dans le gène *CFTR* (responsable de la mucoviscidose). Quatre patients sur 103 (3.9%) ont présenté un variant homozygote pathogène dans les gènes *CYP2C9* (2 patients) et *CYP2C19* (2 patients). Un patient a présenté 2 DS, un variant pathogène dans le gène *LDLR* (Hypercholestérolémie familiale) du groupe 1 et le variant *CYP2C19*2* homozygote pathogène du groupe 3 (patient 7 du tableau X). Cinq patients sur 109 (4.6%) de la cohorte sont porteurs hétérozygote du variant p.Cys282Tyr dans le gène *HFE* (Hémochromatose de type 1). Le variant p.(Gly542*) (rs113993959) dans le gène *CFTR* a été retrouvé chez un patient. Nous n'avons pas retrouvé de patient porteur du variant le plus fréquent (p.Phe508del) dans le gène *CFTR* à l'état hétérozygote.

Parmi les patients porteurs d'une DNS, 5 patients présentaient également un DP (gène *KCNQ5*, *MTO1*, *TRRAP*, *PIGC* et *HUEWI*) soit 35.7% (patient 6, 8, 9, 10 et 13 du tableau X). Parmi les familles où l'on a retrouvé une DNS, 2 étaient consanguines et présentaient un DP de maladie autosomique récessive (variants pathogènes homozygotes dans les gènes *MTO1* et *PIGC*) et une DS à l'état hétérozygote dans les gènes *FBNI* (maladie de Marfan) et *PROC* (coagulopathie) (variant 8 et 10 du tableau 9).

Concernant les DNS du groupe 1, les variants ont été ou vont être vérifiés en séquençage Sanger au sein du laboratoire spécialisé dans la pathologie correspondante. Le variant p.(Gln94*) à l'état hétérozygote dans le gène *TNNI3* (gène responsable de cardiomyopathie dilatée autosomique dominante) chez le patient 1 du tableau 9 a été classé

en classe 4 après avis auprès d'un spécialiste et une consultation de cardiopédiatrie a été demandée. L'échographie transthoracique n'a pas retrouvée de cardiomyopathie dilatée chez ce patient et un suivi régulier est prévu. Le variant p.(Glu418Lys) à l'état hétérozygote dans le gène *LDLR* (gène responsable d'hypercholestérolémie familiale) chez le patient 7 a été classée en classe 4 après avis auprès d'un spécialiste et un bilan lipidique et une consultation d'endocrinologie ont été demandée. Le variant p.(Arg609Cys) à l'état hétérozygote dans le gène *FBNI* (gène responsable du syndrome de Marfan) chez le patient 8 a été classé en classe 4 après avis auprès d'un spécialiste et une consultation génétique spécialisée ainsi qu'une échographie transthoracique ont été demandée. Le variant p.(Val339Met) à l'état hétérozygote dans le gène *PROC* (gène responsable de thrombophilie par déficit en protéine C) chez le patient 10 a été classée en classe 4 après avis auprès d'un spécialiste et une consultation d'hématologie a été demandé. Le variant p.(Pro73Leu) à l'état hétérozygote dans le gène *SERPINC1* (gène responsable de thrombophilie par déficit en antithrombine III) chez le patient 12 a été classé en classe 5 après avis auprès d'un spécialiste et une consultation en hématologie sera demandée.

Concernant les patients hétérozygotes pour la mutation p.(Cys282Tyr) dans le gène *HFE* responsable de l'hémochromatose de type 1, leur présence a été vérifiée en séquençage Sanger au sein du laboratoire des maladies héréditaire du métabolisme au CBPE et validé par le docteur David Cheillan. Le variant p.(Gly542*) dans le gène *CFTR* va être vérifié en séquençage Sanger. Un conseil génétique va être prodigué lors de la consultation de rendu des résultats.

Concernant les patients porteurs de variants dans les gènes de pharmacogénétique, ils vont être vérifiés en séquençage Sanger. De plus, tous les patients ont ou vont bénéficier d'une consultation de pharmacogénétique au centre hospitalo-universitaire de Dijon. Les variants retrouvés dans ces gènes permettent de classer les patients dans le groupe des « métaboliseurs lents » pour les médicaments métabolisés par le cytochrome p450-2C9 et p450-2C19, ce qui peut entraîner des variations de concentration sanguine du principe actif par rapport à un métaboliseur normal. Ceci peut aboutir à un potentiel surdosage ou à un sous-dosage, selon le métabolisme du médicament. Pour les patients 2 et 6 du tableau 9 porteur des variants homozygotes *CYP2C9*2* et *CYP2C9*3*, il existe un risque hémorragique accru avec les anti-inflammatoires non stéroïdiens et la warfarine, un risque d'hypoglycémie accru avec les sulfamides, un risque d'hypotension artérielle avec certains agonistes de l'angiotensine II (losartan, irbesartan) ainsi qu'un risque toxique de certains médicaments utilisés en

neurologie comme l'amitriptyline, la fluoxétine, la phénytoïne ou l'acide valproïque et la fluvastatine. Pour les patients 7 et 9 porteurs de variants homozygotes *CYP2C19*2*, il existe un risque accru d'effets indésirables par surdosage pour les inhibiteurs sélectifs de la recapture de la sérotonine, certains antidépresseurs tricycliques, les inhibiteurs de la monoamine oxydase et la phénytoïne. Il existe également un risque d'effets toxiques avec l'utilisation d'inhibiteurs de la pompe à protons, le voriconazole et un risque de diminution d'efficacité par sous-dosage pour le clopidogrel (pro-drogue). Pour tous ces patients, la présence de ces variants ne présente pas une contre-indication absolue mais une précaution d'emploi.

| Patient | Consanguinité | Gène | Nomenclature HGVS | Conséquence protéique | Type de mutation | Zygotie | Mode de transmission | DP | Groupe de DS |
|---------|---------------|-----------------------------|----------------------------|-----------------------|------------------|--------------|----------------------|----------------------|--------------|
| 1 | Non | <i>TNNI3</i> | NM_000363.4:c.280C>T | p.(Gln94*) | Non-sens | Hétérozygote | AD, AR | Non | 1 |
| 2 | Non | <i>HFE p.(Cys282Tyr)</i> | NM_000410.3:c.845G>A | p.(Cys282Tyr) | Faux-sens | Hétérozygote | AR | Non | 2 |
| 3 | Non | <i>CYP2C9*3 homozygote</i> | NM_000771.3:c.1075A>C | p.(Ile359Leu) | Faux-sens | Homozygote | AR | Non | 3 |
| 4 | Non | <i>HFE p.(Cys282Tyr)</i> | NM_000410.3:c.845G>A | p.(Cys282Tyr) | Faux-sens | Hétérozygote | AR | Non | 2 |
| 5 | Non | <i>HFE p.(Cys282Tyr)</i> | NM_000410.3:c.845G>A | p.(Cys282Tyr) | Faux-sens | Hétérozygote | AR | Non | 2 |
| 6 | Non | <i>CYP2C9*2 homozygote</i> | NM_000771.3:c.430C>T | p.(Arg144Cys) | Faux-sens | Homozygote | AR | Oui (<i>KCNQ5</i>) | 3 |
| 7 | Non | <i>LDLR</i> | NM_000527.4:c.1252G>A | p.(Glu418Lys) | Faux-sens | Hétérozygote | AD | Non | 1 |
| 7 | Non | <i>CYP2C19*2 homozygote</i> | NM_000769.2:c.681G>A | p.(Pro227=) | Synonyme | Homozygote | AR | Non | 3 |
| 8 | Oui | <i>FBNI</i> | NM_000138.4:c.1825C>T | p.(Arg609Cys) | Faux-sens | Hétérozygote | AD | Oui (<i>MTO1</i>) | 1 |
| 9 | Non | <i>CYP2C19*2 homozygote</i> | NM_000769.2:c.681G>A | p.(Pro227=) | Synonyme | Homozygote | AR | Oui (<i>TRRAP</i>) | 3 |
| 10 | Oui | <i>PROC</i> | NM_000312.3 : c.1015G>A | p.(Val339Met) | Faux-sens | Hétérozygote | AD | Oui (<i>PIGC</i>) | 1 |
| 11 | Non | <i>HFE p.(Cys282Tyr)</i> | NM_000410.3:c.845G>A | p.(Cys282Tyr) | Faux-sens | Hétérozygote | AR | Non | 2 |
| 12 | Non | <i>SERPINC1</i> | NM_000488.3_c.218C>T | p.(Pro73Leu) | Faux-sens | Hétérozygote | AD | Non | 1 |
| 13 | Non | <i>HFE p.(Cys282Tyr)</i> | NM_000410.3:c.845G>A | p.(Cys282Tyr) | Faux-sens | Hétérozygote | AR | Oui (<i>HUEWI</i>) | 2 |
| 14 | Non | <i>CFTR</i> | NM_000492.3:c.1624G>T | p.(Gly542*) | Non-sens | Hétérozygote | AR | Non | 2 |

Tableau 9 : Caractéristiques des variants pathogènes et probablement pathogènes des DNS

4.4 Analyse des CNV

Concernant l'analyse des CNV, elle a été réalisée sur 96 patients par le service de bioinformatique de Dijon. 6 CNV ont été retenus (6.25%) (Tableau 10). 3 CNV étaient des VOUS déjà connus pour ces patients détectés par ACPA auparavant et mis en évidence par l'analyse des données de séquençage. 1 CNV était un CNV bénin sur le chromosome X chez une fille également détecté en ACPA. 2 CNV étaient pathologiques.

