



HAL
open science

Valeur ajoutée d'une réanalyse de données de séquençage d'un panel de gènes de déficience intellectuelle

Charlotte Felici

► To cite this version:

Charlotte Felici. Valeur ajoutée d'une réanalyse de données de séquençage d'un panel de gènes de déficience intellectuelle. Sciences pharmaceutiques. 2018. hal-03297591

HAL Id: hal-03297591

<https://hal.univ-lorraine.fr/hal-03297591v1>

Submitted on 23 Jul 2021

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

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



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-thesesexercice-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

UNIVERSITE DE LORRAINE

2018

FACULTE DE PHARMACIE

**MEMOIRE
DU DIPLOME D'ETUDES SPECIALISEES
DE BIOLOGIE MEDICALE**

Soutenu devant le jury interrégional

Le 5 octobre 2018

Par Charlotte FELICI,
née le 20 janvier 1990 à Laxou

Conformément aux dispositions de l'arrêté
du 4 octobre 1988, tient lieu de

**THÈSE
POUR LE DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE**

**Valeur ajoutée d'une réanalyse de données de séquençage
d'un panel de gènes de déficience intellectuelle.**

Membres du jury

Président :

M. le Professeur J.L. MERLIN

PU-PH, Faculté de Pharmacie,
Unité de Biologie des Tumeurs, ICL

Juges :

Mme le Docteur AS. DENOMME-PICHON

AHU, CHRU de Nancy, Directeur de thèse

M. le Professeur J.L. MERLIN

PU-PH, Co-directeur de thèse

Mme le Docteur C. BONNET

MCU-PH, CHRU de Nancy

M. le Professeur P. JONVEAUX

PU-PH, Faculté de Médecine de Nancy

M. le Docteur M. MULLER

PH, CHRU de Nancy

**UNIVERSITE DE LORRAINE
FACULTE DE PHARMACIE
Année universitaire 2017-2018**

Doyen

Raphaël DUVAL

Vice-Doyen

Julien PERRIN

Directrice des études

Marie SOCHA

Conseil de la pédagogie

Présidente, Brigitte LEININGER-MULLER

Vice-Présidente, Alexandrine LAMBERT

Collège d'enseignement pharmaceutique hospitalier

Présidente, Béatrice DEMORE

Commission prospective facultaire

Président, Christophe GANTZER

Vice-Président, Jean-Louis MERLIN

Commission de la recherche

Présidente, Caroline GAUCHER

Chargés de mission

Communication

Innovation pédagogique

Référente ADE

Référent dotation sur projet (DSP)

Marie-Paule SAUDER

Alexandrine LAMBERT

Virginie PICHON

Dominique DECOLIN

Responsabilités

Filière officine

Filière industrie

Filière hôpital

Pharma Plus ENSIC

Pharma Plus ENSAIA

Pharma Plus ENSGSI

Cellule de formation continue et individuelle

Commission d'agrément des maîtres de stage

ERASMUS

Caroline PERRIN-SARRADO

Julien GRAVOULET

Isabelle LARTEAUD

Jean-Bernard REGNOUF DE VAINS

Béatrice DEMORE

Marie SOCHAT

Jean-Bernard REGNOUF DE VAINS

Xavier BELLANGER

Igor CLAROT

Béatrice FAIVRE

François DUPUIS

Mihayl VARBANOV

Doyens honoraires

Chantal FINANCE

Francine PAULUS

Claude VIGNERON

Professeurs émérites

Jeffrey ATKINSON

Jean-Claude BLOCK

Max HENRY

Alain MARSURA

Claude VIGNERON

Professeurs honoraires

Pierre DIXNEUF

Chantal FINANCE

Marie-Madeleine GALTEAU

Thérèse GIRARD

Michel JACQUE

Maîtres de conférences honoraires

Monique ALBERT

Mariette BEAUD

François BONNEAUX

Gérald CATAU

Jean-Claude CHEVIN

Pierre LABRUDE
 Vincent LOPPINET
 Alain NICOLAS
 Janine SCHWARTZBROD
 Louis SCHWARTZBROD

Assistants honoraires

Marie-Catherine BERTHE
 Annie PAVIS

Jocelyne COLLOMB
 Bernard DANGIEN
 Marie-Claude FUZELLIER
 Françoise HINZELIN
 Marie-Hélène LIVERTOUX
 Bernard MIGNOT
 Blandine MOREAU
 Dominique NOTTER
 Francine PAULUS
 Christine PERDIAKIS
 Marie-France POCHON
 Anne ROVEL
 Gabriel TROCKLE
 Maria WELLMAN-ROUSSEAU
 Colette ZINUTTI

Enseignants

**Section
 CNU***

Discipline d'enseignement

Professeurs des universités – praticiens hospitaliers

Danièle BENSOUSSAN-LEJZEROWICZ	82	Thérapie cellulaire
Jean-Louis MERLIN	82	Biologie cellulaire
Jean-Michel SIMON	81	Économie de la santé, législation pharmaceutique
Nathalie THILLYU	81	Santé publique et Épidémiologie

Professeurs des universités

Christine CAPDEVILLE-ATKINSON	86	Pharmacologie
Igor CLAROT	85	Chimie analytique
Joël DUCOURNEAU	85	Biophysique, acoustique, audioprothèse
Raphaël DUVAL	87	Microbiologie clinique
Béatrice FAIVRE	87	Hématologie, biologie cellulaire
LUC FERRARI	86	Toxicologie
Pascal FRIANT-MICHEL	85	Mathématiques, physique
Christophe GANTZER	87	Microbiologie
Frédéric JORAND	87	Eau, santé, environnement
Isabelle LARTAUD	86	Pharmacologie
Dominique LAURAIN-MATTAR	86	Pharmacognosie
Brigitte LEININGER-MULLER	87	Biochimie
Pierre LEROY	85	Chimie physique
Philippe MAINCENT	85	Pharmacie galénique
Patrick MENU	86	Physiologie
Jean-Bernard REGNOUF DE VAINS	86	Chimie thérapeutique
Bertrand RIHN	87	Biochimie, biologie moléculaire

Maîtres de conférences des universités – praticiens

Béatrice DEMORE	81	Pharmacie clinique
Alexandre HARLE	82	Biologie cellulaire oncologique
Julien PERRIN	82	Hématologie biologique
Loïc REPEL	82	Biothérapie
Marie SOCHA	81	Pharmacie clinique, thérapeutique et biotechnique

Maîtres de conférences

Sandrine BANAS	87	Parasitologie
Xavier BELLANGER	87	Parasitologie, mycologie médicale
Emmanuelle BENOIT	86	Communication et santé
Isabelle BERTRAND	87	Microbiologie
Michel BOISBRUN	86	Chimie thérapeutique
Ariane BOUDIER	85	Chimie physique
Cédric BOURA	86	Physiologie
Joël COULON	87	Biochimie
Sébastien DADE	85	Bioinformatique
Dominique DECOLIN	85	Chimie analytique
Roudayna DIAB	85	Pharmacie galénique
Natacha DREUMONT	87	Biochimie générale, biochimie clinique
Florence DUMARCAY	86	Chimie thérapeutique
François DUPUIS	86	Pharmacologie
Reine EL OMAR	86	Physiologie
Adil FAIZ	85	Biophysique, acoustique
Anthony GANDIN	87	Mycologie, botanique
Caroline GAUCHER	86	Chimie physique, pharmacologie
Stéphane GIBAUD	86	Pharmacie clinique
Thierry HUMBERT	86	Chimie organique
Olivier JOUBERT	86	Toxicologie, sécurité sanitaire
Alexandrine LAMBERT	85	Informatique, biostatistiques
Julie LEONHARD	86/01	Droit en Santé
Christophe MERLIN	87	Microbiologie environnementale
Maxime MOURER	86	Chimie organique
Coumba NDIAYE	86	Épidémiologie et santé publique
Marianne PARENT	85	Pharmacie galénique
Caroline PERRIN-SARRADO	86	Pharmacologie
Virginie PICHON	85	Biophysique
Sophie PINEL	85	Informatique en santé (e-santé)
Anne SAPIN-MINET	85	Pharmacie galénique
Marie-Paule SAUDER	87	Mycologie, botanique
Guillaume SAUTREY	85	Chimie analytique
Rosella SPINA	86	Pharmacognosie
Sabrina TOUCHET	86	Pharmacochimie
Mihayl VARBANOV	87	Immuno-virologie
Marie-Noëlle VAULTIER	87	Mycologie, botanique
Emilie VELOT	86	Physiologie, physiopathologie humaines
Mohamed ZAIYOU	87	Biochimie et biologie moléculaire

Professeurs associés

Julien GRAVOULET	86	Pharmacie clinique
Anne MAHEUT-BOSSER	86	Sémiologie

Professeur agrégé

Christophe COCHAUD	11	Anglais
--------------------	----	---------

*Disciplines du Conseil National des Universités :

80 : Personnels enseignants et hospitaliers de pharmacie en sciences physico-chimiques et ingénierie appliquée à la santé

81 : Personnels enseignants et hospitaliers de pharmacie en sciences du médicament et des autres produits de santé

82 : Personnels enseignants et hospitaliers de pharmacie en sciences biologiques, fondamentales et cliniques

85 : Personnels enseignants-chercheurs de pharmacie en sciences physico-chimiques et ingénierie appliquée à la santé

86 : Personnels enseignants-chercheurs de pharmacie en sciences du médicament et des autres produits de santé

87 : Personnels enseignants-chercheurs de pharmacie en sciences biologiques, fondamentales et cliniques

11 : Professeur agrégé de lettres et sciences humaines en langues et littératures anglaises et anglo-saxonnes

SERMENT DES APOTHICAIRES



Je jure, en présence des maîtres de la faculté, des conseillers de l'ordre des pharmaciens et de mes condisciples :

D'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.

D'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement.

De ne jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine ; en aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser des actes criminels.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.



« LA FACULTE N'ENTEND DONNER AUCUNE APPROBATION, NI IMPROBATION AUX OPINIONS EMISES DANS LES THESES, CES OPINIONS DOIVENT ETRE CONSIDEREES COMME PROPRES A LEUR AUTEUR ».

Remerciements

À Monsieur le Professeur Jean-Louis MERLIN,

Président et Co-directeur de thèse,

PU-PH, Faculté de Pharmacie de Nancy, Unité de Biologie des Tumeurs, ICL de Nancy

Pour l'honneur que vous me faites en acceptant la présidence et la co-direction de ce jury.

Je vous remercie de l'intérêt et du temps que vous avez bien voulu consacrer à mon travail.

Veuillez trouver ici l'expression de mon respect et de ma sincère reconnaissance.

À Madame Anne-Sophie DENOMME-PICHON,

Directeur de thèse,

AHU, Faculté de Médecine, Laboratoire de Génétique, CHRU de Nancy

Pour la confiance que tu m'as accordée en me dirigeant dans ce travail, pour ton soutien, tes conseils, ton dynamisme et ton efficacité. Un grand merci également pour tout le temps que tu m'as consacré, ta grande disponibilité, pour ta pédagogie et tes connaissances qui m'ont épaulée tout au long de mes stages en génétique. C'est un plaisir de travailler à tes côtés. Sois assurée de ma profonde et sincère reconnaissance.

À Madame Céline BONNET,

MCU-PH, Faculté de Médecine, Laboratoire de Génétique, CHRU de Nancy

Pour l'honneur que vous me faites en acceptant de juger ce travail. Je vous remercie pour votre aide, vos conseils, votre disponibilité et votre gentillesse.

Veuillez recevoir ici l'expression de toute ma gratitude et de mon profond respect.

À Monsieur le Professeur Philippe JONVEAUX,

PU-PH, Faculté de Médecine, Direction de l'Agence de la Biomédecine

Pour le grand honneur que vous me faites en acceptant de juger cette thèse. Je vous remercie pour vos conseils qui m'ont justement aiguillée lors de mon cursus, votre grande disponibilité et vos précieuses connaissances que vous m'avez partagées.

Soyez assuré de mon grand respect et de ma sincère reconnaissance.

À Monsieur Marc MULLER,

PH, Laboratoire de Génétique, CHRU de Nancy

Pour m'avoir fait l'honneur d'accepter de juger cette thèse. Je te remercie sincèrement de l'intérêt que tu as bien voulu porter à mon travail. Sois assuré de ma profonde reconnaissance.

À l'ensemble des laboratoires de génétique de Nancy, de Metz et de Strasbourg, et tout particulièrement à Aurélie, Myriam, Virginie, Laetitia, Lucille, Michèle, Christine, Inès, Rémy et Sylvie,

Un grand merci pour votre gentillesse et le temps que vous avez consacré à ma formation.

Merci à Christine pour votre gentillesse et votre compréhension lors de la réalisation de ce travail et pour tout ce que vous m'avez appris durant ce dernier stage.

À Papa et Maman

Pour votre amour, votre soutien et vos encouragements,
Pour m'avoir guidée et soutenue dans mes choix,
Pour m'avoir permis de mener ces études dans les meilleures conditions qui soient,
Si je suis là aujourd'hui c'est en immense partie grâce à vous
Cette réussite est aussi la vôtre. J'espère vous rendre fiers. Avec tout mon amour.

À Pauline et Juliette

Pour m'avoir écoutée, supportée, soutenue dans mes nombreux moments d'euphorie, de stress ou de doute tout au long de ces années. Merci pour tous les bons moments de complicité que nous avons partagés et pour ceux restants à venir.

À Mamie

Pour être si présente dans ma vie,
Je te dois une grande partie de ma réussite et je ne te remercierai jamais assez pour tout ce que tu m'as appris.
Merci pour ton soutien et tes attentions qui comptent beaucoup pour moi.
Je t'admire énormément.

À Mémère

Merci pour tous ces bons moments partagés gravés dans mon cœur, pour tous tes enseignements.

À Papi et à Pépère

J'espère que vous auriez été fiers de votre petite fille aujourd'hui. Je vous dédie mon travail.

À Claire, Charlotte, Hélène, Marion et Shanti

Pour tous nos rires, nos soirées et tous ces beaux moments que l'on a vécus ensemble durant ces dernières années et pour tous ceux qui restent à venir ! Merci à Claire pour ta gentillesse et ta présence, à Charlotte pour ton sourire et ton rire communicatif, à Hélène pour ton éternelle bonne humeur et joie de vivre, à Marion pour ton optimisme à toute épreuve et à Shanti pour ta douceur et ta bienveillance. Merci pour vos conseils avisés et pour être présentes quoi qu'il arrive ! Je vous aime !!!

À Juliette et Arnaud

Merci pour votre amitié infaillible, pour tous les souvenirs et les bons moments passés ensemble depuis le lycée ! Vous comptez beaucoup pour moi et j'espère continuer à vous voir très longtemps.

À Cerise

Bientôt 28 ans qu'on se connaît... et toujours aussi complices ! Tellement de souvenirs avec toi, quand on soignait Coco au Synthol, « Ziitroooone », « Eehh ooh ! », « Excusez-moi nous sommes perdues », les descentes de la rue des Sœurs Macarons avec le verglas, les heures passées « au parc », les stratégies pour dormir l'une chez l'autre (ou pour adopter un chien en cachette), nos « mots » de 5 pages pour raconter juste une après-midi au collège... je peux continuer encore longtemps ! Merci pour cette belle amitié qui perdurera encore longtemps j'en suis certaine.

À Matthieu et Antoine

Merci me m'avoir toujours accueillie chez vous les bras ouverts après mes journées éreintantes (Promis la prochaine fois je serai moins fatiguée !). Merci aussi pour tous ces supers moments partagés avec vous, pour toutes ces soirées parisiennes et ces délicieux repas ! J'espère que notre amitié durera encore de longues années.

À Delphine et Claire

Merci pour votre présence bienveillante et apaisante, votre sagesse, et tous ces moments à discuter de choses amusantes ou non (je pense bien sûr aux dysonites aiguës qui font des ravages).

À mes co-internes Clément, Manon, Charlotte,

Merci pour ces quatre belles années, pour avoir rendu mes stages plus sympathiques et pour avoir partagé ces innombrables pauses café avec moi !

À Justine, Mathilde et Prisca

Pour ces bons souvenirs des années de fac, ces soirées, ces vacances et ces week-ends. J'espère que nous arriverons à nous revoir régulièrement.

À Julien

Pour ton amour, ta patience, ton soutien et ta présence dans les bons comme les mauvais moments de ma vie,

Pour savoir me faire rire et me réconforter,

Merci pour ce que tu es. Je t'aime.