Concernant ces 3 CNV pathologiques, ils avaient été détectés en ACPA et classés comme bénin ou VOUS mais devant la réanalyse des données et l'avancement des connaissances sur le caractère pathologique des CNV, ils ont pu être reclassés suite à cette analyse. Tous les CNV ont été vérifiés en PCR quantitative. 2 CNV détectés par l'analyse bioinformatique sur les données de séquençage ont été infirmés en PCR quantitative.

Le premier CNV pathologique correspondait à une délétion 16p11.2 récurrente proximale de 417 kb emportant de nombreux gènes dont le gène *TBX6*, chez une patiente présentant une DI syndromique. Cette délétion est connue comme étant un facteur de prédisposition aux troubles neurodéveloppementaux et explique en partie le phénotype de la patiente. Elle est héritée du père qui est sain. Ce CNV avait été détecté sur la ACPA réalisée en 2007 et classé comme polymorphisme puisque le caractère pathogène en tant que facteur de prédisposition aux troubles neurodéveloppementaux de cette anomalie n'a été démontré dans la littérature qu'en 2008 (50,51).

Le second CNV pathologique était une délétion de 341 kb en 10q24.32 emportant plusieurs gènes dont le gène *TRIM8* récemment impliqué dans les encéphalopathies épileptiques dans la littérature (52,53). Ce CNV détecté sur une ACPA réalisée en 2012 avait été rendu comme VOUS et était hérité du père sain porteur de ce CNV en mosaïque.

Le troisième CNV pathologique était une duplication intragénique *de novo* emportant 5 exons du gène *KDM6A* responsable de 5% des syndromes de Kabuki expliquant le phénotype de la patiente qui présentait une DI syndromique et certains des signes dysmorphiques rapportés dans ce syndrome (54).

Deux CNV *de novo* classés VOUS détectés en ACPA ont également été détectés sur les données du séquençage. Le premier était une délétion de 12 kb dans le gène *PLEKHG4B* qui est un gène non rapporté en pathologie humaine (*guanine nucleotide exchange factor*) chez un jeune patient atteint de TSA. Le deuxième CNV classé VOUS est une délétion de 1

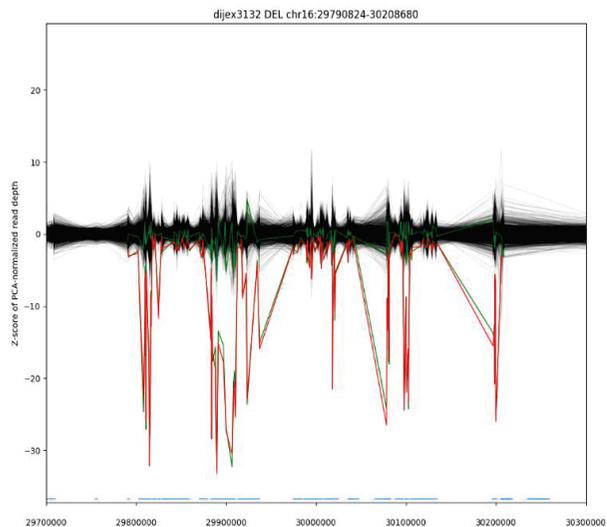
Mb sur le chromosome X emportant plusieurs gènes dont aucun n'est associé à un retard mental chez un patient présentant un retard psychomoteur.

Enfin, une délétion de 608 Kb du gène *DMD* (responsable de la myopathie de Duchenne) sur le chromosome X a été retrouvée chez une fille héritée de sa mère asymptomatique. Ce CNV a été classé bénin puisqu'il s'agit d'un variant pathologique chez les garçons. Ceci correspondant à une découverte secondaire non liée au phénotype de la patiente. Les représentations graphiques des différents CNV sont répertoriées dans la figure 21.

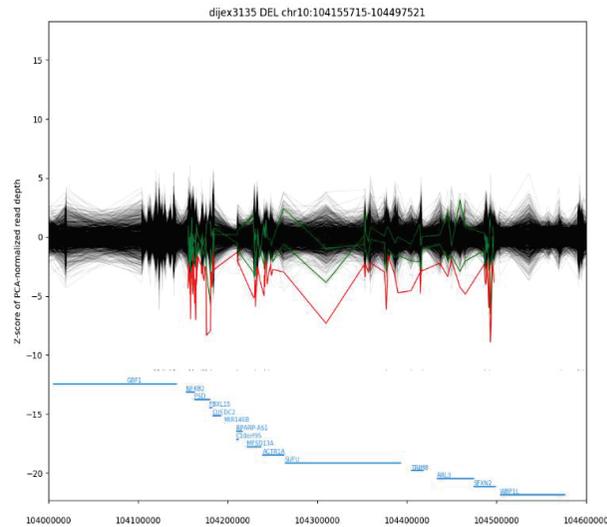
En additionnant les taux diagnostiques obtenus pour l'analyse des SNV et celui obtenu pour l'analyse des CNV, on obtient un taux diagnostique concernant les données primaires de 38.2%.

| Patient | Clinique | Intervalle | Taille du CNV en Kb | Bande cytogénétique | Type de CNV | Contenu en gène | Mode de transmission | Résultat ACPA | Classification CNV |
|---------|--|---------------------------|---------------------|---------------------|-------------|---|-----------------------------------|---------------|--|
| 1 | DI, épilepsie, retard staturopondéral, dysmorphie, anomalie des extrémités | chr16:29790824-30208680 | 417 | 16p11.2 | Délétion | <i>ZG16, KIF22, MAZ, PRRT2, PAGRI, MVP, CDIP T, CDIPTOSP, SEZ6L2, ASPHD1, KCTD13, TME M219, TAOK2, HIRIP3, INO80E, DOC2A, C16orf92, FAM57B, LOC112694756, ALDOA, PPP4C, TBX6, YPEL3, LOC101928595, GDPD3, MAPK3, CORO1A, LOC606724, BOLA2, BOLA2B, SLX1B, SLX1A, SLX1A-SULT1A3, SLX1B-SULT1A4</i> | Hérité du père | Bénil | PIEV (Facteur de susceptibilité) |
| 2 | DI, épilepsie, ectopie rénale | chr10:104155715-104497521 | 341 | 10q24.32 | Délétion | <i>NFKB2, PSD, FBXL15, CUEDC2, MIR146B, RP ARPAS1, C10orf95, MFSD13A, ACTR1A, SUFU, TRIM8, ARL3, SFXN2</i> | Hérité du père (père en mosaïque) | VOUS | Pathologique (gène <i>TRIM8</i>) |
| 3 | DI modérée non syndromique | chrX:44820527-44896936 | 76 | Xp11.3 | Duplication | <i>KDM6A</i> | <i>De novo</i> | VOUS | Pathologique |
| 4 | TSA | chr5:151626-163665 | 12 | 5p15.33 | Délétion | <i>PLEKHG4B</i> | <i>De novo</i> | - | VOUS |
| 5 | RPM, laryngomalacie | chrX:133700072-134715563 | 1015 | Xq26.3 | Délétion | <i>PLAC1, FAM122B, FAM122C, MOSPD1, LINC02243, SMIM10, RTL8B, RTL8C, RTL8A, SMIM10L2B, ETDB, SMIM10L2B-AS1, CT55, ZNF75D, ETDA, ETDC, ZNF449, LOC100506790, SMIM10L2A, INTS6L-AS1, INTS6L</i> | <i>De novo</i> | VOUS | VOUS |
| 6 | DI, épilepsie | chrX:31341713-31950346 | 608 | Xp21.1Xp21.2 | Délétion | <i>DMD</i> | Hérité de la mère | Bénil | Bénil (pathologie récessive lié au chromosome X) |

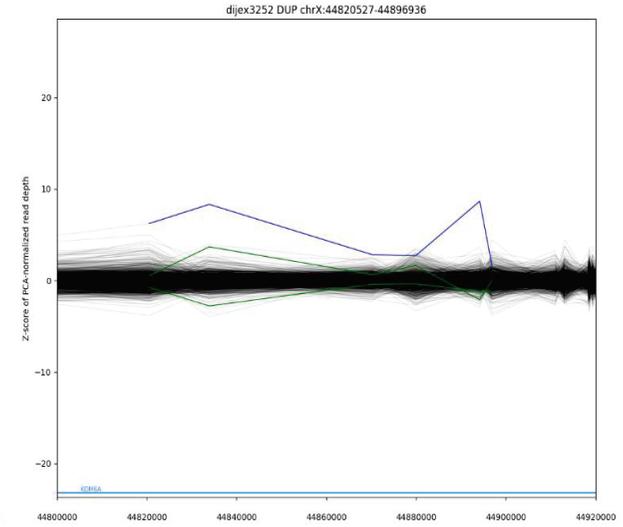
Tableau 10 : Caractéristiques des CNV retenus lors de l'analyse sur les données de séquençage de l'exome