TABLE DES MATIÈRES

TABLE DES MATIÈRES.....	1
LISTE DES ABRÉVIATIONS.....	5
LISTE DES TABLEAUX.....	8
LISTE DES FIGURES.....	9
I INTRODUCTION.....	11
II MATÉRIELS ET MÉTHODES	16
II.1 Individus analysés	16
II.2 Gènes ciblés	17
II.3 <i>Wet lab</i> ou génération des données de séquençage	17
II.3.1 Prétraitement de l'échantillon	17
II.3.2 Préparation de la librairie	18
II.3.2.1 Fragmentation, sélection de taille et purification de l'ADN génomique.....	18
II.3.2.2 Ligation des adaptateurs et des barcodes	20
II.3.2.3 Amplification de la librairie barcodée	22
II.3.2.4 Concentration de la librairie par lyophilisation	22
II.3.2.5 Capture	22
II.3.2.6 Amplification post-capture	23
II.3.2.7 Multiplexage des échantillons.....	24
II.3.3 Séquençage.....	25
II.3.3.1 Préparation du « template » ou matrice de séquençage.....	25
II.3.3.2 Dépôt sur puce et séquençage.....	28
II.4 Pools des parents	29
II.5 <i>Dry lab</i> ou analyse informatique des données de séquençage	30
II.5.1 <i>Base calling</i> ou détermination de la séquence.....	30
II.5.2 Contrôles de qualité primaires	31
II.5.3 Alignement ou <i>mapping</i>	31
II.5.4 Contrôles de qualité secondaires	31
II.5.5 <i>Variant calling</i> ou identification des variants	32
II.5.6 Annotation des variants.....	32

II.5.6.1	Recommandations de l'ACMG	33
II.5.6.2	Bases de données de variants spécifiques de populations.....	33
II.5.6.2.1	dbSNP	33
II.5.6.2.2	ExAC	34
II.5.6.2.3	gnomAD.....	34
II.5.6.3	ClinVar	35
II.5.6.4	OMIM	35
II.5.6.5	Mutalyzer	36
II.5.6.6	VarSome.....	36
II.5.6.7	Outils de prédiction in silico	36
II.5.6.7.1	PolyPhen-2	36
II.5.6.7.2	SIFT	37
II.5.6.8	VEP.....	37
II.5.6.9	Outil de visualisation des variants : IGV	40
II.6	Détection des CNV.....	43
II.7	Analyses bioinformatiques et biologiques du panel de gènes.....	43
II.7.1	Première analyse	43
II.7.1.1	Pipeline Ion Torrent.....	43
II.7.1.2	Pipeline maison.....	44
II.7.2	Réinterprétation des résultats non contributifs.....	45
II.7.2.1	Visualisation des variants.....	46
II.7.2.2	Collecte manuelle de nouvelles informations d'annotations.....	46
II.7.2.3	Priorisation des variants	46
II.7.2.3.1	Confrontation de la ségrégation du variant au type de transmission de la pathologie associée au gène porteur du variant	46
II.7.2.3.2	Confrontation du type de transmission du gène porteur du variant et de la fréquence du variant.....	47
II.7.2.3.3	Consultation des autres critères.....	47
II.7.2.3.1	Partage de données internationales.....	48
II.7.2.3.2	Classification des variants candidats.....	48

II.8	Validation des variants	48
II.9	Projet INDEXES : exploration par séquençage exomique des causes moléculaires de déficiences intellectuelles sévère	49
II.9.1	Présentation générale	49
II.9.2	Inclusions	51
II.9.3	Séquençage et analyse des données.....	52
III	RÉSULTATS	53
III.1	Réanalyse	53
III.2	Bilan des inclusions dans le projet INDEXES	59
III.2.1	Individus proposés par les cliniciens	59
III.2.2	Individus à présenter au comité scientifique	59
III.2.3	Individus à proposer pour inclusion en fonction des résultats du panel	59
III.2.4	Individus à proposer pour inclusion en fonction du phénotype.....	61
IV	DISCUSSION ET PERSPECTIVES	64
IV.1	Discussion.....	64
IV.1.1	Évolution rapide des connaissances	64
IV.1.2	Difficultés et erreurs d'interprétation	64
IV.1.3	Utilisation du partage de données international	65
IV.1.4	Utilisation du même pipeline.....	65
IV.1.5	Limites des paramètres utilisés	66
IV.1.5.1.1	Filtres appliqués	66
IV.1.5.1.2	Pool des parents.....	66
IV.1.5.1.3	Inactivation de l'X	67
IV.1.5.1.4	Pathologies à transmission récessive liée à l'X.....	67
IV.1.6	Mosaïque	67
IV.1.7	La réanalyse des panels dans la littérature	67
IV.1.8	Délai entre les deux analyses.....	68
IV.1.9	Importance des données phénotypiques.....	69
IV.2	Perspectives : bénéfice diagnostic du projet INDEXES	72

V CONCLUSION	73
BIBLIOGRAPHIE.....	74
ANNEXES.....	82
RÉSUMÉ.....	88

LISTE DES ABRÉVIATIONS

3D : 3 dimensions / tridimensionnel
ACMG : American College of Medical Genetics and Genomics
AD : Autosomique Dominant
ADN : Acide DésoxyriboNucléique
ADNc : Acide DésoxyriboNucléique complémentaire
AMPc : Adénosine MonoPhosphate cyclique
AR : Autosomique Récessif
BAM : Binary sAM
BWA : Burrows-Wheeler Aligner
CADD : Combined Annotation Dependent Depletion
CDS : Coding DNA Sequence
CGH : Comparative Genomic Hybridization
Chr. : Chromosome
CHRU : Centre Hospitalier Régional Universitaire
CMOS : Complementary Metal-Oxyde Semiconductor
CNV : Copy Number Variant
CPRC : Contrat de Programme de Recherche Clinique
dbSNP : Single Nucleotide Polymorphism database
D : Délétère
Dfc : Délétère Faible Confiance
DI : Déficience Intellectuelle
dNTP : désoxyriboNucléotides TriPhosphate
DS : Déviation Standard
emPCR : emulsion Polymerase Chain Reaction
ESNG : Ensembl Gene ID
ESNT : Ensembl Transcript ID
ExAC : Exome Aggregation Consortium
FATHMM : Functional Analysis Through Hidden Markov Models
FISH : Fluorescent In Situ Hybridization
GATK : Genome Analysis ToolKit
GERP : Genomic Evolutionary Rate Profiling
gnomAD : genome Aggregation Database
GRCh37 : Genome Reference Consortium human 37
HGVS : Human Genome Variation Society

HTA : HyperTension Artérielle
IGV : Integrative Genomics Viewer
IMC : Indice de Masse Corporelle
INDEXES : INtellectual Disability EXploration by Exome Sequencing
IRM : Imagerie par Résonance Magnétique
ISFET : Ion-Sensitive Field Effect Transistor
LRT : Likelihood Ratio Test
M-CAP Score : Mendelian Clinically Applicable Pathogenicity Score
NCBI : National Center for Biotechnology Information
N : Neutre
NGS : Next Generation Sequencing
OMIM : Online Mendelian Inheritance in Man
ORL : Oto-Rhino-Laryngologique
Pb : Paires de Bases
PD : Probablement Délétère
PEG : PolyÉthylène Glycol
PCR : Polymerase Chain Reaction
PhastCons : phylogenetic Analysis with Space/Time models Conservation score
PM2 : Pathogenic Moderate 2
POCS : Pointes Ondes Continues du Sommeil
PP3 : Pathogenic Supporting 3
PP4 : Pathogenic Supporting 4
PROVEAN : PROtein Variation Effect ANalyzer
PS2 : Pathogenic Strong 2
QF-PCR : Quantitative Fluorescence-Polymerase Chain Reaction
REGOVAR : RÉseau Grand Ouest pour l'interprétation des VAriants génétiques Rares
RCIU : Retard de Croissance Intra Utérin
SAM : Sequence Alignment / Map
SIFT : Sorting Intolerant From Tolerant
SO : Sequence Ontology
SNP : Single Nucleotide Polymorphism
SPRI : Solid Phase Reversible Immobilization
TED : Troubles Envahissants du Développement
TSA : Troubles du Spectre Autistique
UCSC : University of California, Santa Cruz
UTR : Untranslated Transcribed Region
VCF : Variant Calling Format

VEP : Variant Effect Predictor Significance

VUS (ou VOUS) : Variant of Unknown (variant de signification inconnue)

WES : Whole Exome Sequencing (séquençage d'exome complet)

XHMM : eXome Hidden Markov Model

LISTE DES TABLEAUX

TABLEAU I : Correspondance de la valeur du score Q avec la probabilité d'une identification de base exacte.....	30
TABLEAU II : Correspondances entre les conséquences (SO term), leur définition selon Sequence Ontology (SO description), leur référence dans SequenceOntology (SO accession), le terme affiché (Display term) et leur IMPACT (d'après Ensembl).	39
TABLEAU III : Présentation des VUS mis en évidence lors de la réanalyse.	55
TABLEAU IV : Prédiction des différents outils de prédiction <i>in silico</i> consultés via VEP pour le variant présent dans TUBB2B.	58
TABLEAU V : Individus candidats pour le projet INDEXES.....	63
ANNEXE 1 : Bases de données et outils principalement utilisés pour l'alignement, l'identification, l'annotation et la classification des variants.	82
ANNEXE 2 : Liste des 446 analysés, version n°2 du panel.....	83
ANNEXE 3 : Récapitulatif des critères de l'ACMG permettant de classer les variants en 5 catégories : bénin, probablement bénin, VUS, probablement pathogène et pathogène (d'après Richards <i>et al.</i> , 2015).....	84
ANNEXE 4 : Récapitulatif du poids de chaque critère de classification des variants bénins et de leur précautions d'interprétation (selon Richards <i>et al.</i> , 2015).	85
ANNEXE 5 : Récapitulatif du poids de chaque critère de classification des variants pathogènes et de leurs précautions d'interprétation (selon Richards <i>et al.</i> , 2015).	86
ANNEXE 6 : Règles de combinaison des différents critères permettant de classer un variant de séquence en variant bénin (classe 1), variant probablement bénin (classe 2), VUS (classe 3), variant probablement pathogène (classe 4) ou variant pathogène (classe 5) (selon Richards <i>et al.</i> , 2015).....	87

LISTE DES FIGURES

FIGURE 1 : Augmentation du nombre de gènes connus comme étant impliqués dans la DI au cours du temps (à gauche) et rendement diagnostique des DI au cours du temps (modifiée d'après Vissers, Gilissen et Veltman 2016).....	14
FIGURE 2 : Principales étapes de l'analyse du panel de 446 gènes de DI en NGS. ... Erreur ! Signet non défini.	
FIGURE 3 : Taille des fragments d'ADN élués (verticalement) selon le ratio du volume de billes SPRI sur le volume d'ADN (horizontalement) (d'après Broad Institute).	19
FIGURE 4: Aperçu du processus de sélection de taille puis de purification de l'échantillon (d'après Beckman Coulter).....	19
FIGURE 5 : Exemple de profil électrophorétique après fragmentation, sur puce DNA Chip High Sensitivity (Agilent), montrant les unités de fluorescence (FU) en fonction de la taille (en pb).	20
FIGURE 6 : Schéma des étapes de A-tailing et de ligation des adaptateurs aux extrémités 3'A de chaque fragment d'ADN.	21
FIGURE 7 : Exemple de profil électrophorétique sur puce DNA 1000 après amplification de la librairie montrant les unités de fluorescence (FU) en fonction de la taille (en pb).....	22
FIGURE 8 : Schéma du processus de capture par le kit SureSelect Target Enrichment (d'après Agilent Technologies).	23
FIGURE 9 : Exemple de profil électrophorétique sur puce High Sensitivity DNA après capture et amplification montrant les unités de fluorescence (FU) en fonction de la taille (en pb).	24
FIGURE 10 : Exemple de profil électrophorétique de la librairie finalisée sur puce High Sensitivity DNA.....	25
FIGURE 11 : Schéma des différentes configurations pouvant être rencontrées lors de l'émulsion des gouttelettes d'eau en phase huileuse lors d'une PCR en émulsion.	26
FIGURE 12 : Schéma d'une microgouttelette, illustrant le principe de la PCR en émulsion.	27
FIGURE 13 : Schéma du processus de séquençage Ion Torrent (d'après Goodwin, McPherson, et McCombie 2016)	29
FIGURE 14 : Illustration des notions de profondeur et de couverture de séquençage (modifié d'après Lacoste <i>et al.</i> 2017).	32
FIGURE 15 : Diagramme montrant chaque « conséquence » par rapport à la structure génomique (d'après Ensembl).....	38
FIGURE 16 : Exemple de profil d'un variant faux positif (chr7:g.74146817C>T) visualisé sur IGV avec un mauvais alignement des reads.....	41

FIGURE 17 : Exemple de profil d'un variant faux positif (chrX:g.148582573GAA>G) visualisé sur IGV avec un homopolymère de A.....	42
FIGURE 18 : Exemple de profil d'un variant faux positif (chrX:g.110439122TA>T) visualisé sur IGV avec un biais de brin.....	43
FIGURE 19 : Différentes stratégies utilisées pour identifier des gènes responsables de maladies rares par approche exomique (selon Gilissen <i>et al.</i> 2012).....	50
FIGURE 20 : Diagramme de flux des résultats.....	53
FIGURE 21 : Répartition des variants rapportés sur <i>TUBB2B</i> (modifié d'après Geiger <i>et al.</i> 2017).....	57
FIGURE 22 : Représentation des variants associés au syndrome de Pitt-Hopkins.....	61
FIGURE 23 : Visualisation du variant dans <i>CREBBP</i> avec IGV.....	71

I INTRODUCTION

La déficience intellectuelle (DI) est définie comme un trouble neuro-développemental apparaissant dans l'enfance et caractérisé par des déficits du fonctionnement intellectuel et du comportement adaptatif dans les domaines conceptuel, social et pratique. Selon l'American Psychiatric Association (American Psychiatric Association 2013), trois critères sont nécessaires pour diagnostiquer une déficience intellectuelle :

1. les déficits de fonctionnement intellectuel tels que de raisonnement, de résolution de problèmes, de planification, de pensée abstraite, de jugement, d'apprentissage scolaire et d'apprentissage par l'expérience ;
2. les déficits de fonctionnement adaptatif provoquant des difficultés pour l'individu à respecter les normes socioculturelles et de développement et à assumer ses responsabilités sociales ;
3. l'apparition de ces troubles pendant l'enfance.

La prévalence mondiale de la DI est proche de 1 % (Maulik *et al.* 2011). Les hommes sont plus touchés avec un ratio homme/femme qui se situe entre 1,3 et 1,9 (Leonard et Wen 2002). Ce déséquilibre peut s'expliquer par les formes liées à des anomalies sur le chromosome X, qui contient une densité importante de gènes impliqués dans les fonctions cognitives (Chelly *et al.* 2006).

Plusieurs degrés de gravité sont distingués dans la DI, qui peut être qualifiée de légère, modérée, sévère ou profonde. Chaque classe affecte respectivement 85 %, 10 %, 3,5 % et 1,5 % des individus atteints (Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental Disorders; Board on the Health of Select Populations; Institute of Medicine; The National Academies of Sciences, Engineering, et Medicine 2015). Elles peuvent également être classées en DI non syndromiques et DI syndromiques. Ces dernières sont associées à des caractéristiques cliniques ou comorbidités très variables telles que des malformations d'organes, des troubles de croissance, des particularités morphologiques ou des troubles neuro-développementaux non spécifiques comme l'épilepsie, la schizophrénie ou l'autisme (Chelly *et al.* 2006). Les DI non syndromiques sont isolées.

Les étiologies responsables des DI sont très hétérogènes et se composent de facteurs génétiques et environnementaux. Ces derniers sont des facteurs exogènes qui influent sur le développement et le fonctionnement du système nerveux central. Ils interviennent dans la période prénatale (exposition à des substances toxiques comme l'alcool, infection maternelle comme la toxoplasmose ou la rubéole), périnatale (traumatisme à la naissance, hypoxie,

infection materno-fœtales...), ou postnatale (infections, intoxications, traumatisme cérébral...) (Maulik *et al.* 2011). Dans l'étude de Rauch *et al.* réalisée en 2006, 59,3 % des étiologies des DI sont inconnues et la contribution des causes exogènes est estimée à seulement 1,3 %. Les 39,4 % restants représentent des causes génétiques, qu'elles soient chromosomiques, subchromosomiques ou moléculaires (Rauch *et al.* 2006). Depuis 2009, le séquençage de nouvelle génération, également appelé séquençage haut-débit ou *Next Generation Sequencing* (NGS) a permis d'augmenter considérablement le rendement diagnostique en identifiant de nombreux gènes et variations responsables de DI (Figure 1) (Ng *et al.* 2010; Ng *et al.* 2009; Vissers, Gilissen, et Veltman 2016). Ces techniques ont permis de constater qu'une grande proportion de DI est due à des variations rares monogéniques (Ellison, Rosenfeld, et Shaffer 2013). Un très grand nombre de gènes est impliqué (Sian Ellard 2014) et chaque gène de DI pris individuellement ne représente qu'une très petite proportion des étiologies, estimée pour chacune à moins de 1 % (Ellison, Rosenfeld, et Shaffer 2013; McClellan et King 2010a, 2010b). Actuellement, on connaît plus de 1500 gènes impliqués dans les DI (Lelieveld *et al.* 2016). La grande majorité des DI est d'apparition sporadique et s'explique le plus souvent par des variations *de novo* (c'est-à-dire survenues au cours de la méiose et absentes chez les parents) affectant des gènes codant des protéines et impliqués dans le développement (de Ligt *et al.* 2012; Gilissen *et al.* 2014; Veltman et Brunner 2012; Vissers *et al.* 2010).

Aujourd'hui, malgré cette grande avancée, près de 30 % des DI restent sans diagnostic moléculaire (Vissers, Gilissen, et Veltman 2016). Cette proportion s'explique par l'hétérogénéité génétique et phénotypique de la DI. Identifier les causes moléculaires des DI est indispensable (Kaufman, Ayub, et Vincent 2010; Rauch *et al.* 2006; Vissers *et al.* 2010), d'une part en recherche pour mieux comprendre les mécanismes moléculaires impliqués dans la DI et d'autre part en clinique pour améliorer la prise en charge de l'individu atteint, orienter le conseil génétique et identifier de potentielles cibles thérapeutiques.

Depuis 2005, le séquençage haut-débit évolue en devenant plus abordable (Metzker 2010). Son arrivée dans les laboratoires d'analyses médicales a conduit à réorganiser les stratégies de diagnostic moléculaire sur le territoire national. Les investigations habituelles en génétique devant une DI sans orientation diagnostique sont les suivantes :

- 1) Recherche de la maladie de l'X fragile par biologie moléculaire car cela reste la cause la plus fréquente de DI héréditaire avec une incidence estimée à 1 sur 4000 hommes et à 1 sur 5000 à 8000 femmes (Hill *et al.* 2010) ;
- 2) Analyse chromosomique par puces à ADN (par CGH array ou SNP array) ;

- 3) NGS, dont deux approches sont actuellement implantées en laboratoire de diagnostic en France dans le cadre de la DI : l'approche par panel de gènes ciblés ou celle par séquençage de l'exome.

Le laboratoire de génétique du CHRU de Nancy a mis en place le séquençage d'un panel de plus de 400 gènes impliqués dans la déficience intellectuelle.

L'analyse par panel de gènes en NGS est une stratégie dans laquelle un ensemble défini de gènes liés au phénotype de l'individu est analysé. Cette méthode consiste à enrichir des régions d'intérêt appartenant à des gènes qui sont connus ou soupçonnés comme étant impliqués dans la pathologie. Le débit important de cette technologie permet l'analyse simultanée de plusieurs gènes chez plusieurs individus et la rend utilisable pour des maladies présentant une forte hétérogénéité génétique (Collège National des Enseignants et Praticiens de Génétique Médicale 2016).