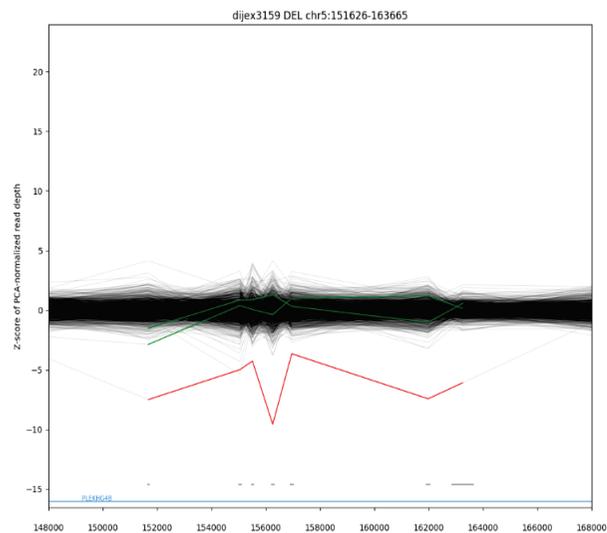
Patient 1 : délétion 16p11.2



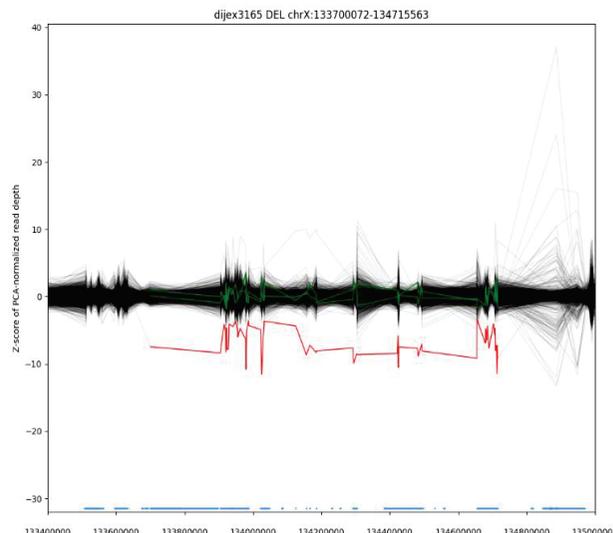
Patient 2 : délétion 10q24.32 (*TRIM8*)



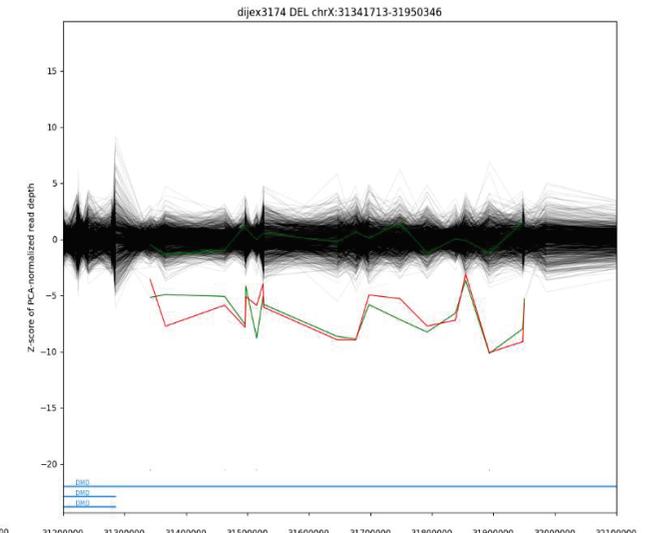
Patient 3 : Duplication Xp11.3 (*KDM6A*)



Patient 4 : délétion 5q15.33 (*PLEKHG4B*)



Patient 5 : délétion Xq26.3



Patient 6 : délétion Xp21.1p21.2 (*DMD*)

Figure 21 : Représentation graphique des CNV retenus

La ligne rouge représente une délétion chez le patient, la ligne bleue une duplication chez le patient et les 2 lignes vertes sont celles des parents. L'axe des abscisses représente les coordonnées génomiques et l'ordonnée, une déviation en z-score par rapport à la base de données de patients de Dijon.

5 DISCUSSION

L'analyse du séquençage de l'exome chez 110 patients lyonnais porteur d'une AD/DI a permis de mettre en évidence chez 42 patients (38.2%) une cause génétique expliquant leur phénotype. Chez 39 d'entre eux (35.5%), un SNV pathogène ou probablement pathogène dans un gène responsable de trouble du neurodéveloppement ou plus largement d'anomalies du développement a été mis en évidence. Chez 3 d'entre eux, un CNV pathogène a permis d'expliquer le phénotype. En plus des données primaires mise en évidence qui expliquent le phénotype du patient, des DNS ont également été mis en évidence chez 14 patients (12.7%) concernant les 3 catégories de gènes analysés. 4.7 % (5/106) patients ont présenté des variants dans des gènes concernant le groupe 1, c'est-à-dire les gènes actionnables de prédisposition à des pathologies à révélation plus tardive. 5.5% (6/109) ont présenté des variants dans des gènes du groupe 2 c'est-à-dire des gènes de pathologies récessives fréquentes ou de pathologies récessives liées à l'X avec un risque de transmettre une pathologie à sa descendance. 3.9% (4/103) ont présenté des variants dans des gènes du groupe 3 (gènes de pharmacogénétique).

5.1 Implication des SNV dans le diagnostic primaire de la cohorte

Comme rappelé précédemment, l'analyse du séquençage de l'exome chez 110 patients lyonnais porteur d'une AD/DI a permis de mettre en évidence chez 39 d'entre eux (35.5%), un SNV pathogène ou probablement pathogène (de classe 5 ou 4 selon les critères de l'ACMG) dans un gène responsable de trouble du neurodéveloppement ou plus largement d'anomalies du développement. Très récemment, Srivastava et al, 2019 (4) ont réalisé une méta-analyse comparant le taux diagnostique du séquençage de l'exome en pratique clinique des différentes cohortes publiées de patients présentant des troubles du neurodéveloppement (Déficience intellectuelle et/ou Trouble du spectre autistique associé ou non à des malformations). Après avoir sélectionné 30 articles, ils retrouvent un taux diagnostique moyen de 36% au total, un taux de 31% pour les troubles neurodéveloppementaux isolés et un taux de 53% pour les troubles neurodéveloppementaux associés à des malformations. Dans notre cohorte, le taux diagnostique du groupe de patients présentant une DI et/ou TSA isolés est de 10,5% et celui du groupe de patients présentant une DI syndromique est de 39,1%. Ces taux

pris séparément sont différents, cela est sûrement dû à la classification des patients dans chacun des groupes puisque que la distinction d'une DI syndromique d'une DI isolée est floue car certains signes peuvent sembler discrets. Ce consortium d'experts rappelle les recommandations de l'ACMG de 2010 concernant l'exploration d'une DI ou un TSA par la réalisation d'une ACPA et d'une recherche d'un syndrome de l'X fragile (sauf pour les filles présentant un TSA isolé) en première intention. Il y a aussi dans certains cas des indications pour tester les gènes *MECP2* et *PTEN* en première intention. Les experts concluent en rappelant que le taux diagnostique de 36 % retrouvé dans la méta-analyse est supérieur à celui de l'ACPA (15-20%) et proposent donc d'adopter le séquençage de l'exome en première intention devant une DI isolée ou syndromique d'autant plus que les algorithmes d'analyse des CNV sont de plus en plus associés aux pipelines bioinformatiques d'analyses des exomes cliniques et permettent donc de détecter les CNV. Ce nouvel algorithme proposé par cette équipe est représenté sur la figure 22.

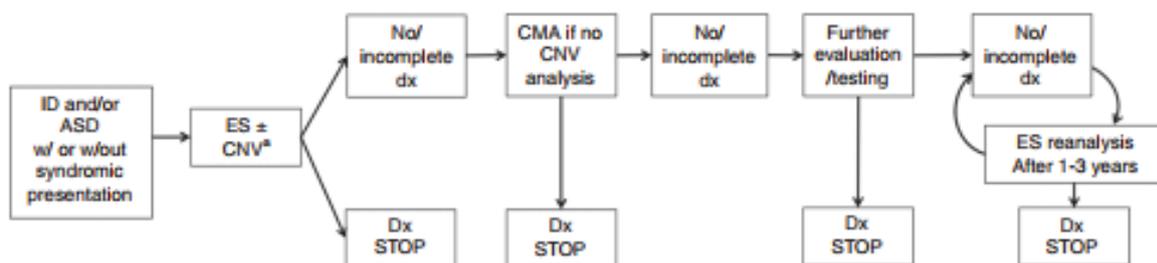


Figure 22 : Algorithme de diagnostic proposé par Srivastava et al. En incorporant le séquençage de l'exome dans l'exploration des DI et des TSA.

Comme précisé dans l'introduction, l'exploration des DI au sein du laboratoire de génétique de Lyon passait auparavant par la réalisation d'un panel de gènes d'environ 500 gènes impliqués dans la DI, si l'ACPA préalablement réalisée, était négative. Le projet FIND a été l'occasion d'opter pour une approche par le séquençage de l'exome pour l'exploration des AD/DI au sein du laboratoire de génétique du CHU de Lyon. Pour comparer, le panel DI comportait 450 gènes et le taux diagnostique de ce panel était d'environ 30%. Concernant le panel épilepsie, il comporte 146 gènes dans la V4 actuelle et le taux diagnostique est de 27.6% (V1-V3). Pour rappel le taux diagnostique de la cohorte du projet FIND concernant les patients présentant une encéphalopathie épileptique, est de 45.5%, bien que ce ne soit pas la

seule indication du panel épilepsie (épilepsie familiale, syndrome épileptique reconnaissable, épilepsie au premier plan...) et qu'une comparaison est donc délicate. Un article récent de Palmer et al, 2018 retrouve un taux diagnostique de 50 % dans une cohorte de 32 patients présentant une encéphalopathie épileptique. Concernant les groupes AD sans DI et le groupe avec des troubles neurologiques au premier plan, ceux-ci ont un effectif trop faible, respectivement de 7 et 4 patients, pour pouvoir en tirer des conclusions significatives.

La formation de ces groupes phénotypiques peut également être discutée puisque la définition d'un DI syndromique ou non syndromique reste floue avec la présence de certains signes malformatifs mineurs ou non détectés. Pour les pathologies pour lesquels les signes neurologiques sont au premier plan, dans certains cas un panel spécifique existe. Pour les AD sans DI, l'hétérogénéité clinique et génétique est telle qu'il est difficile d'analyser les données de ce groupe.

5.2 Implication des CNV dans le diagnostic primaire de la cohorte

Tous les patients de la cohorte ont eu au préalable une ACPA dans le cadre de l'exploration de leur AD/DI qui n'a pas retrouvé de CNV pathogènes ou probablement pathogènes pouvant expliquer leur phénotype. Les patients présentant un CNV de classe 3 (VOUS) pouvaient bénéficier de l'exome. L'analyse des CNV sur les données de l'exome a permis de retrouver les CNV de classe 3 des patients détectés au préalable en ACPA. Cette analyse à l'aveugle a permis de retenir 6 CNV chez 96 patients préalablement détectés par l'ACPA (6,25%). Parmi ces 6 CNV, 3 ont fait l'objet d'une reclassification suite aux données récentes de la littérature et sont ainsi passés de CNV de signification indéterminée (Classe 3 VOUS) à des CNV pathogènes. L'analyse des CNV est donc efficace sur les données de séquençage de l'exome puisqu'elle a permis de mettre en évidence les CNV préalablement détectés en ACPA. Aucun CNV pathogène non préalablement détecté par l'ACPA n'a été mis en évidence lors de cette analyse qui représente tout de même un intérêt pour la détection des « petits CNV » inférieurs à la résolution de l'ACPA ou de CNV présents dans des régions non couvertes par la puce et donc non détectables. L'analyse n'a été effectuée que sur 96 patients puisque certaines données provenant de Dijon étaient incomplètes et certaines analyses étaient trop bruitées pour être interprétées. Cette technique nécessite donc encore des améliorations afin d'être plus performante. Cette analyse montre également l'intérêt de la relecture des

résultats des ACPA avant de poursuivre les explorations génétiques face aux avancées rapides des connaissances dans le domaine de la génétique.

L'addition des taux diagnostiques de la ACPA de 15-20% et de celui du séquençage de l'exome (environ 36%) permet de poser un diagnostic étiologique dans 50% des cas de DI (4) (en enlevant les cas de Trisomie 21 qui représentent environ 9-10 % des DI et le syndrome de l'X-fragile qui représentent environ 1% des DI (55)). Cependant l'implémentation de l'analyse des CNV dans les pipelines bioinformatiques de l'analyse des données de l'exome est de plus en plus fréquente (56). Une étude de Visser et al, 2017 sur 105 patients présentant des troubles du neurodéveloppement bénéficiant d'un séquençage de l'exome avec analyse des CNV retrouvent un taux diagnostique pour l'implication d'un CNV pathogène de 6.4%. De plus, dans cette cohorte, 2 CNV pathogènes qui n'avaient pas été détectés par ACPA ont pu être mis en évidence (région non couverte par les oligonucléotides de la puce) (21). Dans un second article, le taux diagnostique est de 16.7% concernant l'implication d'un CNV dans le phénotype de trouble du neurodéveloppement (57). Ceci conforte encore plus la proposition de Srivastava et al., de proposer l'exome en première intention pour le diagnostic des AD/DI si l'analyse des CNV est possible sur les données de séquençage. De plus, l'analyse des CNV couplée à celle des SNV dans la même technique permet plus facilement de mettre en évidence des pathologies récessives dues à un mécanisme hétérozygote composite expliqué par un SVN sur un allèle du gène et à un CNV sur l'autre. Il est cependant intéressant de noter que la recherche de délétion et de duplication est plus sensible sur des données de séquençage de panels que sur l'exome car la profondeur est plus grande et la couverture meilleure. Ceci permet de mettre notamment en évidence des anomalies en mosaïque plus facilement.

Cependant, l'analyse des CNV sur les données de séquençage de l'exome présente des limites. En effet, il est difficile de détecter des délétions et/ou des duplications dans les régions riches en GC, dans des pseudogènes ou encore les duplications segmentaires. De plus, des CNV présents dans des régions non codantes comme des régions promotrices, des régions régulatrices ou introniques ne pourront pas être détectés car ces régions ne seront pas séquençées. Certaines régions cibles peuvent également présenter une faible déviation par rapport à la norme et ne pas être détectées. Si certaines régions sont mal couvertes, elles peuvent également entraîner des faux positifs. Pour finir, les régions très polymorphiques (notamment les duplications communes) peuvent ne pas être détectées même si cela n'est pas forcément délétère puisque ces duplications sont des polymorphismes (33).

5.3 Recherche des Données Non Sollicitées dans la cohorte

La recherche des DNS a permis de mettre en évidence chez 14 patients (12.7%) des variants pathogènes ou probablement pathogènes concernant les 3 catégories de gènes analysés. 4.7 % (5/106) patients ont présenté des variants dans des gènes concernant le groupe 1, c'est-à-dire les gènes actionnables de prédisposition à des pathologies à révélation plus tardives. 5.5% (6/109) ont présenté des variants dans des gènes du groupe 2, c'est-à-dire des gènes de pathologies récessives fréquentes ou de pathologies récessives liées à l'X avec un risque de transmettre une pathologie à sa descendance. 3.9% (4/103) ont présentés des variants dans des gènes du groupe 3 (gènes de pharmacogénétique).

Cette analyse des DNS demande un temps supplémentaire aussi bien clinique que biologique. En effet, lors de la consultation pré-test, l'information doit être donnée au patient et doit être claire concernant ces DNS. On peut se poser la question de la clarté de l'information devant le nombre de gènes recherchés et leur diversité en termes de types de pathologies allant aussi bien de prédisposition au cancer ou aux cardiopathies que des gènes de pharmacogénétique ou des gènes de pathologies récessives ou liées à l'X à risque de transmission. Un consentement spécifique doit être également recueilli. Le choix est toujours laissé au patient de rechercher ces DNS ou non. Pour être inclus au projet FIND, il fallait cependant accepter la recherche des DNS dans au moins une des 3 catégories de gènes des DNS. Le temps nécessaire à l'explication de ces examens génomiques justifie l'émergence d'un nouveau métier celui de conseiller en génomique pour soutenir le généticien dans cette démarche. Le temps de rendu de résultats est aussi beaucoup plus long si la recherche de DNS est positive puisqu'il faut détailler les conséquences du résultat, le conseil génétique et adresser le patient à une consultation spécialisée et/ou des examens complémentaires. Sur le plan biologique, un temps supplémentaire est nécessaire aussi bien pour l'analyse bio-informatique (notamment des variants de pharmacogénétique) que pour l'interprétation des variants qui nécessite tout d'abord une première analyse par le biologiste qui lit l'exome et ensuite la validation par un biologiste « expert » concernant le gène impliqué. L'implication de ces biologistes experts pour l'interprétation des DNS est primordiale. En effet, dans ce projet, seul les variants rapportés sur la base de données HGMD ou Clinvar comme pathogènes ou probablement pathogènes pouvaient être rendus. Cependant, une expertise était

parfois nécessaire devant certains variants avec des étiquettes contradictoires ou encore non rapportés mais semblant très probablement pathogènes comme des variants non-sens dans certains gènes soumis à l'haploinsuffisance. De plus, le temps biologique est d'autant plus allongé qu'il faut ensuite confirmer par une seconde méthode (séquençage SANGER, long-range PCR). L'étude FIND s'attardera à quantifier ce temps supplémentaire aussi bien clinique et biologique ainsi que le recours secondaire à des examens complémentaires et des consultations spécialisées.

La recherche de ces DNS reste un sujet très débattu en France et à l'étranger. Ceci soulève des questions éthiques autant au niveau individuel que sociétal. Récemment, l'avis 130 du Comité Consultatif National d'Ethique datant du 29 mai 2019 a donné un avis favorable pour la recherche des DS concernant les gènes actionnables pour lesquels une stratégie préventive ou thérapeutique est possible, tout en rappelant la nécessité d'informer les patients et leur famille des conséquences de la recherche de ces DS au moment de la rédaction initiale du consentement au préalable. La Société Française de Médecine Prédictive et Personnalisée a publié récemment des recommandations concernant les données incidentes touchant des gènes de prédisposition au cancer (40). Pour l'instant, il n'existe pas de recommandations européennes proposant une liste de gènes minimale. Chaque laboratoire peut faire le choix d'une stratégie locale consignée dans des procédures. Cependant, le cadre légal français (article 16-10 du code civil et R.1131-5 du code de santé publique) précise que « lorsque l'examen génétique conduit à révéler fortuitement d'autres informations que celles recherchées, le droit en vigueur, pour protéger le patient d'informations inutiles, angoissantes ou dont la révélation n'est pas désirée, *n'est pas en faveur de la transmission d'informations autres que celle initialement recherchée et pour laquelle le patient a consenti à la réalisation de l'examen. Dans ces conditions, il appartient au médecin de déterminer au cas par cas dans le cadre d'un colloque singulier avec son patient la conduite à tenir* ». Le cadre de recherche à travers le protocole FIND permet d'encadrer l'analyse des DNS en attendant une évolution législative à ce sujet.