La sélection de gènes appropriés dans un panel est difficile et nécessite un partenariat étroit entre les cliniciens et les généticiens de laboratoire. La plupart du temps, les panels sont constitués de gènes validés comme étant en cause dans la pathologie étudiée, c'est-à-dire déjà décrits dans la littérature comme responsables de DI. Leur validation nécessite, pour la majorité, des analyses fonctionnelles et/ou des répliques de cas apportant des preuves de leur pathogénicité. La connaissance des mécanismes moléculaires et physiopathologiques facilite l'interprétation des résultats pour ces gènes. Certains gènes associés à des troubles ayant un phénotype chevauchant avec ceux de la DI peuvent être inclus dans le but d'un diagnostic différentiel (Xue *et al.* 2015).

Le nombre de nouveaux gènes découverts comme étant impliqués dans la DI augmente très rapidement et de manière continue au cours de ces dernières années (Figure 1). C'est pourquoi, il est fréquent d'inclure dans les panels des gènes dits candidats ou gènes de recherche, c'est-à-dire des gènes suspectés d'être impliqués dans la DI, sans validation fonctionnelle ou sans réplique de cas. La vocation de ces derniers se situe dans le versant de la recherche car leur identification n'est souvent basée que sur quelques familles ou cas isolés. L'analyse porte sur les séquences exoniques et sur les quelques dizaines de paires de bases des introns flanquant les exons pour couvrir les jonctions introns-exons contenant les sites consensus d'épissage de l'ARN. Il est possible d'ajouter des régions géniques supplémentaires telles que des promoteurs ou des introns d'intérêt clinique (Collège National des Enseignants et Praticiens de Génétique Médicale 2016). Les panels peuvent être « fait maison » par le laboratoire lui-même ou peuvent, dans le cas de certaines pathologies

particulières – en général, celles pour lesquelles un petit nombre de gènes est ciblé – être fournis « prêt à l'emploi » par certains fournisseurs.

Un panel peut compter de moins d'une dizaine à plusieurs centaines de gènes en fonction de la pathologie ciblée et du laboratoire proposant le test. Il existe par exemple un panel ciblant 44 gènes de DI proposés par les CHU de Saint-Étienne ou de Bordeaux et des panels beaucoup plus conséquents comme celui du CHU de Brest contenant plus de 500 gènes de DI (Orphanet). Certains laboratoires préfèrent n'inclure que les gènes fortement associés à la pathologie tandis que d'autres privilégient le rendement diagnostique en incluant le plus de gènes possibles. L'interprétation de ces derniers peut s'avérer difficile et générer de nombreux variants de signification indéterminée (VUS), pour lesquels des recherches plus approfondies sont nécessaires, lesquelles peuvent montrer que ces variants ne sont finalement pas mis en cause dans la symptomatologie de l'individu testé (Xue *et al.* 2015).

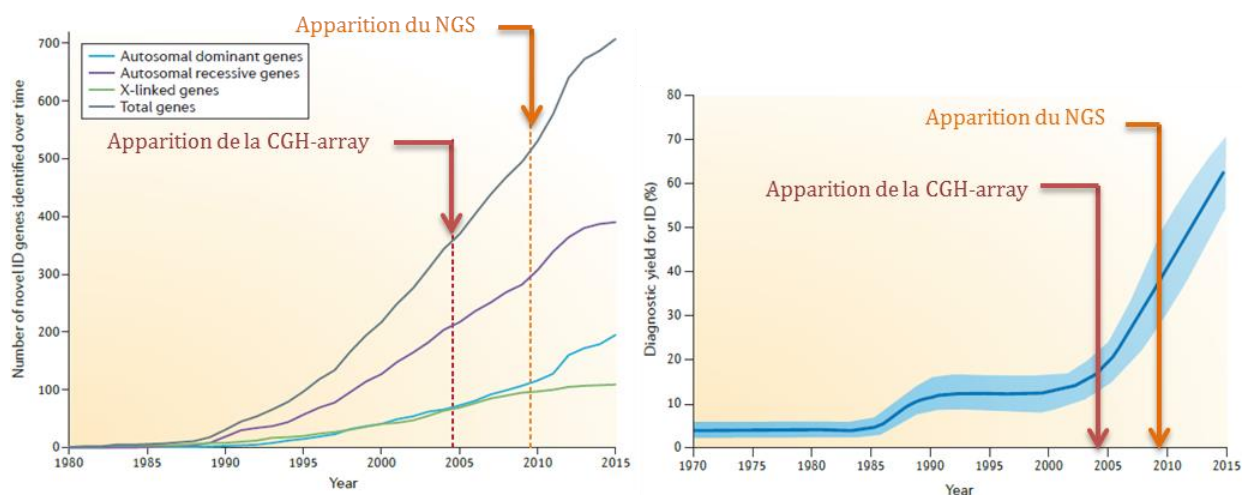


FIGURE 1 : Augmentation du nombre de gènes connus comme étant impliqués dans la DI au cours du temps (à gauche) et rendement diagnostique des DI au cours du temps (modifiée d'après Vissers, Gilissen et Veltman 2016).

Ce travail a pour but d'évaluer l'intérêt d'une réanalyse à un an environ de résultats jusqu'alors non contributifs d'un panel de 446 gènes impliqués dans la déficience intellectuelle effectuée au laboratoire de génétique du CHRU de Nancy. Depuis la mise en place du panel au laboratoire, 66 individus atteints de DI ont bénéficié d'un séquençage de panel de gènes entre fin 2015 et 2017. Après une première interprétation, aucun variant candidat n'a été retenu pour 44 individus. Le but de ce travail est :

1) d'évaluer la valeur ajoutée d'une réanalyse de données de séquençage d'un panel de gènes de déficience intellectuelle ;

- 2) d'essayer de résoudre de nouveaux cas à la lumière des données actuelles de la littérature ;
- 3) d'évaluer l'indication d'une exploration par séquençage exomique par une inclusion dans le projet INDEXES (INtellectual Disability EXploration by Exome Sequencing). Ce dernier vise à explorer, par séquençage exomique en trios (un individu et ses deux parents), dix sujets atteints de DI sévère du CHRU de Nancy grâce à un financement obtenu après dépôt d'un appel d'offre au CPRC.

II MATÉRIELS ET MÉTHODES

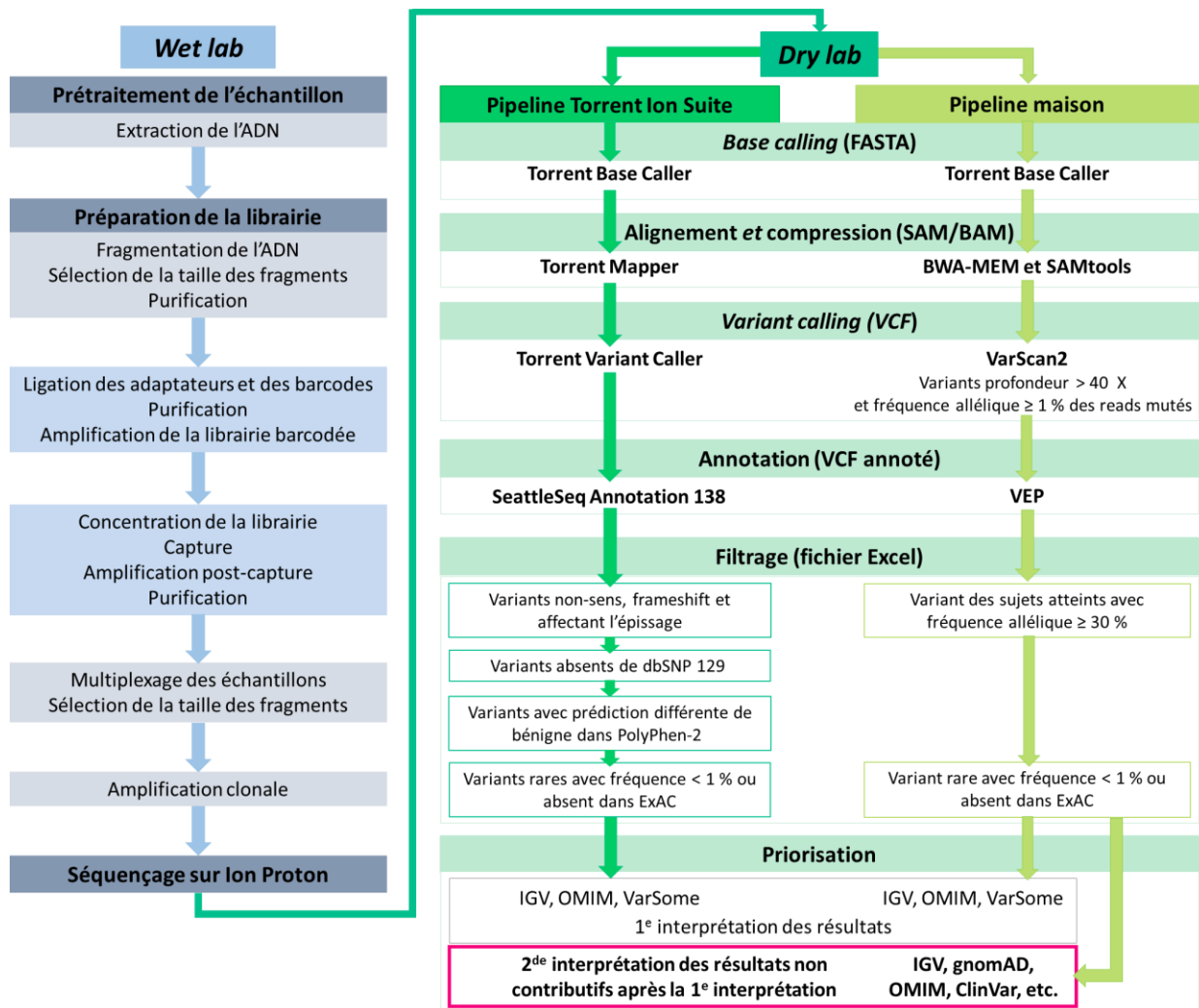


FIGURE 2 : Principales étapes de l'analyse du panel de 446 gènes de DI en NGS.

Le cadre bleu représente la partie *wet lab* qui permet de générer les données de séquençage (étapes à la paillasse). Le *dry lab* permet l'analyse bioinformatique des données pour l'identification des variants. Les parties vertes schématisent les deux analyses indépendantes par deux pipelines différents : le pipeline Ion Torrent Suite (à gauche) et le pipeline maison (à droite). L'encadré gris délimite la première interprétation des résultats. L'encadré rose représente ce travail qui consiste en la réanalyse des résultats obtenus par le pipeline maison lors d'une seconde interprétation à un an.

II.1 Individus analysés

Les 66 individus analysés présentent une DI sans diagnostic étiologique. Ils sont vus en consultation de génétique médicale au CHRU de Nancy où un séquençage d'un panel de

gènes DI leur est prescrit après recueil de leur consentement écrit et/ou celui de leurs parents ou représentants légaux. Un prélèvement sanguin est effectué et adressé au laboratoire de génétique avec le consentement.

Après une première interprétation, le résultat est non contributif pour 44 individus. Leurs données de séquençage sont analysées une seconde fois, environ un an plus tard. Un 45^e individu est choisi au hasard parmi les 22 individus pour lesquels la première analyse a mis en évidence une étiologie moléculaire et est réanalysé en tant que de témoin positif.

II.2 Gènes ciblés

Le panel de gènes utilisé permet d'analyser par séquençage haut débit plus de 400 gènes d'intérêt dans la DI. Il a initialement été mis en place par le laboratoire de génétique moléculaire du CHRU de Strasbourg en 2014. Deux versions du panel sont utilisées. La première version, contenant 439 gènes, est utilisée pour la première série de 16 échantillons. Les quatre séries suivantes utilisent la seconde version qui contient sept gènes supplémentaires, soit un total de 446 gènes. La taille de la région ciblée est de 2 mégabases.

Ce panel se divise en deux sous-ensembles de gènes de DI : un panel principal de diagnostic avec des gènes d'implication prouvée dans la DI et un panel secondaire contenant des gènes candidats. Une dernière catégorie correspond à des gènes impliqués dans l'épilepsie. L'épilepsie est un trouble fréquemment associé aux DI. Sa prévalence chez les individus atteints de DI est plus élevée que dans la population générale et augmente avec la sévérité de la DI (Robertson *et al.* 2015). La liste des gènes du panel est disponible en Annexe 2.

II.3 Wet lab ou génération des données de séquençage

II.3.1 Prétraitement de l'échantillon

L'ADN génomique est extrait à partir d'un prélèvement sanguin périphérique sur tube EDTA. Cette étape s'effectue à l'aide de l'automate QIAcube (QIAGEN) via le kit QIAmp DNA Blood mini, qui, après une lyse cellulaire (en présence d'un tampon de lyse et de protéase), utilise l'adsorption spécifique de l'ADN sur une membrane à base de silice selon des conditions précises de pH et de concentration en certains sels. Les protéines et autres contaminants sont éliminés par différentes étapes de lavages. L'ADN est élué, qualifié et quantifié. Le spectrophotomètre Nanodrop ND-1000 (ThermoFisher Scientific) exploite l'absorbance des acides nucléiques pour le doser et permettre de détecter la présence d'éventuels contaminants. Un dosage plus précis est effectué par utilisation de marqueurs fluorescents

spécifiques de l'ADN avec le fluorimètre Qubit 2.0 (ThermoFisher Scientific, kit dsQUBIT DNA BR), plus sensible et plus spécifique. Une migration électrophorétique sur gel d'agarose permet de vérifier que l'ADN n'est pas dégradé.

II.3.2 Préparation de la librairie

II.3.2.1 *Fragmentation, sélection de taille et purification de l'ADN génomique*

Fragmentation

La technologie Ion Shear (kit Ion Xpress Plus Fragment Library, Life Technologies) casse l'ADN en petits fragments d'environ 130 pb, taille adaptée pour la suite de la technique. Cette technologie utilise une enzyme de restriction qui coupe l'ADN au niveau d'une séquence caractéristique et le morcelle.

Sélection et purification

Des billes Agencourt AMPure XP (Beckman Coulter Genomics), également appelées billes SPRI (Solid Phase Reversible Immobilization), sont utilisées pour sélectionner les fragments selon leur taille et les purifier. Ces billes sont composées de polystyrène et de magnétite et sont recouvertes de groupements carboxyles. Elles sont paramagnétiques, c'est-à-dire aimantées uniquement dans un champ magnétique, ce qui les empêche de s'agglomérer dans la solution de polyéthylène glycol (PEG) et de sel (ions sodium) dans laquelle elles sont en suspension. Le PEG provoque l'agrégation de l'ADN, se trouvant sous forme dissoute dans une solution aqueuse, et les ions sodium protègent les charges négatives du squelette phosphaté de cet ADN. L'ADN se lie avec les groupes carboxyles chargés négativement. Un aimant est placé à l'extérieur du tube pour créer un fort champ magnétique externe. Les billes liées à l'ADN sont attirées par le champ et se « collent » à la paroi du tube. L'élution s'effectue à l'aide d'eau ou d'un tampon d'élution (solution aqueuse) tout en maintenant le champ magnétique.

Il est possible de sélectionner la taille des fragments retenus à l'aide de la solution de billes SPRI. En effet, plus les fragments d'ADN sont grands, plus leur charge est importante, ce qui favorise leur interaction avec les billes par rapport aux fragments de plus petite taille. Ces derniers restent dans le surnageant et sont soit éliminés, soit conservés, selon la taille que l'on souhaite récupérer. Plus le ratio du volume de billes SPRI sur celui d'ADN est petit, plus les fragments d'ADN retenus sont grands (Figure 3).

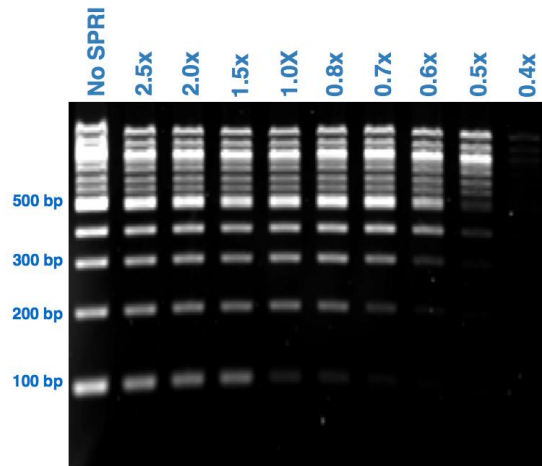


FIGURE 3 : Taille des fragments d'ADN élués (verticalement) selon le ratio du volume de billes SPRI sur le volume d'ADN (horizontalement) (d'après Broad Institute).

Les fragments de la taille souhaitée sont sélectionnés avant de purifier l'échantillon (Figure 4).

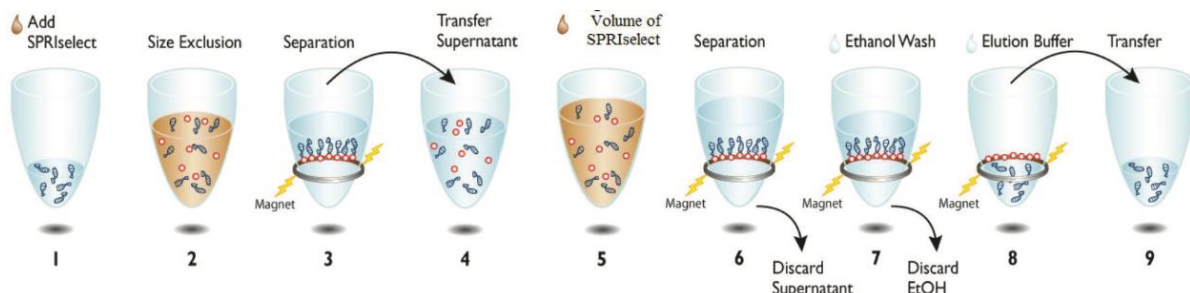


FIGURE 4: Aperçu du processus de sélection de taille puis de purification de l'échantillon (d'après Beckman Coulter).

Étape 1 : L'ADN est en solution aqueuse.