Une autre problématique éthique soulevée par la recherche de ces DNS est celle de la notion de diagnostic pré-symptomatique. En effet, la plupart des patients explorés dans le cadre d'AD/DI sont des enfants. Or, découvrir une prédisposition à une pathologie sans lien avec la demande de l'examen entre dans le cadre du diagnostic pré-symptomatique. Chez les mineurs, celui-ci est autorisé s'il existe des mesures de prévention dans l'enfance. Ceci est

discutable pour certains gènes bien que cela puisse représenter une information importante pour les apparentés dans l'immédiat.

Sur le plan individuel, afin de mieux cerner les attentes et les préférences des patients et des professionnels de santé autour du séquençage de l'exome, l'étude SEQUAPRE réalisé entre 2015 et 2016 au sein des centres de références d'anomalies du développement de Dijon et Lyon a apporté quelques réponses. En effet, une première étude quantitative sur 513 patients a montré que ceux-ci étaient favorables à une information exhaustive concernant les données de signification inconnue et les DNS (58). Les résultats de l'étude FIND en cours visant à déterminer les attentes et le vécu des patients concernant la recherche des DNS lors du séquençage de l'exome permettront d'éclairer ce point à travers une étude qualitative (observation des consultations médicales et entretiens avec suivi des familles) et une étude quantitative avec passation de questionnaires et d'échelles standardisées (qualité de vie et dépression). Une étude basée sur des groupes focus de patients porteur de pathologies sans diagnostic ou de pathologies liées au gène de la liste de l'ACMG sur 21 familles ont montré que 8 d'entre elles étaient favorables à l'obtention de ces DNS et que 11 ne l'étaient pas en raison des conséquences psychologiques de tels résultats (2 familles ambivalentes). Ces entretiens ont permis de souligner l'importance de la transmission de l'information avant la réalisation du test génétique concernant ces DNS, de l'autonomie du patient dans le choix de les rechercher ou non, de l'importance d'un délai de réflexion et de la qualité des interactions entre les patients et leurs familles et les professionnels de santé (59).

Récemment, une étude rétrospective concernant les DNS menée par le CHU de Dijon sur 700 exomes montre des résultats similaires aux nôtres. Ils retrouvent la présence de variants pathogènes dans des gènes de prédisposition du groupe 1 dans 3.7 % des cas, des variants pathogènes dans des gènes du groupe 2 (porteur hétérozygote de pathologies récessives ou liées à l'X) dans 3.8% des cas et des variants pharmacogénétique (gène *CYP2C9* et *CYP2C19*) responsables d'un phénotype de métaboliseur lent dans 57 cas sur 700 (60). Cette étude souligne également le temps supplémentaire pour l'analyse de ces DNS avec un temps moyen de 6,3 minutes pour les gènes du groupe 1 et 1,3 minutes pour les gènes du groupe 2 et une moyenne de 37 minutes pour la réinterprétation des variants.

Aux Etats-Unis, le « BabySeq Project » montre une volonté de médecine prédictive et personnalisée dès la naissance chez des personnes saines. En effet, cette étude portait sur les impacts médicaux, comportementaux et économiques du séquençage du génome de nouveau-

nés sains ou admis en soins intensifs. Neuf pourcent des nouveau-nés présentaient des pathologies à révélation pédiatriques, 88% étaient porteurs hétérozygotes de variants dans des gènes de pathologies récessives et 5% présentait des variants des gènes de pharmacogénétique. Cette étude conclut donc sur l'efficacité de cette approche par rapport au dépistage néonatal actuel pour prévenir du risque de certaines maladies (61).

Pour conclure, la recherche des DNS a un impact sur l'organisation des soins avec le recours à des consultations génétiques et spécialisées, la nécessité de pipelines bio-informatiques supplémentaires et du temps d'interprétation et de validation biologique supplémentaires. De plus, cela nécessite des besoins de renforts cliniques avec notamment l'apparition de nouveaux métiers comme celui de conseiller en génomique. Il est nécessaire de poursuivre les recommandations françaises pour préciser et mieux encadrer ces recherches et de poursuivre celles sur les besoins et les attentes des patients ainsi que les conséquences à plus long terme du rendu de ces DNS.

5.4 Les Limites du séquençage de l'exome

Malgré la démonstration de l'efficacité diagnostique du séquençage de l'exome dans les AD/DI, celui-ci ne permet pas de détecter tous les types de variants responsables de pathologies. En effet, l'exome ne permet pas de détecter les remaniements chromosomiques équilibrés, les expansions de triplets, les variants introniques, les délétions ou duplications ne concernant qu'un ou deux exons (ceux-ci sont mieux détectés par les panels de gènes) ou encore les variants dans les régions non ou peu couvertes.

Le Plan France médecine génomique 2025 a fait le choix du séquençage du génome pour le diagnostic de certaines maladies. Les AD/DI devraient bientôt entrer dans les indications pour la prise en charge diagnostique de ces pathologies par ces plateformes. En attendant, le projet pilote « DEFIDIAG » va très prochainement pouvoir débuter. Ce projet, regroupant 12 centres d'inclusions français, a pour objectif de comparer le taux diagnostique causal d'origine génétique du séquençage de l'exome contre celui de la stratégie nationale de référence actuellement en place qui comporte une recherche du syndrome de l'X Fragile, la réalisation d'une ACPA et du panel DI 44 chez des patients présentant une DI bénéficiant pour la première fois d'une évaluation génétique. Les autres objectifs de cette étude seront aussi de comparer la stratégie du séquençage du génome en cas index (séquençage

uniquement du patient) VS le séquençage du génome en trio (séquençage du patient et de ses parents). Une analyse de l'impact économique, médicale et psychosociale sera également réalisée. Ce projet permettra de déterminer la meilleure stratégie de séquençage à haut débit à adopter pour le diagnostic des DI en France.

Si à l'avenir une stratégie de séquençage du génome en première intention était adoptée, cela engendrerait un remaniement profond de l'organisation à la fois des consultations de génétique clinique et à la fois des laboratoires de biologie moléculaire. En effet, l'équipe clinique devrait alors être renforcée avec des conseillers en génétique formés en génomique afin de soutenir les médecins et avoir un temps alloué à l'information précise et au recueil du consentement du patient. Les consultations de rendu de résultats seront également plus longues notamment si des DNS sont retrouvés afin d'expliquer ces résultats aux patients et de les orienter correctement vers les différents spécialistes. Ces spécialistes devront faire face à de nouveaux patients qui ne sont pas encore symptomatiques et devront également assurer un suivi à long terme ainsi qu'un dépistage familial.

Du point de vue du laboratoire, la stratégie du séquençage du génome en première intention nécessiterait un remodelage profond organisationnel avec la nécessité de former des techniciens de laboratoire au séquençage haut débit ainsi que d'obtenir de nouveaux séquenceurs à plus haut débit afin de répondre à la demande. Des infrastructures bio-informatiques ainsi qu'une équipe renforcée sont également nécessaires afin de répondre aux besoins informatiques incluant la capacité de parallélisation des analyses, la puissance de calcul importante, la mémoire vive et la capacité de stockage importante (<http://anddi-rares.org/assets/files/recommandations-gestion-donn%C3%A9es-ngs-anpgm.pdf>). De plus, une collaboration entre les différents laboratoires experts est indispensable pour rendre des résultats fiables concernant les DNS. Un financement conséquent est également indiscutable puisque le prix du séquençage d'un génome est bien élevé que celui d'une ACPA. En se référant aux tarifs des actes innovants hors nomenclature, un exome solo revient à 2205.9 euros alors qu'une ACPA revient à 1080 euros. Si on ne prend en compte que les réactifs alors un exome trio (qui est la stratégie qui a été adoptée ici) revient à 1500 euros (500 euros par personne) et la CGH à environ 250 euros.

De plus, une autre problématique soulevée est celle de la gestion des données puisque cette stratégie aurait pour conséquences une augmentation massive de l'ordre de téraoctets des données génétiques à conserver. Cette conservation nécessite une puissance informatique et la

création de logiciels spécifiques adaptés pour leur prise en charge. Il faudra déterminer quel type d'informations et combien de temps les conserver. A l'heure du règne des bases de données, du partage des informations sur internet et des réseaux sociaux, des réflexions éthiques sont primordiales pour assurer l'anonymisation et la protection de ces données (62). La gestion de ces masses de données ou « big data » est une question qui ne concerne pas seulement les données issues du séquençage de l'ADN humain et soulève la problématique de l'hégémonie du GAFA (Google, Amazon, Facebook et Apple) qui n'ont pas pour tradition de travailler avec des médecins ou des biologistes.

5.5 L'intérêt de la réanalyse des données du séquençage de l'exome

Les résultats obtenus sur l'analyse des CNV ont montré l'intérêt de réanalyser les données génomiques obtenues par ACPA au vu de l'évolution de la littérature scientifique et de la découverte de nouveaux gènes qui a permis de reclasser 3 CVN responsables de la pathologie des patients. Cette stratégie de réanalyse d'exams négatifs a également été appliquée sur les données de séquençage et plusieurs équipes ont déjà démontré son efficacité. Dans le cas de patients atteints d'AD, la réanalyse des données de séquençage permet d'augmenter le taux diagnostique d'environ 10% (63–65). Ces nouveaux diagnostics peuvent être dus à de nouvelles publications décrivant de nouveaux gènes impliqués en pathologie, à une meilleure description phénotypique des patients, à une réanalyse avec un pipeline bioinformatique plus récent qui permet de mettre en évidence de nouveaux variants ou des CNV et la mise à jour des bases de données. De plus, le partage de données sur des plateformes internationales comme « *Genematcher* » permet également, grâce à la collaboration entre plusieurs équipes, de valider fonctionnellement de nouveaux variants dans des gènes candidats.

Cette stratégie de réanalyse des données peut donc être appliquée sur les données de l'ACPA devant la présence de VOUS avant la réalisation d'un exome ou d'un génome. Si celui-ci est négatif, une réanalyse des données de séquençage à 1 an peut être envisagé après une réévaluation clinique. Ceci demandera un temps supplémentaire clinique pour réévaluer le phénotype mais également une réanalyse bio-informatique avec un pipeline plus récent et un temps biologique supplémentaire avec une réinterprétation des données au vu des avancées de la littérature. Dans leur étude sur 54 patients souffrant d'AD variées ayant bénéficiés d'une

réanalyse des données de l'exome à 12 mois, Ewans et al, 2019 montre un coût supplémentaire d'environ 586 \$US par patient (64).

5.6 L'avenir du Séquençage à Haut Débit

Les technologies préalablement mentionnées nécessitent avant de séquencer l'ADN une étape d'amplification. Cela est responsable d'erreurs de copies et de biais lors de la lecture des séquences. Depuis quelques années, le séquençage de 3^{ème} génération permet d'obtenir des fragments d'ADN de grande longueur et de s'affranchir de l'étape d'amplification. On parle également de « *single molecule sequencing* ». Deux nouvelles technologies dites de séquençage en temps réel ont su se démarquer sur le marché. La première a été développée par l'entreprise Pacific Bioscience (Menlo Park, USA) et permet d'observer de façon optique en temps réel la fluorescence lors de l'incorporation des nucléotides par une polymérase immobilisée au fond d'un puit. Le séquençage est dit en « *zero mode waveguide* ». Cette technologie permet de séquencer des fragments de 10 kb voire même 100 kb et peut identifier en plus des variants nucléotidiques, des remaniements de structure, des régions répétées et permet d'obtenir des informations sur la méthylation de l'ADN. Ces inconvénients sont le débit plus limité, le taux d'erreur plus élevé ainsi que son coût plus élevé par base par rapport aux séquenceurs de 2^{ème} génération (66). La deuxième approche est celle du séquençage de l'entreprise Oxford Nanopore Technologie (Oxford, Royaume-Uni) qui a lancé en 2014 un appareil portatif au format d'une clé USB baptisé le MinION qui permet de séquencer à la fois l'ADN et l'ARN. Avec cette technologie, la séquence d'un fragment d'ADN simple brin est déterminée lors de son passage dans un nanopore grâce à une enzyme (hélicase). La mesure de l'intensité du courant électrique lors du passage d'un nucléotide à travers ce pore permet de déterminer la séquence de l'ADN. Cette technologie permet des fragments allant jusqu'à 150 kb, cependant son taux d'erreur est lui aussi, pour l'instant, encore très important (67). Concernant leur application médicale, ces technologies méritent des améliorations avant de pouvoir remplacer les séquenceurs de 2^{ème} génération mais peuvent être une aide supplémentaire notamment dans le domaine de la recherche.

Le séquençage à haut débit permet également une nouvelle approche dite « *multiomics* » pour mieux appréhender de nombreuses pathologies humaines, notamment les maladies complexes, les cancers ou les pathologies à pénétrance incomplète. Cette approche permet d'analyser les données à la fois issues du génome, du transcriptome, du méthylome,

du microbiome, du protéome et du métabolome. L'intégration de ces données permet de mettre en évidence des processus biologiques ou des voies de signalisations causales de la pathologie. Cela permet également d'appréhender la régulation génétique ainsi que l'implication de certains facteurs environnementaux (68). Cette approche permettra à l'avenir de mieux comprendre de nombreuses pathologies humaines.

6 CONCLUSIONS

L'objectif de cette thèse était d'analyser les résultats du séquençage de l'exome chez les 110 patients porteurs d'AD/DI des Hospices Civils de Lyon inclus dans le projet FIND. L'analyse des résultats a porté d'une part sur les données primaires expliquant le diagnostic des patients à la fois par l'analyse des SNV et par celle des CNV, d'autre part sur l'analyse des DS portant sur des gènes actionnables pour lesquels une action préventive et/ou curative est possible à partir des données du séquençage de l'exome.

Trente-neuf variants pathogènes ont été mis en évidence dans des gènes connus responsables d'AD/DI expliquant le phénotype des patients. Le taux diagnostique concernant l'analyse des SNV dans cette cohorte de patients pour le diagnostic primaire est de 34.5 % ce qui correspond aux données de la littérature à ce sujet. De plus, le diagnostic primaire porte sur l'analyse des CNV qui a été réalisée sur les données du séquençage de l'exome pour la première fois à Lyon. Cette analyse à l'aveugle a permis de reclasser 3 CNV préalablement identifiés comme VOUS en ACPA en CNV pathogène à la suite des données récentes de la littérature. Cette partie des résultats souligne 2 points : d'une part, l'analyse des CNV est efficace sur les données de séquençage de l'exome et d'autre part, une relecture des résultats d'ACPA est nécessaire avant de poursuivre les explorations génétiques face aux avancées rapides des connaissances dans le domaine de la génétique.

L'analyse des données secondaires a permis de mettre en évidence l'intérêt médical de cette stratégie. En effet, des variants dans des gènes actionnables responsables de cardiopathies, d'hypercholestérolémies familiales ou de coagulopathies ont été mis en évidence et ont permis de mettre en place un dépistage familial ainsi qu'une surveillance particulière de ces patients. De plus, des patients porteurs hétérozygotes de variant dans les gènes *HFE* et *CFTR* ont pu bénéficier d'un conseil génétique et d'un dépistage familial. De plus, 4 patients ont présenté des variants pathogènes dans des gènes de pharmacogénétique ce qui permettra d'adapter leurs traitements pharmaceutiques à l'avenir afin d'éviter d'éventuels effets indésirables. De plus, l'analyse des DNS a nécessité une mise en place organisationnelle particulière aussi bien pour l'analyse bio-informatique et l'interprétation des données que pour la vérification des variants qui a nécessité une collaboration avec les laboratoires référents concernant chaque pathologie. Ceci a permis de souligner les difficultés auxquelles le laboratoire devra répondre à l'avenir devant l'avènement du séquençage haut débit et de la médecine prédictive et personnalisée.

La place de l'exome dans le diagnostic des AD/DI reste encore à définir notamment dans le contexte de l'avènement du plan France médecine génomique 2025 qui prévoit dans l'avenir le séquençage du génome pour le diagnostic des AD/DI. De plus, des recommandations et des réflexions éthiques à propos de la recherche des DNS sont encore nécessaires.

7 REFERENCES

1. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* janv 2016;17(1):9-18.
2. Déficiences intellectuelles. :1176.
3. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet.* 14 mai 2010;86(5):749-64.
4. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med Off J Am Coll Med Genet.* 11 juin 2019;
5. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA.* 12 nov 2014;312(18):1870-9.
6. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med Off J Am Coll Med Genet.* 2016;18(7):696-704.
7. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-Djebbar S, et al. Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin Genet.* 2016;89(6):700-7.
8. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 15 nov 2012;367(20):1921-9.
9. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet Lond Engl.* 10 nov 2012;380(9854):1674-82.
10. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 17 oct 2013;369(16):1502-11.
11. Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, et al. De novo mutations in moderate or severe intellectual disability. *PLoS Genet.* oct 2014;10(10):e1004772.
12. Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. *Ann Neurol.* oct 2014;76(4):473-83.

13. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 12 nov 2014;312(18):1880-7.
14. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet Lond Engl*. 4 avr 2015;385(9975):1305-14.
15. Nolan D, Carlson M. Whole Exome Sequencing in Pediatric Neurology Patients: Clinical Implications and Estimated Cost Analysis. *J Child Neurol*. 2016;31(7):887-94.
16. Monroe GR, Frederix GW, Savelberg SMC, de Vries TI, Duran KJ, van der Smagt JJ, et al. Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. *Genet Med Off J Am Coll Med Genet*. 2016;18(9):949-56.
17. Tarailo-Graovac M, Wasserman WW, Van Karnebeek CDM. Impact of next-generation sequencing on diagnosis and management of neurometabolic disorders: current advances and future perspectives. *Expert Rev Mol Diagn*. 2017;17(4):307-9.
18. Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, et al. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med Off J Am Coll Med Genet*. 2016;18(11):1090-6.
19. Kuperberg M, Lev D, Blumkin L, Zerem A, Ginsberg M, Linder I, et al. Utility of Whole Exome Sequencing for Genetic Diagnosis of Previously Undiagnosed Pediatric Neurology Patients. *J Child Neurol*. 2016;31(14):1534-9.
20. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet EJHG*. 2017;25(2):176-82.
21. Vissers LELM, van Nimwegen KJM, Schieving JH, Kamsteeg E-J, Kleefstra T, Yntema HG, et al. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet Med Off J Am Coll Med Genet*. 2017;19(9):1055-63.
22. Evers C, Stauffer C, Granzow M, Paramasivam N, Hinderhofer K, Kaufmann L, et al. Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Mol Genet Metab*. 2017;121(4):297-307.
23. Chérot E, Keren B, Dubourg C, Carré W, Fradin M, Lavillaureix A, et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet*. mars 2018;93(3):567-76.
24. Córdoba M, Rodríguez-Quiroga SA, Vega PA, Salinas V, Perez-Maturo J, Amartino H, et al. Whole exome sequencing in neurogenetic odysseys: An effective, cost- and time-saving diagnostic approach. *PloS One*. 2018;13(2):e0191228.