Étapes 2 et 3 : Les plus grands fragments sont éliminés en utilisant la quantité appropriée de billes SPRI pour qu'ils se fixent à leur surface.

Étape 4 : Les plus petits fragments, que l'on souhaite récupérer, se trouvent dans le surnageant qui est transféré dans un nouveau tube (étape 4).

Étape 5 : Les billes sont jetées. On ajoute un nouveau volume de billes magnétiques SPRI au surnageant afin de fixer les petits fragments d'ADN que l'on veut purifier.

Étape 6 : Les billes se fixent au support aimanté entraînant avec elles les petits fragments d'ADN.

Étape 7 : On élimine le surnageant. Deux lavages avec de l'éthanol sont effectués pour éliminer les contaminants.

Étape 8 : L'ADN est élué à l'aide d'un tampon aqueux.

Étape 9 : L'éluât contenant les fragments d'ADN de taille voulue est conservé.

La concentration et la taille moyenne des fragments sont vérifiées à l'aide de l'automate Bioanalyzer 2100 (Agilent) avec le kit DNA High Sensitivity (Agilent). Cet automate effectue une migration électrophorétique des fragments récupérés à l'aide d'une puce constituée de puits connectés entre eux par des micro-canaux en verre. La puce est tapissée d'un gel constitué d'un polymère et d'un fluorochrome. Après dépôt des échantillons, des électrodes s'emboîtent dans chaque puits de la puce et créent un gradient de tension permettant aux fragments d'ADN de migrer. Le polymère permet le tamisage moléculaire : les molécules les plus petites migrent plus rapidement que celles de grande taille. Le fluorochrome contenu dans le gel s'intercale entre les doubles brins d'ADN, rendant possible la détection par un laser des fragments par fluorescence induite. La concentration est déterminée à partir de l'intensité de fluorescence. Les données générées sont traduites en images comparables à celles d'un gel (bandes) et d'un électrophorégramme (pics). Un étalon externe appelé *ladder*, contenant des composants de taille connue, est déposé dans un puits dédié et permet d'obtenir une courbe d'étalonnage du temps de migration des fragments par rapport à leur taille. Deux marqueurs internes, de taille connue, un « *lower marker* » de petite taille et un « *upper marker* » de grande taille, sont déposés dans les puits de dépôt des échantillons et du standard externe. Ce sont des standards internes utilisés pour aligner les données de l'échantillon de manière relative à celles du standard externe (Figure 5). Ils migrent en même temps que les échantillons et que le standard externe.

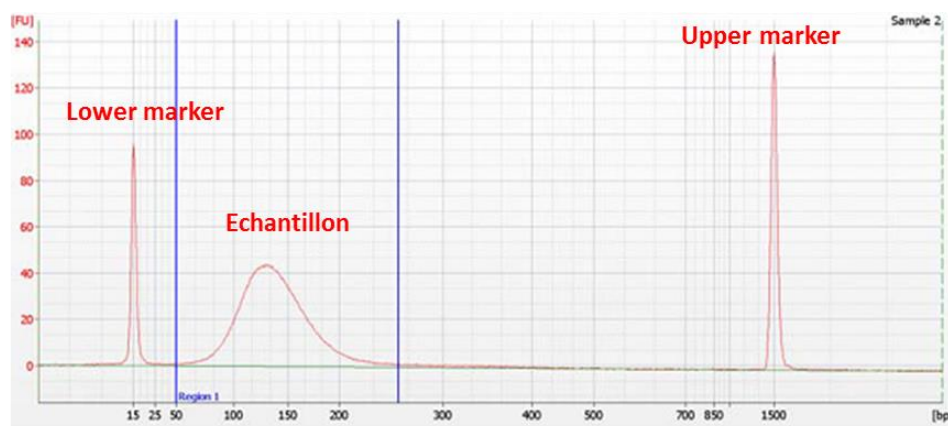


FIGURE 5 : Exemple de profil électrophorétique après fragmentation, sur puce DNA Chip High Sensitivity (Agilent), montrant les unités de fluorescence (FU) en fonction de la taille (en pb).

Les fragments de l'échantillon doivent avoir une taille comprise entre 50 et 250 pb.

II.3.2.2 Ligation des adaptateurs et des barcodes

Cette étape s'effectue à l'aide des kits Ion Xpress Plus Fragment Library et Ion Xpress Barcode Adapters (Life Technologies). Les adaptateurs Ion Xpress Barcode contiennent une

courte séquence nucléotidique (environ 10 pb) qui correspond à un code-barres (ou *barcode* ou *index*) spécifique à chaque échantillon. Ces barcodes permettent le multiplexage de tous les échantillons de la série : à l'issue du séquençage, les séquences sont réattribuées à chaque échantillon grâce à leur barcode.

Les extrémités des fragments d'ADN sont réparées par une polymérase (Nick Repair Polymerase), à laquelle est ajouté un mélange de désoxyribonucléotides. Une étape de A-tailing ajoutant une désoxyadénosine à l'extrémité 3' de chaque molécule d'ADN, est effectuée grâce à une terminale transférase. Ces deux dernières étapes permettent une ligature efficace des adaptateurs aux extrémités 3'A de chaque molécule d'ADN par l'action d'une ADN ligase (Figure 6).

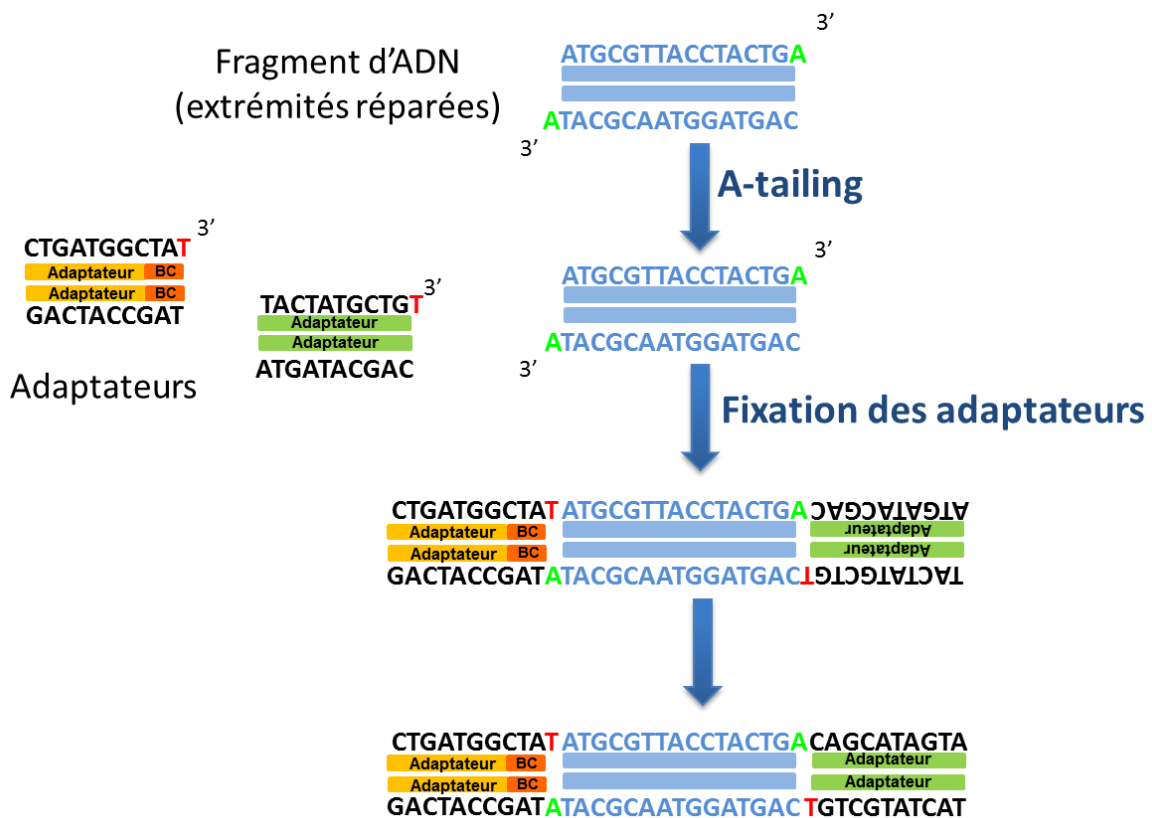


FIGURE 6 : Schéma des étapes de A-tailing et de ligation des adaptateurs aux extrémités 3'A de chaque fragment d'ADN.

BC : Barcodes.

Les échantillons sont purifiés à l'aide des billes Agencourt AMPure XP sur le même principe que vu précédemment, afin d'éliminer les adaptateurs libres.

II.3.2.3 Amplification de la librairie barcodée

Cette étape s'effectue à l'aide des kits SureSelect Target Enrichment Kit PTN Hyb (Agilent) et Herculase II Fusion DNA Polymerase (Agilent). Des amorces (ou *primers*) complémentaires des adaptateurs d'ADN de nos échantillons s'hybrident sur ces derniers et l'ajout d'un mélange de dNTP et d'une ADN polymérase permet l'amplification de la librairie par PCR classique.

Les échantillons sont purifiés à l'aide des billes Agencourt AMPure XP sur le même principe que vu précédemment, afin d'éliminer le reste d'enzyme, les dNTP et les amorces libres.

La concentration et la taille des fragments sont vérifiées à l'aide de l'automate Bioanalyzer 2100 avec le kit DNA 1000 (Agilent) comme vu précédemment (Figure 7).

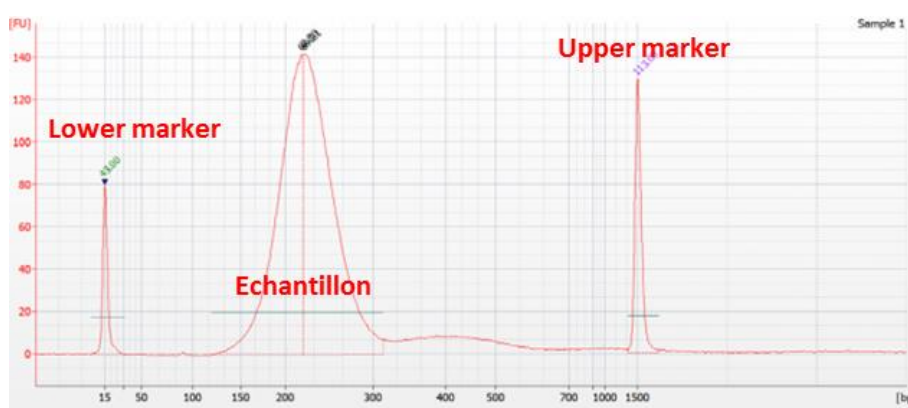


FIGURE 7 : Exemple de profil électrophorétique sur puce DNA 1000 après amplification de la librairie montrant les unités de fluorescence (FU) en fonction de la taille (en pb).

On observe un seul pic autour de 220 pb.

II.3.2.4 Concentration de la librairie par lyophilisation

Si la concentration de la librairie est inférieure à 221 ng/μL, l'échantillon est concentré à l'aide du concentrateur SpeedVac (ThermoFisher Scientific). Il s'agit d'un automate qui centrifuge l'échantillon sous vide, à une température provoquant l'évaporation de l'eau contenue dans le tube.

II.3.2.5 Capture

Dans cette analyse, seuls certains gènes précis sont séquencés. Cette étape permet de sélectionner les séquences correspondantes à ces gènes parmi la totalité de l'échantillon. Des sondes ARN spécifiques, complémentaires des fragments d'ADN cibles sont utilisées pour capturer ceux-ci par hybridation. Le design de ces sondes a été réalisé par le

laboratoire de génétique moléculaire du CHRU de Strasbourg (Dr A. PITON). Elles sont utilisées avec le kit SureSelect Target Enrichment PTN Hyb.

Les fragments d'ADN sont dénaturés pour obtenir de l'ADN simple brin. Les sondes ARN s'hybrident avec la séquence simple brin ciblée. Ces sondes sont biotinylées, c'est-à-dire combinées à des résidus de biotine. Leur capture s'effectue par ajout, après incubation, de billes magnétiques recouvertes de streptavidine, une protéine qui se lie à la biotine avec une affinité exceptionnellement élevée. La streptavidine fixe les sondes ARN biotinylées sur les billes magnétiques (Figure 8). Un aimant, placé à l'extérieur du tube pour retenir les billes contre la paroi, permet d'éliminer la fraction non capturée en jetant le surnageant.

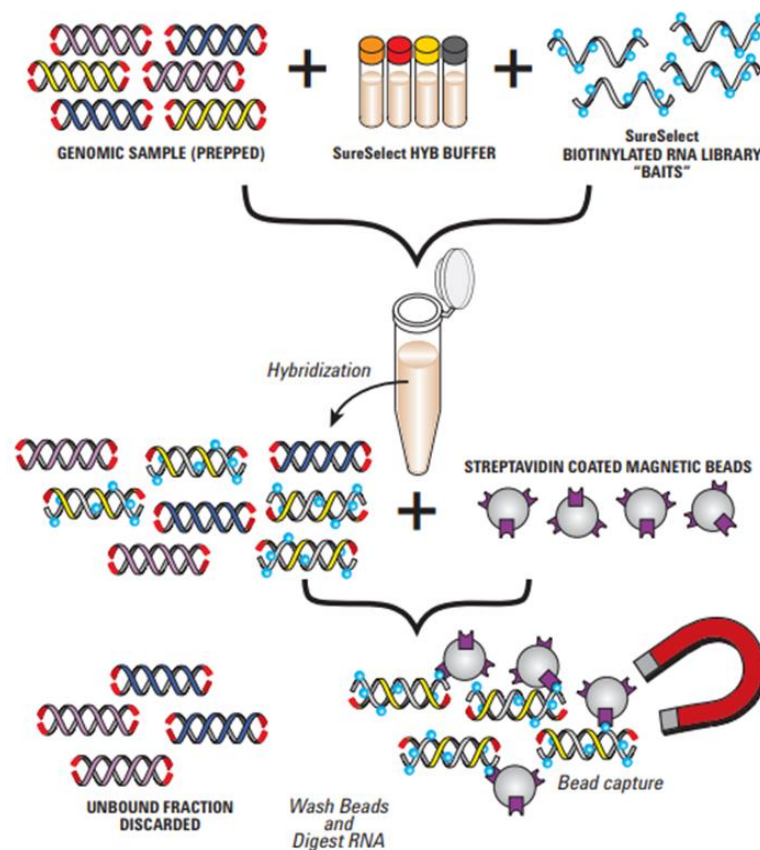


FIGURE 8 : Schéma du processus de capture par le kit SureSelect Target Enrichment (d'après Agilent Technologies).

II.3.2.6 Amplification post-capture

Cette étape s'effectue à l'aide des kits SureSelect Target Enrichment Kit PTN Hyb et Herculase II Fusion DNA Polymerase. Des amorces complémentaires des adaptateurs des fragments d'ADN toujours fixés aux billes de streptavidine s'hybrident sur ces derniers. L'ajout d'un mélange de dNTP et d'une ADN polymérase amplifie les fragments par PCR classique.

Les échantillons sont purifiés à l'aide des billes Agencourt AMPure XP sur le même principe que vu précédemment, afin d'éliminer le reste d'enzyme, les dNTP et les amorces libres. L'ensemble des billes est éliminé à cette étape.

La qualité des fragments capturés et amplifiés est vérifiée à l'aide de l'automate Bioanalyzer 2100 avec le kit DNA High Sensitivity, de la même manière qu'auparavant avec la puce DNA 1000 (Figure 9).

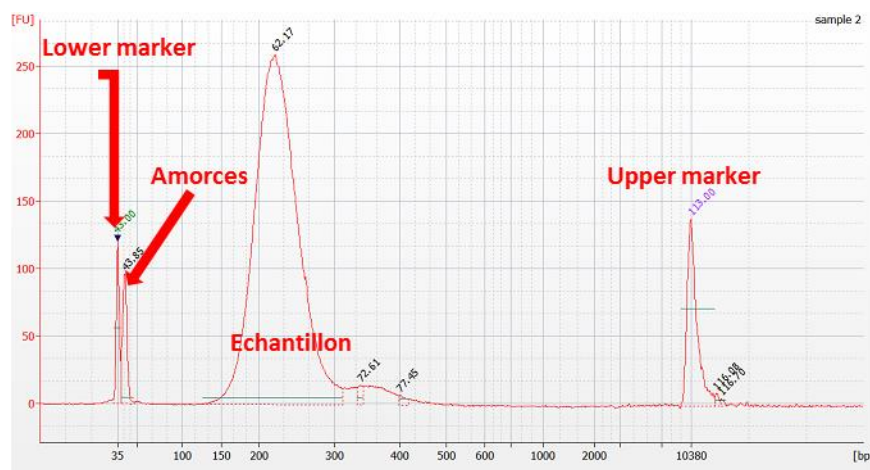


FIGURE 9 : Exemple de profil électrophorétique sur puce High Sensitivity DNA après capture et amplification montrant les unités de fluorescence (FU) en fonction de la taille (en pb).

On observe un pic « échantillon » autour de 220 pb. Un pic autour de 40 pb correspond aux amorces restant dans le tube.

II.3.2.7 Multiplexage des échantillons

Durant cette étape, l'ensemble des échantillons barcodés sont regroupés dans un même tube, en quantité équimolaire.

La taille des fragments d'ADN de la librairie doit se trouver aux alentours 220 pb (\pm 40 bases). Ils sont pour cela sélectionnés grâce à une électrophorèse sur un gel d'agarose prémoulé à 2 %. Le principe est le même que celui d'une électrophorèse classique avec, en plus, la présence, à un endroit précis sur le gel, d'un « puits de récupération » se trouvant au niveau de la bande correspondant aux fragments de 220 pb, permettant de les récupérer.

Les échantillons sont purifiés à l'aide des billes Agencourt AMPure XP. La qualité des fragments est vérifiée à l'aide de l'automate Bioanalyzer 2100 avec le kit DNA High Sensitivity comme vu précédemment (Figure 10).

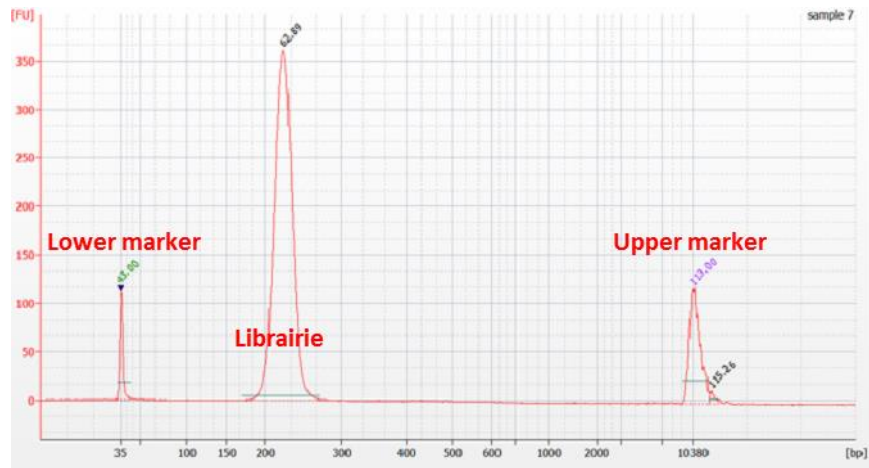


FIGURE 10 : Exemple de profil électrophorétique de la librairie finalisée sur puce High Sensitivity DNA.

II.3.3 Séquençage

Les individus sont séquencés par séries de 16 échantillons sur séquenceur Ion Proton (Life Technologies). Une série s'effectue sur deux puces, soit huit échantillons par puce. Pour une seule initialisation de l'Ion Proton, les deux puces sont chargées en parallèle et séquencées. Au total, 66 individus, quatre témoins positifs ainsi que 10 pools de parents sont séquencés en cinq séries.

II.3.3.1 Préparation du « template » ou matrice de séquençage

Cette étape d'amplification clonale et d'enrichissement est réalisée à l'aide du système OneTouch Ion 2 et du kit Ion PI Template OT2 200 v2. Elle est essentielle pour obtenir un signal suffisant et détectable lors du séquençage.

L'amplification clonale se fait sur un principe de PCR en émulsion (emPCR). Il s'agit d'une phase aqueuse contenant l'ensemble des produits nécessaires à la PCR qui est émulsionnée et forme des gouttelettes aqueuses dans une phase huileuse. Dans les conditions optimales de ce processus, une gouttelette encapsule des billes ISP (Ion Sphere Particle), un fragment d'ADN et des réactifs de PCR contenant des nucléotides, du tampon, des ions magnésium, des amorces complémentaires des adaptateurs et de la Taq polymérase. D'autres configurations sont possibles comme des gouttelettes contenant une ou plusieurs billes avec plusieurs fragments de librairie ou sans aucun fragment (Figure 11). La configuration optimale est la configuration monoclonale avec une bille et une unique molécule d'ADN. La librairie est préalablement quantifiée afin de pouvoir optimiser le ratio

ISP/fragments d'ADN pour qu'il soit proche de un. Une gouttelette fonctionne comme un microréacteur de PCR et amplifie les fragments d'intérêt de façon exponentielle (Figure 12).

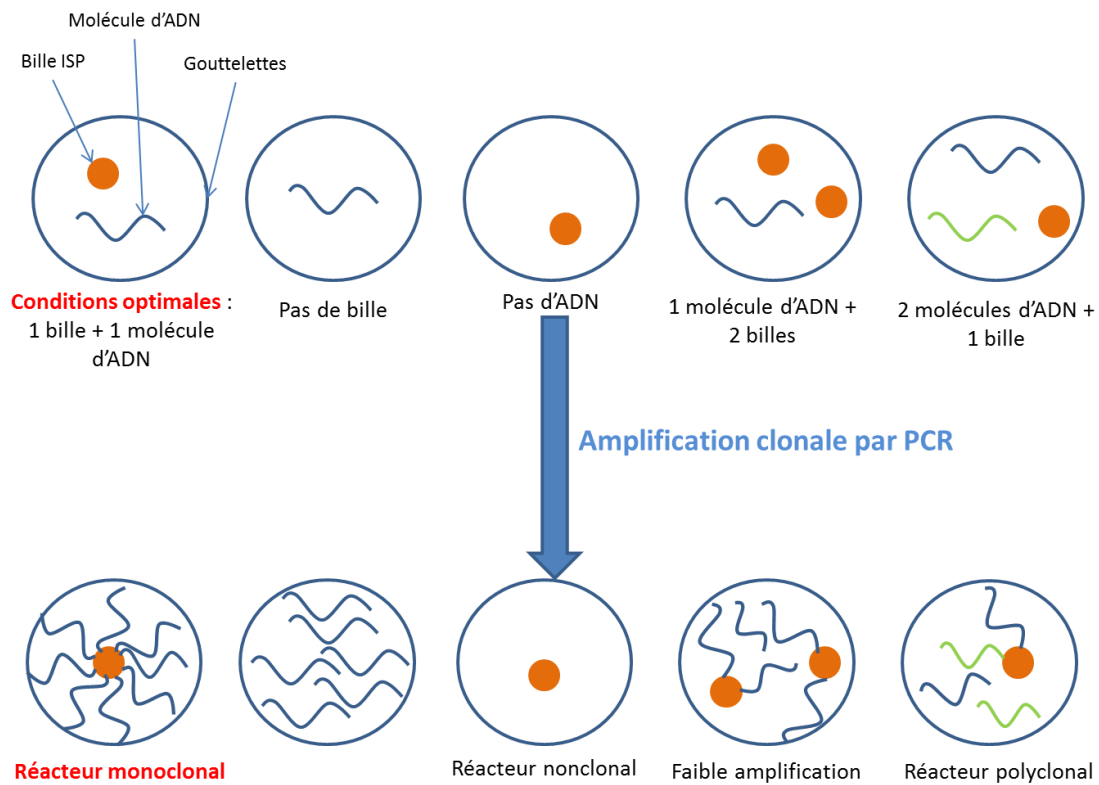


FIGURE 11 : Schéma des différentes configurations pouvant être rencontrées lors de l'émulsion des gouttelettes d'eau en phase huileuse lors d'une PCR en émulsion.

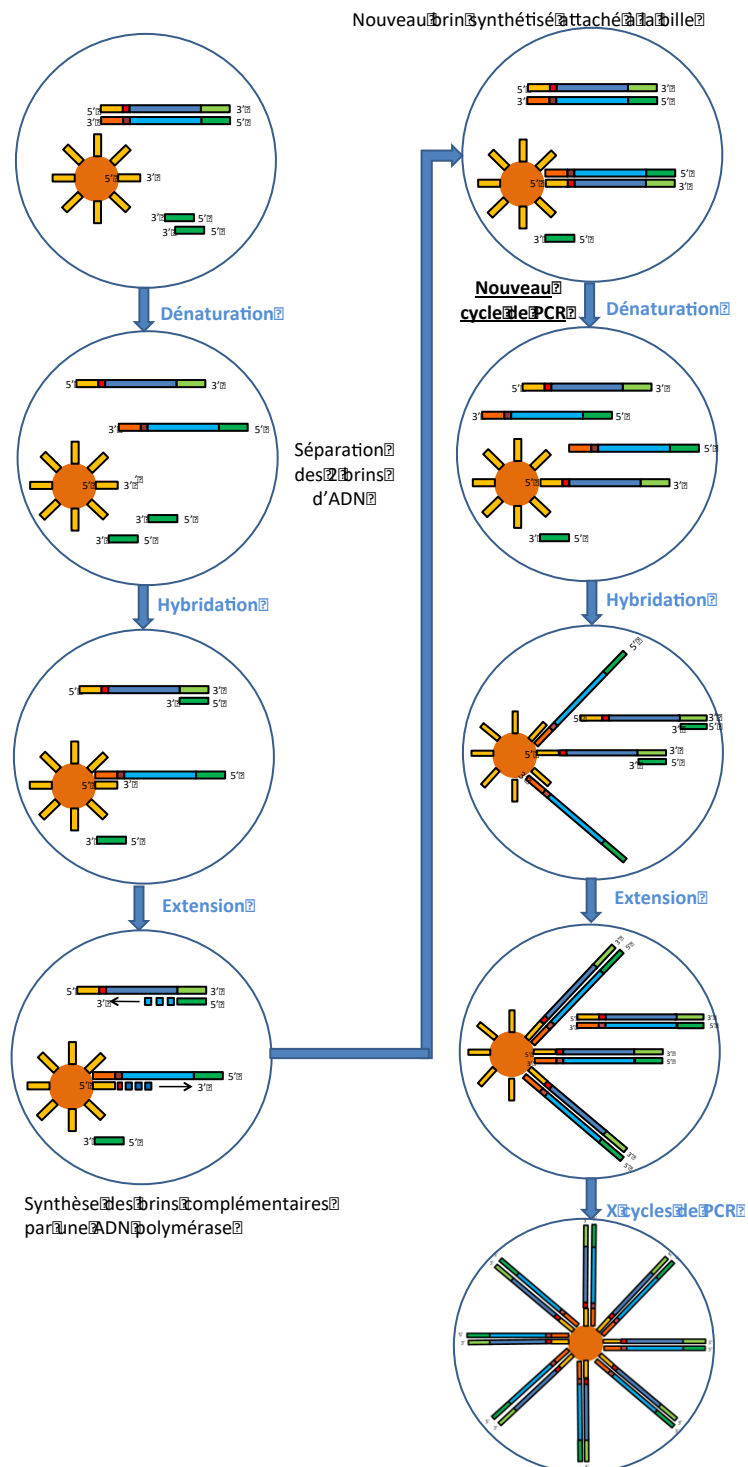


FIGURE 12 : Schéma d'une microgouttelette, illustrant le principe de la PCR en émulsion.

À la surface des billes se trouvent des adaptateurs complémentaires de ceux liés à un des brins du fragment d'intérêt. Après séparation des 2 brins d'ADN, chacun s'hybride avec son adaptateur complémentaire : un avec celui de la bille, l'autre avec celui présent dans la microgouttelette. La polymérase synthétise les brins complémentaires. Les 4 brins d'ADN obtenus subissent à leur tour un nouveau cycle de PCR, et ainsi de suite jusqu'à l'amplification souhaitée (plusieurs millions de copies par bille).

Après un contrôle qualité des ISP par un dosage avec le fluorimètre Qubit 2.0, une étape d'enrichissement a lieu. Les amorces du mix de PCR, non fixées sur les billes ISP sont biotinyllées. Cela permet, d'isoler les billes ISP portant des amplicons à leur surface grâce à l'ajout de billes magnétiques recouvertes de streptavidine.

II.3.3.2 Dépôt sur puce et séquençage

Le kit Ion PI Sequencing 200 v2 (Life Technologies) est utilisé à cette étape. Avant de lancer le séquençage, le Proton subit des lavages au chlorite et à l'eau pour nettoyer les tubulures de l'appareil puis une initialisation afin d'homogénéiser son pH dans l'ensemble des réactifs utilisés. L'échantillon est déposé sur une puce semi-conductrice PI, placée ensuite dans le séquenceur Ion Proton. Sur ces puces, la capacité annoncée par le fabricant est de 10 gigabases de données brutes.

Le séquençage est basé sur une propriété biochimique des ADN polymérase : le relargage d'un ion hydrogène à chaque incorporation de nucléotide lors de la polymérisation. La libération de proton est responsable d'une variation de pH qui est détectée directement dans chaque puits de la puce. La puce semi-conductrice de séquençage est composée de plusieurs millions de puits associés à un capteur au niveau de la couche mince (technologie des semi-conducteurs) qui détecte les variations de pH et les transforme en variations électriques (variations de voltage). Chaque puits a un diamètre ne lui permettant d'accueillir qu'une seule bille (Figure 13). Le brin complémentaire de celui fixé à la surface de la bille est synthétisé. Des bains successifs contenant un seul nucléotide (A, T, G ou C) sont injectés dans ces puits. Si le nucléotide contenu dans le bain est complémentaire de la base du fragment d'ADN fixé sur la bille, il est incorporé dans le brin en cours de synthèse grâce à la polymérase et libère un ion hydrogène. Sinon, rien ne se passe. Les puits sont rincés et un nouveau nucléotide est testé. Le brin complémentaire est synthétisé par ce système de « bains » successifs de chaque nucléotide, générant la séquence du fragment d'intérêt fixé sur la bille.

Le séquençage est de type « single-end read » : le fragment est séquencé en totalité à l'aide d'une seule amorce universelle complémentaire des adaptateurs fixés en 3' de la séquence d'intérêt. Une séquence clé (TCAG), située juste avant la séquence d'intérêt, permet d'apprécier l'intensité de signal correspondant à l'incorporation d'une base et d'étalonner chaque puits.

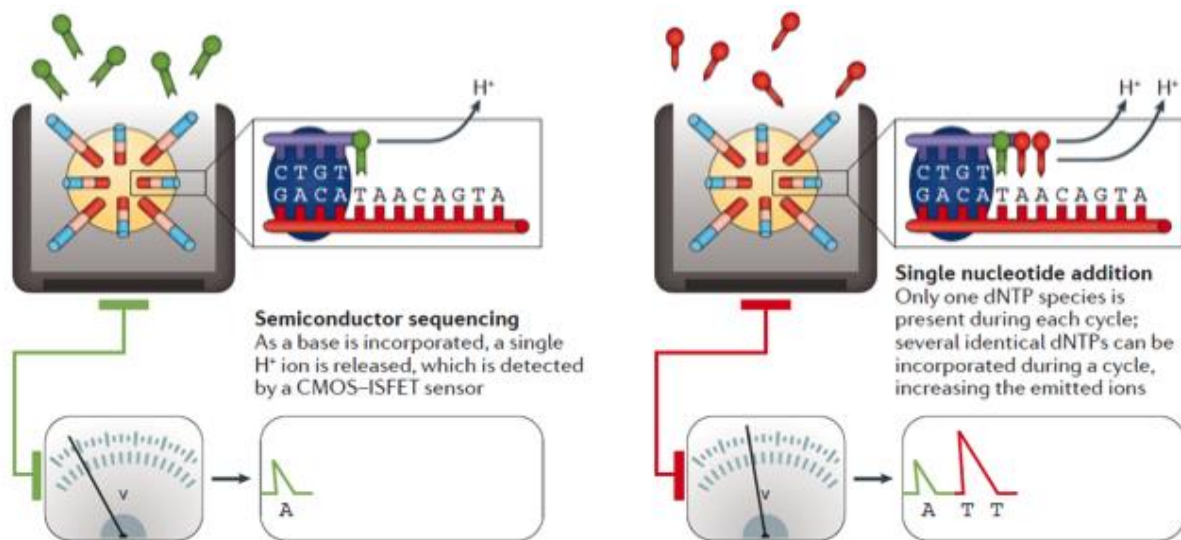


FIGURE 13 : Schéma du processus de séquençage Ion Torrent (d'après Goodwin, McPherson, et McCombie 2016)

Une bille occupe un seul puits de réaction. L'ajout successif des différents nucléotides chacun leur tour dans les puits permet une réaction d'élongation standard pour la synthèse du brin complémentaire au fragment cible. Lorsqu'une base est incorporée, un seul ion hydrogène est généré, entraînant une variation de pH détectée par un semi-conducteur complémentaire à l'oxyde de métal (CMOS) relié à un dispositif de transistor à effet de champs sensible aux ions (IFSET). Une fois qu'un nucléotide est introduit, les bases non incorporées sont éliminées par lavage et le cycle recommence. Plusieurs nucléotides identiques peuvent être incorporés pendant un cycle, augmentant la quantité d'ions émis donc la variation de pH.

Le séquençage d'un run dure en moyenne 2 h30 min. La manipulation complète d'une série dure une semaine.

II.4 Pools des parents

Pour chaque série, des échantillons provenant des deux parents de chaque patient – si l'ADN est disponible – sont manipulés et séquencés en parallèle. L'ensemble des échantillons provenant des pères sont poolés, formant le « pool des pères » traité comme un seul échantillon patient. Tous les ADN de ce pool ont le même code barre. Il en est de même pour l'ensemble des échantillons provenant des mères. Sur chacune des deux puces, un huitième échantillon correspondant à un des pools parents est déposé : le pool des pères sur une puce et le pool des mères sur l'autre puce.

Le séquençage des pools des parents fournit une aide à l'interprétation des variants en renseignant sur la ségrégation de ces derniers. Si un variant est mis en évidence dans le

pool des pères, sa probabilité d'appartenir à un père autre que celui du patient est très faible (= fréquence du variant dans la population × nombre de père dans le pool). Il en est de même pour le pool des mères. Ainsi le variant peut être :

- hérité d'un des deux parents : en faveur d'un polymorphisme sans effet sur le phénotype si le parent est en bonne santé ou d'un mode de transmission dominant si le parent est atteint ;
- hérité des deux parents : en faveur d'un mode de transmission récessif ;
- *de novo* : en faveur d'un mode de transmission dominant.

II.5 *Dry lab* ou analyse informatique des données de séquençage

II.5.1 *Base calling* ou détermination de la séquence

À la fin du *run* (série) de séquençage, les signaux générés par les variations de pH sont convertis en données de séquences, c'est-à-dire en bases A, T, C, G. Des millions de séquences, appelées *reads* ou lectures sont générées sous la forme d'un fichier texte au format FASTA.

Un score de qualité, le score Q ou score Phred, est associé à chacune des bases identifiées. Il représente la probabilité d'erreur d'identification de la base : $Q = -10 \log_{10} P$, avec P la probabilité d'erreur d'identification de la base (Tableau I). Le fichier associé est un fichier FASTQ renseignant le score de qualité associé à chaque base de la séquence. Ce score de qualité est codé par des symboles où chaque symbole correspond à une valeur de qualité.

Score Q	P ou probabilité d'erreur d'identification de la base	Probabilité d'un appel de base exact
10	$\frac{1}{10}$	90 %
20	$\frac{1}{100}$	99 %
30	$\frac{1}{1000}$	99,9 %
40	$\frac{1}{10\ 000}$	99,99 %
50	$\frac{1}{100\ 000}$	99,999 %

TABLEAU I : Correspondance de la valeur du score Q avec la probabilité d'une identification de base exacte.

Un score Q égal à 10 (Q10) a une probabilité P d'identification d'une base incorrecte de 1 sur 10, c'est-à-dire que dans chaque séquence de 10 pb, on peut s'attendre à une erreur.