25. Palmer EE, Schofield D, Shrestha R, Kandula T, Macintosh R, Lawson JA, et al. Integrating exome sequencing into a diagnostic pathway for epileptic encephalopathy: Evidence of clinical utility and cost effectiveness. *Mol Genet Genomic Med.* 2018;6(2):186-99.
26. Kim SH, Kim B, Lee JS, Kim HD, Choi JR, Lee S-T, et al. Proband-Only Clinical Exome Sequencing for Neurodevelopmental Disabilities. *Pediatr Neurol.* 27 févr 2019;
27. Monies D, Abouelhoda M, AlSayed M, Alhassnan Z, Alotaibi M, Kayyali H, et al. The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum Genet.* 2017;136(8):921-39.
28. Hu X, Li N, Xu Y, Li G, Yu T, Yao R-E, et al. Proband-only medical exome sequencing as a cost-effective first-tier genetic diagnostic test for patients without prior molecular tests and clinical diagnosis in a developing country: the China experience. *Genet Med Off J Am Coll Med Genet.* 2018;20(9):1045-53.
29. Snoeijen-Schouwenaars FM, van Ool JS, Verhoeven JS, van Mierlo P, Braakman HMH, Smeets EE, et al. Diagnostic exome sequencing in 100 consecutive patients with both epilepsy and intellectual disability. *Epilepsia.* janv 2019;60(1):155-64.
30. Collège national des Enseignants et Praticiens de Génétique Médicale. Génétique médicale, enseignement thématique. Elsevier Masson. Issy-les-moulineaux; 2016. 347 p. (DFGSM2-3 Médecine).
31. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet.* 2015;16(3):172-83.
32. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* mai 2011;12(5):363-76.
33. Hehir-Kwa JY, Pfundt R, Veltman JA. Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Rev Mol Diagn.* 2015;15(8):1023-32.
34. Derhourhi M. Nouvelle technique de détection simultanée des variant ponctuels et des copy number variants dans l'obésité monogénique. :197.
35. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med Off J Am Coll Med Genet.* juill 2013;15(7):565-74.
36. van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, et al. Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. *Eur J Hum Genet EJHG.* juin 2013;21(6):580-4.
37. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet.* 3 oct 2013;93(4):631-40.

38. Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res.* mars 2015;25(3):305-15.
39. Piton A, Redin C, Mandel J-L. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am J Hum Genet.* 8 août 2013;93(2):368-83.
40. Pujol P, Perre PV, Faivre L, Sanlaville D, Corsini C, Baertschi B, et al. Guidelines for reporting secondary findings of genome sequencing in cancer genes: the SFMPP recommendations. *Eur J Hum Genet.* déc 2018;26(12):1732-42.
41. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cité 9 août 2019]. Disponible sur: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
42. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl.* 1 août 2014;30(15):2114-20.
43. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. 16 mars 2013 [cité 9 août 2019]; Disponible sur: <http://arxiv.org/abs/1303.3997>
44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl.* 15 août 2009;25(16):2078-9.
45. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* mai 2011;43(5):491-8.
46. SnpSift [Internet]. [cité 9 août 2019]. Disponible sur: <http://snpeff.sourceforge.net/SnpSift.html>
47. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med Off J Am Coll Med Genet.* mai 2015;17(5):405-24.
48. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet.* févr 2009;84(2):148-61.
49. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet.* 14 août 2011;43(9):838-46.
50. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med.* 14 févr 2008;358(7):667-75.

51. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet.* 15 févr 2008;17(4):628-38.
52. Assoum M, Lines MA, Elpeleg O, Darmency V, Whiting S, Edvardson S, et al. Further delineation of the clinical spectrum of de novo TRIM8 truncating mutations. *Am J Med Genet A.* 2018;176(11):2470-8.
53. Sakai Y, Fukai R, Matsushita Y, Miyake N, Saitsu H, Akamine S, et al. De Novo Truncating Mutation of TRIM8 Causes Early-Onset Epileptic Encephalopathy. *Ann Hum Genet.* juill 2016;80(4):235-40.
54. Bögershausen N, Gatinois V, Riehmer V, Kayserili H, Becker J, Thoenes M, et al. Mutation Update for Kabuki Syndrome Genes KMT2D and KDM6A and Further Delineation of X-Linked Kabuki Syndrome Subtype 2. *Hum Mutat.* 2016;37(9):847-64.
55. Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A.* 1 oct 2006;140(19):2063-74.
56. Retterer K, Scuffins J, Schmidt D, Lewis R, Pineda-Alvarez D, Stafford A, et al. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genet Med Off J Am Coll Med Genet.* août 2015;17(8):623-9.
57. Charng W-L, Karaca E, Coban Akdemir Z, Gambin T, Atik MM, Gu S, et al. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Med Genomics.* 19 2016;9(1):42.
58. Chassagne A, Pélissier A, Houdayer F, Cretin E, Gautier E, Salvi D, et al. Exome sequencing in clinical settings: preferences and experiences of parents of children with rare diseases (SEQUAPRE study). *Eur J Hum Genet EJHG.* mai 2019;27(5):701-10.
59. Houdayer F, Putois O, Babonneau ML, Chaumet H, Joly L, Juif C, et al. Secondary findings from next generation sequencing: Psychological and ethical issues. Family and patient perspectives. *Eur J Med Genet.* oct 2019;62(10):103711.
60. Thauvin-Robinet C, Thevenon J, Nambot S, Delanne J, Kuentz P, Bruel A-L, et al. Secondary actionable findings identified by exome sequencing: expected impact on the organisation of care from the study of 700 consecutive tests. *Eur J Hum Genet EJHG.* août 2019;27(8):1197-214.
61. Ceyhan-Birsoy O, Murry JB, Machini K, Lebo MS, Yu TW, Fayer S, et al. Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. *Am J Hum Genet.* 3 janv 2019;104(1):76-93.
62. Wagner JK, Mozersky JT, Pyeritz RE. « Use it or lose it » as an alternative approach to protect genetic privacy in personalized medicine. *Urol Oncol.* févr 2014;32(2):198-201.
63. Nambot S, Thevenon J, Kuentz P, Duffourd Y, Tisserant E, Bruel A-L, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies

and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med Off J Am Coll Med Genet.* 2018;20(6):645-54.

64. Ewans LJ, Schofield D, Shrestha R, Zhu Y, Gayevskiy V, Ying K, et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet Med Off J Am Coll Med Genet.* 2018;20(12):1564-74.
65. Li J, Gao K, Yan H, Xiangwei W, Liu N, Wang T, et al. Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation. *Gene.* 5 juin 2019;700:168-75.
66. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics.* oct 2015;13(5):278-89.
67. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol [Internet].* 25 nov 2016 [cité 12 oct 2019];17. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124260/>
68. Hasin Y, Seldin M, Lusk A. Multi-omics approaches to disease. *Genome Biol [Internet].* 5 mai 2017 [cité 26 mai 2018];18. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5418815/>

8 ANNEXE

Liste des gènes/variants à étudier définie dans le cadre du projet FIND (3 catégories) :

Groupe 1) Gènes actionnables : *ACTC1, APC, BRCA1, BRCA2, CACNA1S, COL3A1, DSC2, DSG2, DSP, FBN1, KCNH2, KCNQ1, LMNA, MEN1, MLH1, MSH2, MSH6, MYBPC3, MYH11, MYH7, MYL2, MYL3, MYLK, NF2, PCSK9, PKP2, PMS2, PRKAG2, PTEN, RB1, RET, RYR1, RYR2, SCN5A, SDHAF2, SDHB, SDHC, SDHD, SMAD3, STK11, TGFBR1, TGFBR2, TMEM43, TNNT2, TNNT3, TP53, TPM1, TSC1, TSC2, VHL, WT1, LDLR, APOB, MUTYH, GLA, ACVRL1, BMPR1A, CACNA1C, CACNB2, CDC73, CDH1, CNBP, DMPK, ENG, EPCAM, FH, FLCN, GCH1, GPD1L, HCN4, HMBS, KCNE1, KCNE2, KCNE3, KIT, MET, MLH3, PDGFRA, PLN, PROC, PRKAR1A, PROS1, PTCH1, RBM20, SCN1B, SCN3B, SERPINC1, SGCD, SMAD4, SMARCB1, TGFB3, ATP7B, BCHE, BLM, CASQ2, COQ2, COQ9, CPT2, GAA, HAMP, HFE (uniquement rs1800562), HFE2, IDUA, LDLRAP1, PAH, PCBD1, PTS, QDPR, SERPINA1, SLC25A13, SLC37A4, SLC7A9, DMD, EMD, OTC, F2 (rs1799963), F5(rs6025), F8, F9, VWF*