II.5.2 Contrôles de qualité primaires

Différentes métriques, en plus du Q score, permettent d'apprécier la qualité du séquençage. Un rapport de données de qualité, appelé FastQC report, permet de contrôler visuellement et rapidement différents points essentiels tels que le score de qualité par base et par séquence, la distribution de la taille des séquences, la proportion de nucléotides indéterminés et celle de chaque nucléotide (A, T, C, G), le pourcentage de GC, la proportion de reads dupliquées et d'adaptateurs.

II.5.3 Alignement ou *mapping*

Durant cette étape, l'ensemble des reads sont positionnés sur une séquence dite de référence. Ici, il s'agit de la version GRCh37 du génome humain de référence. On obtient un fichier texte au format SAM (pour Sequence Alignment / Map) qui occupe une place très importante, de l'ordre du Go à une centaine de Go. C'est pourquoi, il est converti au format BAM (pour Binary sAM), la forme binaire et compressée du fichier SAM.

II.5.4 Contrôles de qualité secondaires

De nouveaux contrôles de qualité sont disponibles à l'issue de l'alignement, fournis par le logiciel Torrent Ion Suite. On vérifie l'enrichissement – les pourcentages de reads alignés et de reads alignés sur la cible – ainsi que deux paramètres importants : la profondeur de couverture (*depth*) et la couverture de séquence (*coverage*) (Figure 14).

La profondeur de couverture, exprimée en nombre de fois (X), est le nombre de reads alignés sur une base par le séquenceur. Elle peut être exprimée par une valeur pour une base précise ou par une moyenne lorsqu'il s'agit d'une séquence.

La couverture de séquence est le pourcentage de bases séquencées par rapport au nombre total de bases de la région cible.

La couverture de séquençage à une profondeur minimale donnée est le pourcentage de bases séquencées par rapport au nombre total de bases de la région cible, avec une profondeur X donnée (Collège National des Enseignants et Praticiens de Génétique Médicale 2016).

Ces mesures reflètent la qualité de l'analyse d'une manière globale. Si leurs valeurs sont insuffisantes, le risque de résultat faussement négatif est accru, entraînant une baisse de sensibilité de la technique.

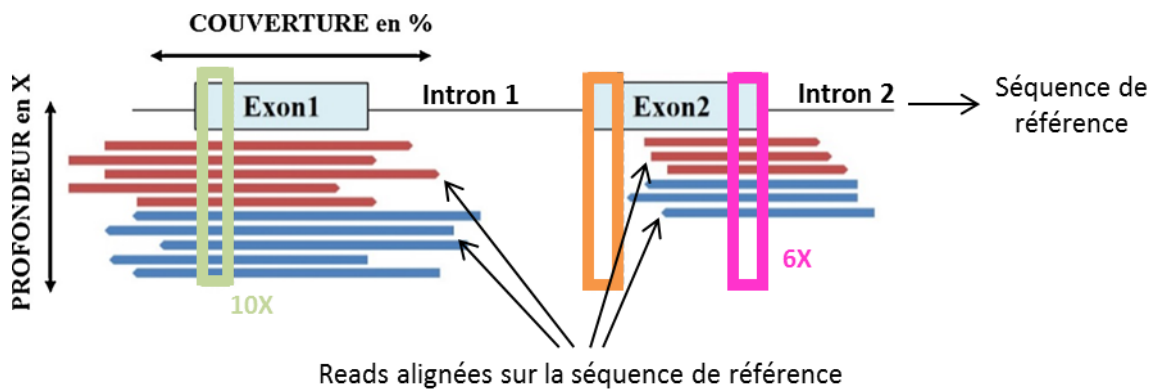


FIGURE 14 : Illustration des notions de profondeur et de couverture de séquençage (modifié d'après Lacoste et al. 2017).

Les flèches rouges représentent les reads sens et les bleus les reads antisens. Les exons 1 et 2 correspondent aux régions ciblées par le séquençage. La couverture de séquençage se lit horizontalement et la profondeur se lit verticalement. L'exon 1 est couvert à 100 % car tous les nucléotides qui le composent ont été lus au moins une fois. La zone encadrée en vert a une profondeur de 10X. En effet, 10 reads « s'empilent » dans l'encadré. Au contraire, l'exon 2 n'est pas couvert à 100 % : la zone orange n'est chevauchée par aucun read : elle n'est pas couverte. La fin de l'exon 2 est couverte avec une profondeur de 6X (encadré rose).

II.5.5 Variant calling ou identification des variants

À cette étape, les différences entre la séquence de référence et les reads générés par le séquençage sont identifiées. Ces différences sont appelées variants ou variations. Tous les variants détectés sont regroupés dans un fichier texte au format VCF (pour Variant Calling Format).

II.5.6 Annotation des variants

L'annotation consiste, pour chaque variant précédemment identifié, à enrichir le VCF en informations grâce à la consultation de multiples bases de données. Cela permet d'appréhender ces variants et de leur créer une « carte d'identité » en renseignant, par exemple, le gène touché, sa fonction, les pathologies associées, la fréquence du variant dans la population, le type de variant, sa localisation (exonique, intronique...), l'effet prédit par différents outils *in silico*, etc. À partir de ces données, il est possible d'appliquer des filtres spécifiques pour éliminer la majorité des variants bénins ou probablement bénins. Les données recueillies pour l'annotation et les filtres utilisés sont détaillés par la suite pour chacune des analyses du panel.

Les variants restants sont priorisés par le biologiste en prenant en compte un ensemble de critères suggérant un caractère pathogène ou bénin. Ces critères permettent d'étiqueter le variant selon les termes de l'ACMG.

Voici un descriptif des recommandations, des bases de données et outils utilisés pour l'annotation des variants.

II.5.6.1 Recommandations de l'ACMG

En 2015, l'ACMG (American College of Medical Genetics) a publié des normes et lignes directrices pour l'interprétation des variants avec l'objectif d'homogénéiser le rendu des résultats intra- et inter-laboratoires. Elles recommandent l'utilisation d'une terminologie standard pour classer les variants identifiés en cinq classes :

- variant bénin (classe 1) ;
- variant probablement bénin (classe 2) ;
- variant de signification inconnue ou VUS (classe 3) ;
- variant probablement pathogène (classe 4) ;
- variant pathogène (classe 5).

L'assignation d'un variant à l'une de ces cinq catégories se déduit d'un faisceau d'arguments avec un poids plus ou moins important dans l'interprétation du variant. L'ACMG fournit deux séries de critères : une pour la classification des variants pathogènes ou probablement pathogènes (Annexe 5) et une autre pour les variants bénins ou probablement bénins (Annexe 4). Les arguments en faveur de la pathogénicité d'un variant sont qualifiés de « très forts » (PVS ou *very strong*), « forts » (PV ou *strong*), « modérés » (PM ou *moderate*) ou « faibles » (PP ou *supporting*). Les arguments en faveur de la bénignité d'un variant sont qualifiés de « suffisants » (BA ou *stand-alone*), « forts » (BS ou *strong*) ou « faibles » (BP ou *supporting*). Ces arguments s'appuient, entre autres, sur des données épidémiologiques, de prédiction *in silico*, de ségrégation, fonctionnelles et alléliques (Annexes 3, 4 et 5). Les arguments retenus sont pondérés (Annexe 6) afin d'assigner une classification au variant. Un variant probablement pathogène ou probablement bénin signifie que le variant est pathogène ou bénin avec une probabilité de 90 % (Richards *et al.* 2015).

II.5.6.2 Bases de données de variants spécifiques de populations

II.5.6.2.1 dbSNP

La base de données dbSNP (Single Nucleotide Polymorphism database) rassemble un ensemble de SNP et petites variations de séquence (substitutions, insertions ou délétions

d'une taille maximale de 50 pb ainsi que des microsatellites), qui apparaissent assez fréquemment dans une population pour être qualifiés de polymorphes. Si une variation contient plus de 50 nucléotides, elle peut être soumise dans dbVar, une base de données similaire à dbSNP mais pour les variations structurelles génomiques. Pour chaque variation sont renseignés le contexte de séquence, les fréquences populationnelles du variant si elles sont disponibles, sa zygosity et la ou les méthodes expérimentales, protocoles et conditions utilisés pour évaluer ce variant par les personnes qui l'ont soumis. Les soumissions individuelles pour un même variant sont regroupées en un dossier auquel est attribué une référence dbSNP (numéro rs). Ce dossier contient des données de synthèse comme la signification clinique dans ClinVar, l'association avec le phénotype, l'origine somatique ou germinale de l'allèle, etc.

II.5.6.2.2 ExAC

ExAC (Exome Aggregation Consortium) est une base de données de variants spécifiques de populations. Elle fournit des données de séquençage d'exome pour plus de 60 000 individus appartenant à diverses populations et contient plus de 10 millions de variants de séquence. Cette base de données renseigne sur la fréquence allélique des variants génétiques codant une protéine. Ces données épidémiologiques peuvent être utilisées pour filtrer les variants candidats dans les maladies mendéliennes. Les variants génétiques délétères sont habituellement rares dans la population générale par pression de sélection négative (Lek *et al.* 2016). Une fréquence allélique dans une population de contrôle plus élevée que la prévalence du trouble est un des arguments forts en faveur de l'interprétation d'un variant comme étant bénin dans le cas d'une maladie mendélienne rare (Richards *et al.* 2015). Cependant, plus de 20 % des données présentes dans ExAC proviennent d'individus atteints de schizophrénie, un trouble qui partage les mêmes gènes et voies moléculaires que la DI, les troubles du spectre autistique et l'encéphalopathie épileptique (Li *et al.* 2016). Les filtres basés sur ExAC doivent être utilisés avec précaution car ils peuvent être à l'origine de faux négatifs.

II.5.6.2.3 gnomAD

La base de données gnomAD (genome Aggregation Database) est une version plus récente d'ExAC. Elle contient les informations de cette dernière et regroupe au total les données de plus de 123 000 exomes et 15 000 génomes d'individus non apparentés séquencés lors de diverses études (Lek *et al.* 2016).

II.5.6.3 ClinVar

ClinVar est une base de données de variants que le NCBI (National Center for Biotechnology Information) met à disposition gratuitement. En juillet 2018, la base contenait plus de 430 000 variants sur plus de 30 000 gènes différents rapportés chez des individus ou familles, dans un cadre clinique ou de recherche, en fonction de leurs caractéristiques cliniques et de leur mode d'hérédité. Parfois, des données fonctionnelles basées sur des preuves expérimentales sont disponibles et permettent d'éclairer l'interprétation basée sur la clinique. Ces variants sont rapportés par des laboratoires de diagnostic clinique ou de recherche, des groupes d'experts et des groupes établissant des recommandations professionnelles de tous les pays. Elles peuvent également être enrichies par des bases de données, notamment des bases de données spécifiques de certains locus. ClinVar renseigne sur l'interprétation du variant par les différentes personnes qui soumettent ce variant (selon les cinq termes recommandés par l'ACMG). Si des interprétations données par des auteurs ne sont pas concordantes pour un même variant, ClinVar enregistre l'interprétation comme conflictuelle. Pour chaque soumission, les preuves appuyant l'interprétation du variant sont consultables (Landrum *et al.* 2016). À chaque variant est attribué un niveau de « revue de statut » (*status review*). Ce niveau dépend des critères d'affirmation et des preuves fournies, de l'accord sur l'interprétation entre les soumissions et du type de personne ayant soumis le variant (le *status review* est plus élevé s'il s'agit d'un groupe d'experts ou d'un groupe établissant des recommandations professionnelles). Ce niveau est représenté graphiquement par un système d'étoiles dont le nombre va de zéro à quatre – quatre étant le meilleur niveau, – et représente le degré de confiance de l'interprétation clinique du variant (Landrum *et al.* 2016; Landrum *et al.* 2014).

II.5.6.4 OMIM

OMIM (Online Mendelian Inheritance in Man) est une base de données de gènes, de maladies et de phénotypes. Il constitue le référentiel principal d'informations sur les gènes, les phénotypes humains et les relations génotype-phénotype. Les informations de la littérature y sont synthétisées et résumées par un groupe d'experts (Amberger *et al.* 2015). La base est constamment mise à jour. En juillet 2018, OMIM répertoriait plus de 15 900 gènes, dont 3900 avec un phénotype associé et plus de 6000 phénotypes pour lesquels la base moléculaire est connue. Elle permet d'apprécier la corrélation génotype-phénotype pour l'individu analysé et le mode de transmission.

II.5.6.5 Mutalyzer

Mutalyzer est une suite de logiciels libres qui permettent de convertir, générer, vérifier la nomenclature d'un variant selon celle recommandée par l'HGVS (Human Genome Variation Society) (Wildeman *et al.* 2008).

II.5.6.6 VarSome

VarSome est un agrégateur de connaissances sur la génomique humaine. Il contient des informations provenant de diverses bases de données publiques comme dbSNP, ClinVar, gnomAD, RefSeq, Ensembl, PubMed, etc. On peut y chercher des informations sur un variant, un gène, une publication, une maladie, un phénotype, un transcrit, une position, une région, etc.

II.5.6.7 Outils de prédiction *in silico*

II.5.6.7.1 PolyPhen-2

PolyPhen-2 (Polymorphism Phenotyping v2) est un logiciel permettant de prédire l'impact possible des substitutions d'acides aminés sur la stabilité et la fonction des protéines humaines. Il prend en compte des critères structuraux et évolutifs qui lui permettent d'estimer la probabilité pour un variant faux-sens non synonyme d'être délétère. Plusieurs alignements et annotations de séquences sont effectués à l'aide de plusieurs bases de données de séquences protéiques. Cela permet à PolyPhen de calculer un score à l'aide d'une approche bayésienne. Pour cela il s'appuie sur divers critères comme la participation de l'acide aminé touché à une liaison covalente, son appartenance à un site ou une région d'intérêt, la fréquence d'observation de cette variation dans cette famille protéique, les interactions potentielles avec son environnement proche (à partir de la structure tridimensionnelle de la protéine), etc. Le score obtenu est compris entre 0,0 et 1,0 – de l'impact le plus faible au plus fort – et classe le variant selon des seuils : bénin (*benign*), possiblement délétère (*possibly damaging*) ou probablement délétère (*probably damaging*). L'utilisateur peut choisir entre deux corpus d'apprentissage pour la prédiction : HumDiv et HumVar. Le premier, HumDiv, regroupe tous les allèles de la base de données UniProtKB ayant un effet délétère connu dans les maladies mendéliennes, ainsi que les différences supposées non délétères, entre les protéines humaines et leurs homologues chez les mammifères proches. Il est adapté pour l'étude des allèles rares à des locus potentiellement impliqués dans phénotypes complexes pour lesquels même les allèles faiblement délétères doivent être considérées comme pathogènes. HumVar, quant à lui, se compose de toutes les mutations humaines pathogènes répertoriées dans UniProtKB et de tous les SNP humains non synonymes courants (c'est-à-dire avec une fréquence de l'allèle mineur supérieure à 1 %). L'utilisation

de ce modèle est préconisée dans le diagnostic des maladies mendéliennes dans lesquelles il est important de distinguer les mutations avec un effet fortement délétère de celles avec un effet faiblement délétère, devant être considérées comme non pathogènes (Adzhubei, Jordan, et Sunyaev 2013).

II.5.6.7.2 SIFT

SIFT (Sorting Intolerant From Tolerant) prédit si une substitution d'acides aminés dans une protéine affecte la fonction de la protéine en utilisant l'homologie de séquence. À partir de la séquence protéique d'intérêt, SIFT choisit des protéines apparentées, possédant des séquences similaires. Il aligne la séquence protéique analysée sur ces dernières et les compare. Cet outil repose sur l'hypothèse que les acides aminés importants sont conservés dans les protéines d'une même famille au cours de l'évolution et prédit que les changements aux positions bien conservées peuvent être nocifs. Pour une position et un type de changement d'acide aminé donnés, il attribue un score reflétant la probabilité normalisée que le changement soit toléré. Les substitutions avec des scores inférieurs à 0,05 sont prédites comme délétères (*damaging*). Sinon, elles sont considérées comme neutres ou tolérées (*tolerated*) (Ng et Henikoff 2003).

II.5.6.8 VEP

VEP (Variant Predictor Effect) est un outil du projet Ensembl, qui est un système bioinformatique qui donne accès de manière rapide à de multiples données scientifiques publiques. Son but est d'accélérer les recherches en génomique. Dans Ensembl, les gènes et les transcrits ont un identifiant appelés respectivement ENSG et ESNT (Zerbino *et al.* 2018).

VEP est un outil open source et gratuit, permettant d'analyser, d'annoter et de prioriser les variants (McLaren *et al.* 2016). Il peut s'utiliser pour des requêtes rapides via son interface web mais également être utilisé en ligne de commande. VEP est utilisé par le pipeline maison, qui ne reprend qu'une partie des informations disponibles, à savoir la « conséquence » ainsi que l'« IMPACT » du variant tel qu'il est indiqué par VEP.

La « conséquence » d'un variant indiquée par VEP utilise les termes de « Sequence Ontology » (SO) (Figure 15) (Eilbeck *et al.* 2005). Cette dernière regroupe un vocabulaire contrôlé et structuré formant un ensemble commun de termes et définitions pour l'annotation génomique afin de faciliter l'échange, l'analyse et la gestion des données génomiques en harmonisant le langage des utilisateurs.

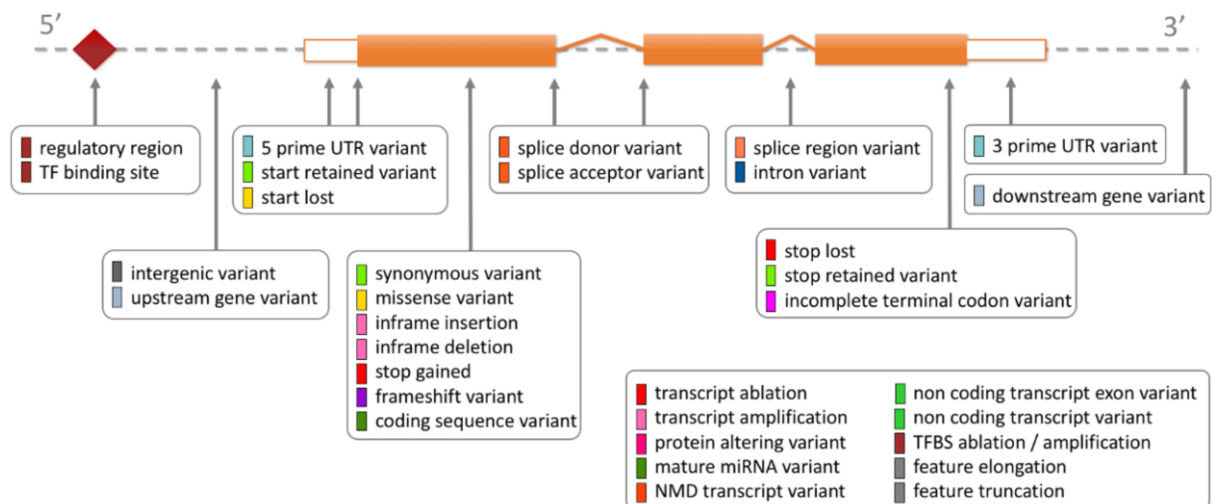


FIGURE 15 : Diagramme montrant chaque « conséquence » par rapport à la structure génomique (d'après Ensembl).

Un « IMPACT » est associé à chaque variant selon sa conséquence. Pour une conséquence donnée, il y a un IMPACT correspondant (Tableau II). Il dépend uniquement du type de variant et non pas du variant lui-même. L'IMPACT peut être HIGH, MODERATE, LOW ou MODIFIER.

- L'annotation HIGH est associée aux variants supposés fortement perturbateurs comme par exemple ceux entraînant une perte de fonction de la protéine.
- L'annotation MODERATE est associée aux variants sans perte de fonction mais pouvant affecter l'efficacité de la protéine.
- L'annotation LOW est utilisée pour les variants peu susceptibles d'impacter la protéine.
- Les variants habituellement dans les séquences non codantes sont classés comme MODIFIER.

* SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice region variant	LOW
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Incomplete terminal codon variant	LOW
start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains	SO:0002019	Start retained variant	LOW
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	SO:0001567	Stop retained variant	LOW
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid	SO:0001819	Synonymous variant	LOW
coding_sequence_variant	A sequence variant that changes the coding sequence	SO:0001580	Coding sequence variant	MODIFIER
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA	SO:0001620	Mature miRNA variant	MODIFIER
5_prime_UTR_variant	A UTR variant of the 5' UTR	SO:0001623	5 prime UTR variant	MODIFIER
3_prime_UTR_variant	A UTR variant of the 3' UTR	SO:0001624	3 prime UTR variant	MODIFIER
non_coding_transcript_exon_variant	A sequence variant that changes non-coding exon sequence in a non-coding transcript	SO:0001792	Non coding transcript exon variant	MODIFIER
intron_variant	A transcript variant occurring within an intron	SO:0001627	Intron variant	MODIFIER
NMD_transcript_variant	A variant in a transcript that is the target of NMD	SO:0001621	NMD transcript variant	MODIFIER
non_coding_transcript_variant	A transcript variant of a non coding RNA gene	SO:0001619	Non coding transcript variant	MODIFIER
upstream_gene_variant	A sequence variant located 5' of a gene	SO:0001631	Upstream gene variant	MODIFIER
downstream_gene_variant	A sequence variant located 3' of a gene	SO:0001632	Downstream gene variant	MODIFIER
TFBS_ablation	A feature ablation whereby the deleted region includes a transcription factor binding site	SO:0001895	TFBS ablation	MODIFIER
TFBS_amplification	A feature amplification of a region containing a transcription factor binding site	SO:0001892	TFBS amplification	MODIFIER
TF_binding_site_variant	A sequence variant located within a transcription factor binding site	SO:0001782	TF binding site variant	MODIFIER
regulatory_region_ablation	A feature ablation whereby the deleted region includes a regulatory region	SO:0001894	Regulatory region ablation	MODERATE
regulatory_region_amplification	A feature amplification of a region containing a regulatory region	SO:0001891	Regulatory region amplification	MODIFIER
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO:0001907	Feature elongation	MODIFIER
regulatory_region_variant	A sequence variant located within a regulatory region	SO:0001566	Regulatory region variant	MODIFIER
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO:0001906	Feature truncation	MODIFIER
intergenic_variant	A sequence variant located in the intergenic region, between genes	SO:0001628	Intergenic variant	MODIFIER

TABLEAU II : Correspondances entre les conséquences (SO term), leur définition selon Sequence Ontology (SO description), leur référence dans Sequence Ontology (SO accession), le terme affiché (Display term) et leur IMPACT (d'après Ensembl).

Ces termes sont affichés par ordre de gravité (du plus sévère au moins sévère).

II.5.6.9 Outil de visualisation des variants : IGV

IGV (Integrative Genomics Viewer) est un outil permettant de visualiser les fichiers BAM (Robinson *et al.* 2011). Il présente l'ensemble des reads alignés sur la séquence de référence et les variants associés. La couverture est disponible pour chaque nucléotide. Les reads sens (*forward*) et les reads antisens (*reverse*) se distinguent par des flèches vers la droite ou vers la gauche, respectivement.

Pour chacun des individus analysés au laboratoire de génétique du CHRU de Nancy, une capture d'écran IGV de chaque variant filtré est effectuée à la fois sur le variant de l'individu ainsi qu'au niveau de la position correspondante au niveau du « pool des pères » et du « pool des mères ». Ces trois captures d'écran sont regroupées sous la forme d'un fichier PDF pour chacun des variants. Dans ce fichier, il est noté le chromosome portant le gène, la position chromosomique du variant et la nature de la variation. À côté de chaque capture d'écran on trouve la profondeur au variant, le nombre de reads portant la référence, le nombre de reads portant le variant et éventuellement le nombre de reads portant une autre variation, avec pour chaque donnée le nombre de reads sens et antisens.

La visualisation des fichiers BAM des individus permet de détecter visuellement et d'éliminer certains variants faux positifs. Plusieurs cas de figures sont possibles.

- Reads mal alignées sur la séquence de référence (Figure 16).
- Une séquence répétée ou homopolymère est fréquemment source d'erreur avec la technologie Ion Torrent. En effet, l'augmentation relative de la variation de pH lors de l'incorporation du même nucléotide plusieurs fois de suite n'est pas proportionnelle au nombre de nucléotides incorporés. Reprenons l'exemple du stretch de A de Buermans et den Dunnen (2014) (Buermans et den Dunnen 2014). Par rapport à un seul « A », un allongement de « AA » multiplie le pH par $2/1$ soit par 2. Le passage de « AA » à « AAA » multiplie le pH par $3/2$, soit par 1,5, le passage de « AAA » en « AAAA » par $4/3$, soit 1,3, celui de « AAAA » en « AAAAA » par $5/4$, soit 1,25 et celui de « AAAAA » en « AAAAAA » par $6/5$, soit 1,2 seulement, et ainsi de suite. Plus la longueur de l'homopolymère augmente, plus l'augmentation relative de la variation de pH diminue, diminuant la probabilité d'identification correcte d'une région répétée. (Figure 17).
- Biais de brin : cela se produit lorsque les génotypes sur les brins sens et les brins antisens diffèrent (Guo *et al.* 2012). Par exemple, le variant étudié se situe presque exclusivement sur les brins antisens (seulement sur les brins bleus) et est quasiment absent des brins sens. Dans ce cas, il s'agit d'un biais de brin donc d'un faux positif.

À noter que le nombre exact de reads sens et antisens portant la variation est visible (Figure 18).



FIGURE 16 : Exemple de profil d'un variant faux positif (chr7:g.74146817C>T) visualisé sur IGV avec un mauvais alignement des reads.

Elles correspondent aux barres horizontales rouge pâle (read sens mal alignés) et bleu pâle (reads antisens mal alignés).



**FIGURE 17 : Exemple de profil d'un variant faux positif (chrX:g.148582573GAA>G)
visualisé sur IGV avec un homopolymère de A.**

On observe que la délétion (rectangle blanc barré d'un trait noir) est de taille variable et qu'elle se situe au bout d'une répétition de A.



FIGURE 18 : Exemple de profil d'un variant faux positif (chrX:g.110439122TA>T) visualisé sur IGV avec un biais de brin.

La délétion est présente sur 86 reads dont 86 reads sens et zéro read antisens.

II.6 Détection des CNV

L'analyse des CNV est effectuée lors de la première analyse des résultats à l'aide de l'outil XHMM (eXome Hidden Markov Model) (Fromer et Purcell 2014). Elle ne sera pas détaillée car elle n'est pas effectuée lors de la seconde interprétation des résultats du panel de gènes.

II.7 Analyses bioinformatiques et biologiques du panel de gènes

II.7.1 Première analyse

Il est conseillé d'effectuer deux analyses indépendantes des données brutes avec deux pipelines utilisant des outils et/ou des paramètres différents. La première est effectuée via un pipeline Ion Torrent et la deuxième via un pipeline « maison ».

II.7.1.1 Pipeline Ion Torrent

Ce pipeline est mis au point par le fournisseur du séquenceur. Il s'agit du logiciel Torrent Suite 5.0.3 (ThermoFisher Scientific). L'alignement des séquences sur le génome humain de référence (version GRCh37) s'effectue grâce au module TMap (Torrent Mapper).

L'identification des variants est réalisée grâce au plugin Variant Caller du Torrent Browser et produit un fichier VCF non annoté. L'annotation du VCF est réalisée avec SeattleSeq Annotation 138 (Ng *et al.* 2009) qui ajoute pour chaque variant une description contenant notamment le nom du gène concerné, la nature du variant (non-sens, faux-sens, synonyme, non codants, etc), sa présence ou non dans la base de données dbSNP, sa fréquence dans la base de données ExAC, la prédiction *in silico* par PolyPhen-2 pour les variants faux-sens. Le fichier VCF annoté est converti en fichier tabulé lisible sur Excel via l'outil VCFtoTab-delimited de Galaxy.

Les variants sont filtrés selon les critères suivants :

- nature de la mutation : exclusion des variants de type synonyme, non-codant et intronique (pour conserver les variants non-sens, frameshift et affectant l'épissage) ;
- base dbSNP version 129 : exclusion des variants présents dans la base ;
- PolyPhen-2 : exclusion des variations prédites bénignes ;
- fréquence dans ExAC : exclusion des variants avec une fréquence supérieure à 1 %.

Les variants restants sont priorisés par une analyse manuelle en visualisant les variants dans IGV et en consultant les bases de données OMIM et VarSome.

II.7.1.2 Pipeline maison

Il s'agit d'un pipeline informatique mis au point par le bioinformaticien du service pour l'analyse des données de séquençage. L'étape d'alignement s'effectue sur la version GRCh37 du génome humain de référence à l'aide de l'outil BWA (Burrows-Wheeler Aligner) (Li et Durbin 2009) en utilisant les paramètres par défaut. L'outil SAMtools (Li *et al.* 2009) est utilisé pour compresser le fichier SAM en fichier BAM.

VarScan2 est utilisé pour identifier les variants et produit un fichier au format VCF (Koboldt *et al.* 2012). Il ne garde que les variants avec une profondeur de couverture supérieure à 40X et une fréquence allélique supérieure ou égale à 1 %.

L'annotation des variants est réalisée avec VEP. À la sortie du pipeline, plusieurs informations sont disponibles dans un tableur au format xls :

- variant hérité ou *de novo* et présent ou non dans la base de données ExAC ;
- identifiant du fichier PDF contenant les captures d'écran IGV ;
- nom du gène, chromosome, position génomique ;
- allèle de référence, allèle alternatif et type de variant (SNV, délétion, insertion) ;
- contexte nucléotidique c'est-à-dire 5 pb en amont et 5 pb en aval du variant ;
- présence d'homopolymère au niveau de la séquence concernée ;

- numéros (identifiant dbSNP) ;
- profondeur de couverture au niveau du variant, nombre de reads porteurs de l'allèle de référence et porteurs du variants, et pourcentage de reads porteurs du variant ;
- détail du nombre de reads sens et antisens porteurs ou non du variant (permettant également de détecter un biais de brin éventuel comme expliqué précédemment pour la visualisation sur IGV) ;
- identifiants ESNG et ESNT ;
- positions relatives
 - o de la paire de base au niveau de la séquence d'ADNc (séquence contenant les régions 5' et 3' UTR) ;
 - o de la paire de base au niveau de la séquence codante (CDS) (contenant la séquence qui est traduite en protéine, entre le codon initiateur et le codon stop, donc exclut les régions UTR) ;
 - o de l'acide aminé sur la protéine ;
- acide aminé et codon de référence d'une part, acide aminé et codon alternatifs d'autre part ;
- fréquence allélique du variant dans ExAC ;
- conséquence du variant et IMPACT associé ;
- prédictions *in silico* de SIFT et PolyPhen-2.

Les variants sont enregistrés dans une base de données. Plusieurs filtres sont appliqués l'aide de requêtes SQL et de scripts Perl permettant de conserver seulement les variants répondant aux critères suivants :

- Fréquence allélique des variants de patients supérieure ou égale à 30 % ;
- Fréquence inférieure à 1 % dans la population de la base de données ExAC.

Les variants restants sont priorisés grâce à une analyse « manuelle » à l'aide de l'outil IGV, de la base de données OMIM et de l'outil VarSome.

II.7.2 Réinterprétation des résultats non contributifs

Pour chaque individu ayant un résultat non contributif pour le panel de gènes de DI, une réanalyse manuelle des variants est effectuée à partir des variants issus du pipeline maison listés dans le tableur au format xlsx. Après l'ajout manuel de quelques informations d'annotation, les variants sont réinterprétés. Voici le déroulement de la réanalyse.

II.7.2.1 Visualisation des variants

La réanalyse manuelle débute par une visualisation du fichier PDF comprenant les captures d'écran du variant dans IGV chez l'individu analysé et dans les pools des parents. Les variants faux positifs sont éliminés.

Il est possible d'identifier le mode d'apparition du variant (*de novo* ou hérité) en calculant le nombre de reads correspondant à un seul père ou mère, en supposant la proportion de reads provenant de chaque parent comme homogène. On considère un variant hétérozygote lorsque 30 à 70 % des reads à une position donnée sont porteurs du variant. Si le variant est présent sur un nombre de reads supérieur à 30 % des reads d'un seul parent dans le pool, on peut estimer que le variant est porté par l'un des pères ou mères. Dans ce cas, le variant est considéré hérité du père et/ou de la mère. Dans le cas contraire, le variant est considéré comme *de novo*. Mathématiquement :

- soit P le nombre de parents dans le pool ;
- soit X le nombre total de reads dans le pool ;
- soit V le nombre de reads porteurs du variant dans le pool.

Le variant est considéré comme étant hérité si :

$$V > 0,3 \times \frac{X}{P}$$

II.7.2.2 Collecte manuelle de nouvelles informations d'annotations

- 1) Recherche de la fréquence allélique dans gnomAD.
- 2) Recherche des numéros de transcrits (numéro NM) via le module Position Converter de Mutalyzer.
- 3) Recherche des transcrits Ensembl correspondants (numéro ENST).
- 4) Recherche de la présence du variant dans ClinVar. Si le variant est présent, son interprétation et le trouble concerné sont renseignés.
- 5) Recherche du phénotype associé au gène porteur du variant dans OMIM.

II.7.2.3 Priorisation des variants

II.7.2.3.1 Confrontation de la ségrégation du variant au type de transmission de la pathologie associée au gène porteur du variant

- Pour les troubles à transmission autosomique dominante et dans le cas où les deux parents sont sains, seuls les variants d'apparition *de novo* sont retenus.

Dans le cas où les parents ne sont pas sains, les variants hérités du ou des parents atteints sont retenus.

- Pour les troubles à transmission autosomique récessive, seuls les variants homozygotes ou hétérozygotes composites sont retenus.
- Pour les troubles à transmission dominante liée à l'X et dans le cas où les deux parents sont sains, seuls les variants d'apparition *de novo* sont retenus.

Dans le cas où la mère est atteinte, les variants hérités de la mère sont retenus chez les individus des deux sexes.

Dans le cas où le père est atteint, les variants hérités du père sont retenus chez les individus de sexe féminin.

- Pour les troubles à transmission récessive liée à l'X, les variants dans les gènes concernés par ce mode de transmission sont retenus, uniquement chez les patients de sexe masculin et uniquement s'ils sont d'apparition *de novo* ou hérités de la mère.

II.7.2.3.2 Confrontation du type de transmission du gène porteur du variant et de la fréquence du variant

Les variants sont éliminés si leur fréquence allélique dans gnomAD est supérieure à un seuil fixé. Les seuils utilisés ont été choisis en fonction du mode d'hérédité.

- Pour les variants se trouvant dans des gènes à transmission dominante, autosomique ou liée à l'X, seuls ceux avec une fréquence allélique inférieure à 0,0001 sont retenus. Cela signifie que les variants présents chez plus de 1/10 000 individus sont éliminés.
- Pour les variants se trouvant dans des gènes à transmission autosomique récessive, seuls les variants avec une fréquence allélique inférieure à 0,01 sont retenus. Il faut qu'ils se trouvent à l'état homozygote ou hétérozygote composite chez l'individu testé, avec un variant hérité du père et un de la mère ou avec un allèle hérité de l'un des parents et un allèle *de novo*. Cela signifie que les variants homozygotes ou hétérozygotes composites trouvés chez plus de $1/100 \times 1/100 \times 1/4 = 1/40\ 000$ individus sont éliminés.
- Enfin, dans le cas d'un gène à transmission récessive liée à l'X, le variant n'est retenu que si sa fréquence est inférieure à 0,0001. Cela signifie que les variants trouvés chez plus de 1/10 000 individus sont éliminés.