Groupe 2) Gènes importants pour le conseil génétique : *ACSL, AFF2, ALG13, AP1S2, ARHGEF6, ARHGEF9, ARX, ATP6AP2, ATP7A, ATRX, ATXN3L, BCOR, BRWD3, CACNA1F, CASK, CCDC22, CDK6, CDKL5, CLCN4, CLIC2, CNKSR2, CUL4B, DCX, DDX3X, DKC1, DLG3, DP71, EIF2S3, FGD1, FLNA, FMR1, FRMPD4, FTSJ1, GDI1, GK, GPC3, GRIA3, HCCS, HCFC1, HDAC8, HPRT1, HSD17B10, HUWE1, IDS, IGBP1, IKBKG, IL1RAPL1, IQSEC2, KDM5C, KIAA2022, KLF8, KLHL15, L1CAM, LAMP2, LAS1L, MAOA, MBTPS2, MECP2, MED12, MID1, MSL3, NAA10, NDUFA1, NHS, NLGN3, NLGN4X, NSDHL, OCRL, OFD1, OPHN1, OTC, PAK3, PCDH19, PDHA1, PHF6, PHF8, PLP1, PORCN, PQBP1, PRPS1, PTCHD1, RAB39B, RBM10, RLIM, RPS6KA3, SHROOM4, SLC16A2, SLC35A2, SLC6A8, SLC9A6, SMC1A, SMS, SOX3, SYN1, SYP, TAF1, TIMM8A, TMLHE, TSPAN7, UBE2A, UPF3B, USP27X, USP9X, WDR45, ZDHHC15, ZDHHC9, ZMYM3, ZNF711, ZNF81, SMN1, CYP21A2, HFE (uniquement rs1800562), CFTR, FANCB. Pour le gène CFTR seules les variations suivantes ont été recherchées : rs199826652, rs113993960, rs113993959, rs80034486, rs76713772, rs80224560, rs74597325, rs75527207, rs79660178, rs75096551, rs77188391, rs121908745, rs76151804, rs77932196, rs74767530, rs77010898, rs121908761, rs121908799, rs121908747, rs397508266, rs75961395, rs74503330, rs121909011, rs397508393, rs267606722, rs78756941, rs121908744, rs74551128, rs77284892, rs121908769, rs75039782*

Groupe 3) Variants pharmacogénomiques : *CYP2C19 (uniquement rs4986893, rs28399504, rs4244285, rs3892097, rs56337013, rs72552267, rs41291556, rs12248560), CYP2C9 (uniquement rs1799853, rs1057910)*

Seuls les variants pathogènes (classe 5) ou probablement pathogènes (classe 4) des bases de données ClinVar ou HGMD Pro sont rapportés.

Nom, prénom du candidat : Faucher Marie

CONCLUSIONS

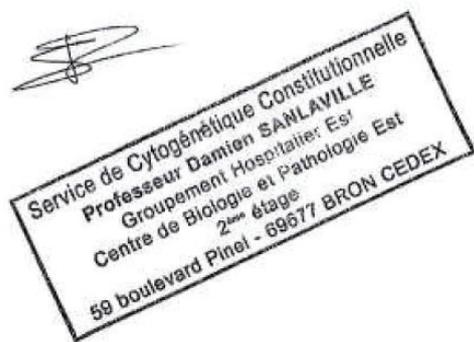
Les anomalies du développement et la déficience intellectuelle (AD/DI) concernent environ 3 % de la population. L'arrivée du séquençage de l'exome a permis d'augmenter le taux diagnostique en mettant en évidence des variants nucléotidiques (SNV) pathogènes dans des gènes impliqués dans ces pathologies. Bien qu'encore délicate, l'analyse de variation du nombre de copie (CVN) est en cours de développement à partir des données d'exome. A côté du diagnostic primaire, le séquençage de l'exome permet de détecter des données dites « secondaires » (DS) pouvant donner lieu à des actions préventives et/ou curatives. Afin d'évaluer les attentes des patients et les modalités organisationnelles de l'analyse des DS détectées par l'exome, le projet FIND regroupant 3 CHU, dont les Hospices Civils de Lyon, a vu le jour.

L'objectif de ce travail est d'analyser les résultats du séquençage de l'exome en trio chez les 110 patients porteurs d'AD/DI des Hospices Civils de Lyon inclus dans le projet FIND avec d'une part une évaluation du taux diagnostique de cette approche et d'autre part une évaluation du nombre de DS et leur impact pour les familles.

Concernant l'analyse des données primaires, le taux diagnostique était de 34.5%. Trente-neuf variants responsables de leurs symptômes ont été détectés chez 38 patients dans les gènes suivants : *ARID1B, BCL11A, CAMK2A, CDK13*2, CYFIP2, DYRK1A, DDX3X, DSP, EFTUD2, GNB1, GNB5, GRM1, GRIN2B, HEXA, HUWE1, KANSL1, KCNQ5, KMT2B, LRIG2, MED13L, MORC2, MTO1, NCKAP1, PACS1, PC, PCGF2, PIGC, PORCN, PPP2R5D, RAC1, RAPSN, SLTRK6, STAG1, TCF4, TRRAP, WDR45*2 et ZMYND11*. Concernant l'analyse des CNV, 6 CNV ont été retenus (analyse sur 84 patients ; résultats provisoires). Tous les CNV avaient été détectés par la CGH-array préalablement. Trois CNV initialement classés comme variant de signification indéterminée ont été reclassés en CNV pathogènes suite à l'analyse récente de la bibliographie. Concernant les DS, 15 variants pathogènes ont été mis en évidence chez 14 patients (12.7% des patients). Ces variants concernaient les gènes actionnables *TNNI3, LDLR, FBN1, PROC* et *SERPINC1*, les gènes de *HFE* (5 patients porteurs de variants p.(Cys282Tyr)) et *CFTR*. De plus, 4 patients étaient porteurs de variants pathogènes dans les gènes de pharmacogénétique *CYP2C9* et *CYP2C19* responsable d'un phénotype de « métaboliseurs lents ».

Le taux diagnostique de l'exome concernant l'analyse des SNV dans notre cohorte correspond au taux décrit dans une récente méta-analyse de la littérature (36%). L'analyse des DS permettra d'adapter la prise en charge des patients et de leur apparentés. L'analyse des CNV est efficace sur les données de l'exome et a permis de retrouver tous les CNV des patients détectés au préalable en CGH-array. Il est très probable qu'à l'avenir le séquençage de l'exome soit l'examen de première intention pour l'exploration des AD/DI puisqu'il présente un taux diagnostique supérieur à celui de la CGH-array (36% contre 12 à 15%) et qu'il permettra à la fois d'analyser les variants monogéniques et les CNV d'un patient. En France, le Plan France Médecine Génomique 2025, récemment mis en place, permettra à l'avenir, lorsque les AD/DI deviendront une préindication, le séquençage du génome entier de ces patients.

Le Président de jury,
Professeur Damien SANLAVILLE
Signature



VU,
Le Doyen de la Faculté de Médecine
et de Maïeutique Lyon-Sud Charles Mérieux

UNIVERSITÉ CLAUDE BERNARD
FACULTÉ DE MÉDECINE ET
DE MAÏEUTIQUE LYON SUD
CHARLES MÉRIEUX

Professeur Carole BURILLON

Vu et permis d'imprimer
Lyon, le 30/09/2019

Le séquençage de l'exome pour le diagnostic des anomalies du développement chez 110 patients dans le cadre du protocole FIND

FAOUCHER Marie ; Thèse Médecine: Lyon 2019 ; n°304

Résumé

Les anomalies du développement et la déficience intellectuelle (AD/DI) concernent environ 3 % de la population. L'arrivée du séquençage de l'exome a permis d'augmenter le taux diagnostique et de mettre en évidence de nouveaux gènes impliqués dans ces pathologies. L'exome permet de faire le diagnostic primaire mais détecte également des données dites « non sollicités » (DNS) donnant lieu à des actions préventives et/ou curatives. De plus, l'analyse de variation du nombre de copie (CNV), actuellement étudié grâce à l'ACPA, est également possible via les données de séquençage de l'exome. L'objectif de cette thèse est d'analyser les résultats du séquençage de l'exome en trio chez 110 patients porteurs d'AD/DI des Hospices Civils de Lyon inclus dans le projet FIND.

Concernant l'analyse des données primaires, 39 variants pathogènes ont été détectés chez 38 patients dans des gènes impliqués dans les AD/DI. Le taux diagnostique est de 34.5%. Concernant les données non sollicitées, 15 variants pathogènes chez 14 patients ont été mis en évidence (12.7% des patients). Ces variants concernaient les gènes actionnables *TNNI3*, *LDLR*, *FBN1*, *PROC* et *SERPINC1*, les gènes de *HFE* et *CFTR* ainsi que dans les gènes de pharmacogénétique *CYP2C9* et *CYP2C19*. 3 CNV premièrement classés comme variant de signification indéterminée lors de la réalisation de l'ACPA ont été reclassés en CNV pathologiques suite à la nouvelle analyse de la bibliographie récente.

Le taux diagnostique de l'exome retrouvé dans notre cohorte correspond au taux décrit dans une récente méta-analyse de la littérature (36%). L'exploration des DNS permet d'adapter la prise en charge des patients via un conseil génétique approprié et une surveillance spécifique. L'analyse des CNV est efficace sur les données de l'exome. La nouvelle analyse des données de l'ACPA est nécessaire avant de poursuivre les explorations génétiques au vu de l'évolution rapide de la littérature scientifique. Il est très probable qu'à l'avenir le séquençage de l'exome soit l'examen de première intention pour l'exploration des AD/DI.

MOTS CLES : Déficience intellectuelle, anomalie du développement, séquençage de l'exome, diagnostic primaire, données non sollicitées, CNV

JURY : Présidente : Monsieur le Professeur Damien SANLAVILLE
Membres : Monsieur le Professeur Vincent DESPORTES DE LA FOSSE
Monsieur le Professeur Patrick EDERY
Madame le Professeur Véronique DAVID
Monsieur le Docteur Gaëtan LESCA

DATE DE SOUTENANCE : 25 octobre 2019

ADRESSE DE L'AUTEUR : 6A1 rue Feuillat, 69003 Lyon
marie.faucher@gmail.com