II.7.2.3.3 Consultation des autres critères

Les variants restants sont conservés en tant que variant candidat ou éliminés selon un faisceau d'arguments :

- IMPACT : l'annotation HIGH est en faveur d'un caractère délétère du variant ;
- Outils de prédiction *in silico* : SIFT et PolyPhen-2. Les annotations « *damaging* » ou « *probably damaging* » fournissent un argument en faveur du caractère délétère du variant ;
- ClinVar : un variant rapporté dans ClinVar et classé comme pathogène ou probablement pathogène pour le trouble qui nous intéresse est un argument en faveur du caractère pathogène du variant ;
- Phénotype OMIM : une corrélation génotype-phénotype importante, est un argument en faveur du caractère délétère du variant. Le spectre phénotypique peut être beaucoup plus large pour un gène que celui actuellement connu et décrit. Une faible corrélation ne suffit pas à elle seule à éliminer le variant.

Le statut de candidat d'un variant ne peut parfois pas être exclu ou confirmé sur la base de ces seuls critères. D'autres recherches sont effectuées au cas par cas dans la littérature, en utilisant VEP, etc.

II.7.2.3.1 Partage de données internationales

La réplification de cas consiste en l'observation de multiples patients non apparentés, mais phénotypiquement similaires, présentant des variants susceptibles dans le même gène. C'est un argument clé dans la validation des variants. La plate-forme web GeneMatcher facilite l'appariement de cas compatibles via la fédération de plusieurs projets. Il s'agit d'un site web en libre accès qui met en contact des scientifiques du monde entier. Il permet de relier les personnes soumettant le même gène d'intérêt en renseignant ou non un ou plusieurs variants. Son but est de permettre d'identifier des cas avec des phénotypes rares ayant des variants dans le même gène candidat (Sobreira *et al.* 2015). Des appels à collaboration ont été soumis dans GeneMatcher pour trouver des cas supplémentaires présentant des variants délétères dans les mêmes gènes et des phénotypes se chevauchant lorsqu'il s'agissait de gènes peu connus.

II.7.2.3.2 Classification des variants candidats

Les variants candidats sont classés selon les recommandations de l'ACMG (Richards *et al.* 2015).

II.8 Validation des variants

Les variants candidats retenus par le biologiste, ainsi que la ségrégation familiale (*de novo* ou hérité des parents), sont systématiquement confirmées par un séquençage de Sanger. Dans un souci d'identitovigilance, cette analyse se fait – pour l'individu et ses parents – sur de l'ADN extrait d'un deuxième prélèvement sanguin indépendant.

II.9 Projet INDEXES : exploration par séquençage exomique des causes moléculaires de déficiences intellectuelles sévère

II.9.1 Présentation générale

Le projet INDEXES (INtellectual Disability Exploration by Exome Sequencing) porte sur l'identification des causes génétiques de DI sévère par séquençage d'exome chez des individus pour lesquels les techniques traditionnelles et le panel de 446 gènes de DI n'étaient pas contributifs. Il s'agit d'un projet monocentrique financé à la hauteur de 30 000 euros dans le cadre d'un CPRC (2017). Ce financement permettra d'analyser trente exomes répartis en dix trios (un individu et ses deux parents), augmentant les chances de mettre en évidence des variants pathogènes, voire de nouveaux gènes et voies moléculaires impliquées dans la DI.

L'objectif est d'identifier des variations *de novo* responsables de DI sévères chez les individus inclus. Si aucune variation *de novo* n'est identifiée, l'analyse sera réalisée selon une hypothèse de transmission autosomique récessive ou lié au chromosome X (Figure 19). Pour se faire, des pipelines vont être créés et pourront être ré-utilisés ultérieurement dans des explorations similaires.

La stratégie en trio consiste à séquençer en parallèle le cas probant et ses parents. Elle est plus coûteuse qu'un séquençage du cas probant seul, mais est appropriée pour mettre en évidence des mutations *de novo*. On détecte, de la même façon, les variants hérités et les variants *de novo*. La majorité des DI pouvant s'expliquer par des mutations *de novo*, il a été choisi l'approche dite elle aussi *de novo* pour l'identification des variants candidats dans ce projet. Cela permet d'augmenter le rendement diagnostique. En effet, dans le cas où aucun des deux parents n'est affecté, cela peut diviser le nombre de variants potentiellement en cause par dix (Harripaul *et al.* 2017).

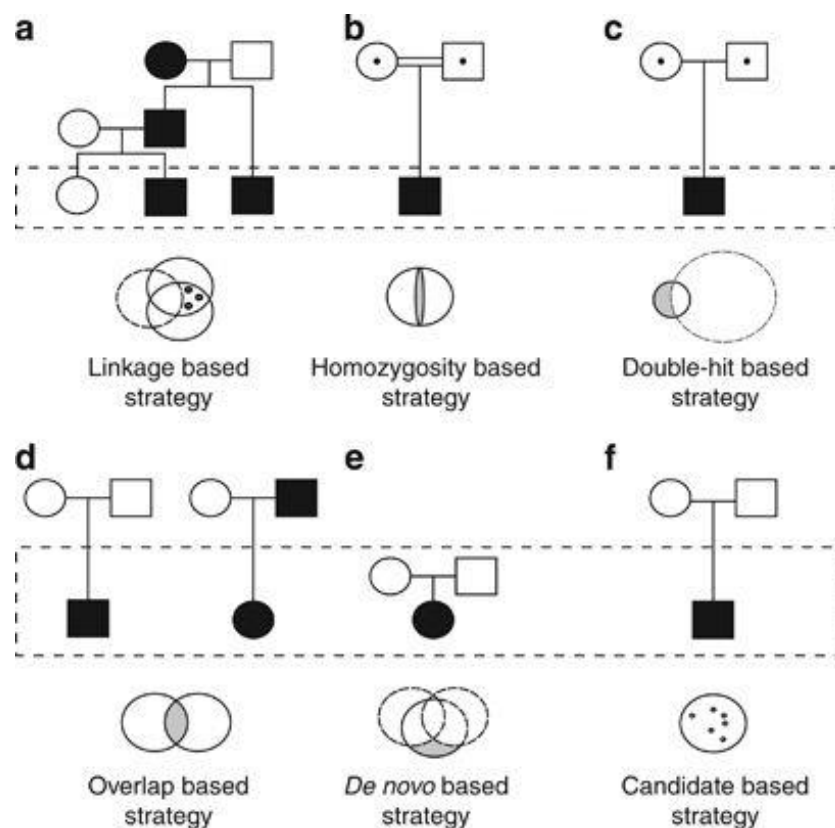


FIGURE 19 : Différentes stratégies utilisées pour identifier des gènes responsables de maladies rares par approche exomique (selon Gilissen et al. 2012).

Les symboles remplis de noir représentent les individus affectés, les symboles vides correspondent aux individus sains et les symboles avec un point illustrent les individus porteurs. Le rectangle en pointillé cerne les individus séquencés dans chaque stratégie. Les cercles sous chaque arbre symbolisent les ensembles de variants identifiés par chacune des stratégies. Les parties grisées correspondent aux variants présents chez les individus atteints, potentiellement pathogènes. Les parties non grisées représentent les variants des apparentés non atteints.

- Stratégie basée sur la ségrégation familiale : lors de troubles monogéniques héréditaires.
- Stratégie basée sur l'homozygotie : lorsque les deux parents sont apparentés et l'individu est atteint d'un trouble récessif rare.
- Stratégie de double-hit : lorsqu'un seul individu est disponible et que l'on suspecte un mode de transmission récessif sans notion de consanguinité.
- Stratégie basée sur le chevauchement : lorsque plusieurs individus issus de familles distinctes, non apparentées, présentent un phénotype similaire.
- Stratégie *de novo* : lors de troubles hétérogènes sur le plan génétique, survenant la plupart du temps *de novo*. D'une part, le nombre de gènes impliqués dans la DI est trop grand pour avoir une chance raisonnable de trouver deux individus dont la DI est

associée au même gène. D'autre part, le nombre de variants par individu est trop important pour mettre en évidence sans le génotype parental le variant *de novo* pathogène. Dans ce cas on utilise un séquençage exomique en trio, permettant de filtrer tous les variants hérités. Il s'agit de la stratégie utilisée dans le projet INDEXES.

- f) Stratégie par gène candidat : lorsqu'aucun échantillon n'est disponible pour les autres membres de la famille.

Ces différentes stratégies peuvent être combinées.

En France, de nombreux laboratoires hospitaliers ont mis en place le séquençage d'exome. Le laboratoire de génétique du CHRU de Nancy possède le matériel et les compétences pour analyser et interpréter les exomes. Ce projet pourrait servir de projet pilote pour le laboratoire et permettre la mise en place du séquençage d'exome en recherche et pour le diagnostic en routine.

II.9.2 Inclusions

L'inclusion se fera au cours d'une consultation de génétique prévue dans le cadre de la prise en charge diagnostique habituelle au CHRU de Nancy. Les critères d'inclusion sont les suivants :

1. DI sévère avec ou sans présentation syndromique associée, mais sans diagnostic évident ;
2. absence d'antécédents familiaux de DI (chez les parents) pour augmenter la probabilité d'identifier des variants *de novo* ;
3. négativité des explorations chromosomiques (caryotype), subchromosomiques (CGH array) et moléculaires (recherche de la maladie de l'X fragile et panel diagnostique de 446 gènes), prévues lors de la prise en charge habituelle des individus atteints de DI ;
4. formulaire de consentement spécifique à l'étude signé ;
5. indication de séquençage exomique retenue par le comité scientifique de génétique.

Si un individu présente une des caractéristiques suivantes, il ne pourra pas être inclus dans l'étude :

1. DI chez l'un des parents car cela est en contradiction avec l'hypothèse privilégiée de mutation *de novo* ;
2. DI laissant présumer une étiologie non génétique (infectieuse, iatrogène, toxique) ;
3. DI avec diagnostic connu ou évocateur pouvant faire l'objet d'une étude moléculaire ciblée ;

4. fausse filiation identifiée lors de la CGH array ou lors de l'analyse du panel de 446 gènes impliqués dans la DI ;
5. quantité ou qualité d'ADN en banque insuffisante pour la réalisation du séquençage exomique ;
6. impossibilité de prélever l'individu ou l'un de ses parents ;
7. refus explicite de participation à l'étude de l'un des parents.

II.9.3 Séquençage et analyse des données

Après inclusion, les aliquots anonymisés des ADN des individus et de leurs deux parents seront transmis à la plateforme IntegraGen pour réaliser le séquençage exomique. Les données génomiques brutes obtenues seront transmises au laboratoire de génétique qui effectuera l'analyse à l'aide de pipelines développés à partir de divers logiciels, algorithmes et de bases de données. Enfin, les variants candidats identifiés seront systématiquement vérifiés par séquençage de Sanger ou par PCR quantitative. Ces résultats seront confirmés sur un second prélèvement effectué lors d'une nouvelle consultation.

III RÉSULTATS

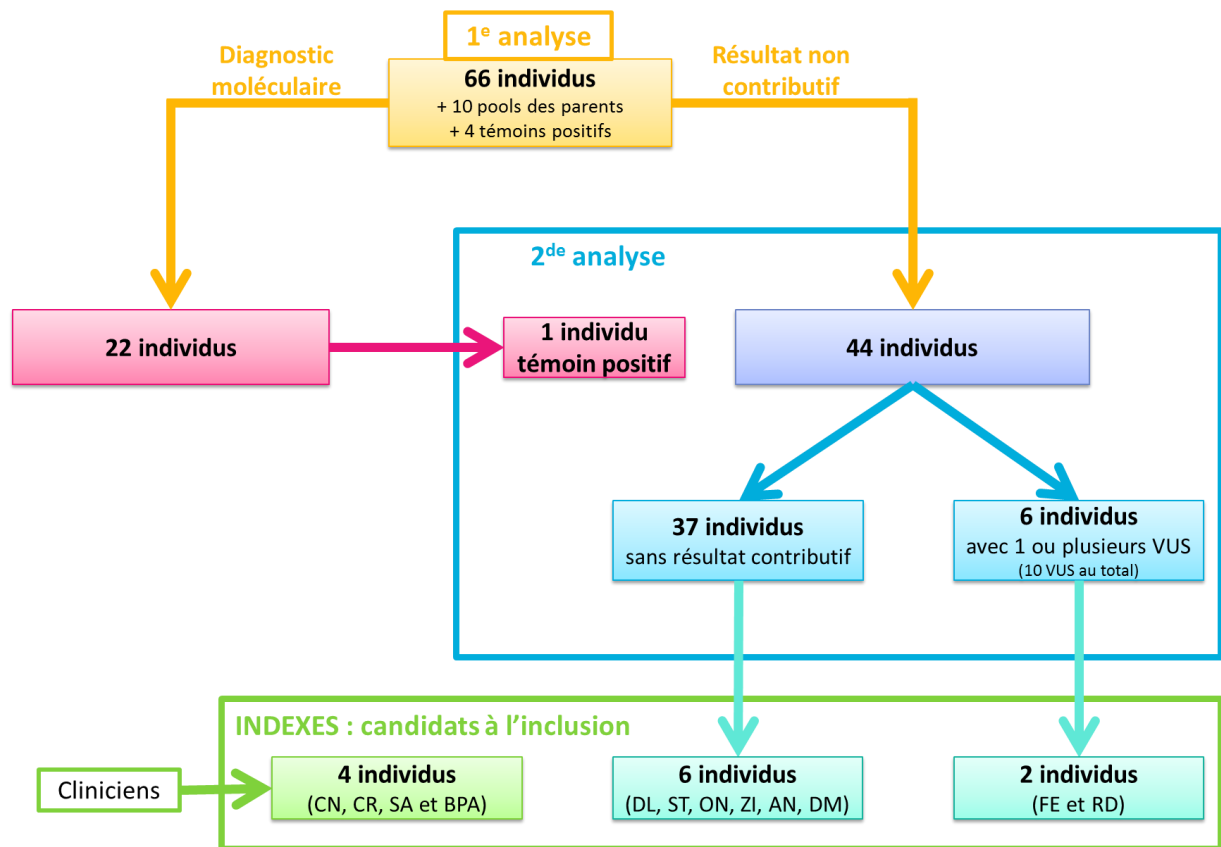


FIGURE 20 : Diagramme de flux des résultats.

III.1 Réanalyse

Au laboratoire de génétique, un panel de 446 gènes de DI a été séquencé chez 66 sujets, 4 témoins positifs et 10 pools de parents.

Un à deux ans après une première analyse non contributive d'un panel de gènes de DI chez 44 des 66 individus, 735 variants ont été réinterprétés, soit en moyenne 17 ($\pm 12,6$) variants par patient, avec un minimum de 5 et un maximum de 81. Aucun nouveau variant délétère ou probablement délétère n'a été identifié. Dix VUS ont été mis en évidence chez six individus (Tableau III).

Identifiant patient	Gène Chromosome	-Notation génomique HGVS -Transcrit -ADnc HGVS -Protéine HGVS	Type	Phénotype individu	gnomAD	ClinVar	-SIFT -PolyPhen-2	Mode de transmission : intitulé du phénotype OMIM	Ségrégation	Sanger	Remarque
FE (1990)	CPA6 Chr. 8	-g.68396042C>T -NM_020361.4 -c.799G>A -p.(Gly267Arg)	Faux sens	DI sévère avec autisme et absence de langage, épilepsie, stéréotypies vocales, difficultés de coordination motrice, apnées avec cyanose buccale (d'origine non épileptique), hypotonie précoce	0,002045	PB (1) P (2) VUS (2)	-D -N	-AR/AD : épilepsie familiale du lobe temporale -AR : épilepsie fébrile familiale	Hérité du père	En cours	Si le séquençage de Sanger confirme la ségrégation, proposer pour inclusion dans INDEXES
		-g.68419039G>C -NM_020361.4 -c.619C>G -p.(Gln207Glu)	Faux sens		0,001412	PB (1) P (2) VUS (3)	-D -N		Hérité du père	En cours	
RD (2012)	TCF4 Chr. 18	-g.52942850_52942851insA -NM_001243235.1 -c.308_309insT -p.(Leu103Phefs*24)	Décalage du cadre de lecture	DI sévère, retard sévère du développement psychomoteur, retard sévère de langage, épilepsie, microcéphalie, dysplasie mineure cérébelleuse inférieure bilatérale, corps calleux hypoplasique, élargissement du système ventriculaire et non visualisation de la commissure blanche antérieure	/	/	-N -N	-AD : syndrome de Pitt Hopkins -AD : dystrophie cornéenne	De novo	Confirmé	Évaluer la corrélation génotype-phénotype avec le clinicien.
GM (2008)	AUTS2 Chr. 7	-g.69364416C>T -NM_001127231.1 -c.454C>T -p.(Arg152*)	Non-sens	DI modérée, microcéphalie (- 2,7 DS), POCS (Pointes Ondes Continues du Sommeil), retard sévère de langage, difficultés de compréhension et de raisonnement, particularités morphologiques (oreilles larges, iris gauche avec anomalie pigmentation, lèvres épaisses, petit pectus excavatum, petit nez bulbeux), difficultés psychomotrice, intolérance à la frustration, troubles du comportement important gênant le quotidien (agitation, hétéro-agressivité, opposition), troubles du sommeil	/	/	-N -N	AD : déficience intellectuelle	/	En cours	/
FC (2004)	SETD2 Chr. 3	-g.47129769A>T -NM_014159.6 -c.5143-32T>A -p.?	Intronique	Dyspraxie, troubles des apprentissages, obésité, asymétrie de vision, macrocéphalie (+ 2 DS), obésité, croissance statur pondérale avancée	0,000033	/	-N -N	AD : syndrome de Luscan-Lumish	De novo	En cours	Demande prélèvement sur tube PAXgene pour analyse du transcriptome car bonne corrélation génotype phénotype.

DA (2006)	ANK3 Chr. 10	-g.61941110C>A -NM_001204404.1 -c.2110G>T -p.(Ala704Ser)	Faux sens	Retard important de langage, TSA majeurs, troubles du sommeil, croissance avancée, macrocéphalie, âge osseux évalué à 10 ans lorsque le patient en avait 8, épilepsie (crises convulsives hyperthermiques), particularités morphologique	0,000325	/	-D -N	AR : déficience intellectuelle	De novo	En cours	Pas d'ADN du père.
		-g.61946485G>T -NM_001204404.1 -c.2022C>A -p.(Asn674Lys)	Faux sens		0,000004	/	-Dfc -PD		De novo	En cours	
	EIF4 Chr. 4	-g.99830840G>A -NM_001331017.1 -c.91C>T -p.(Arg31*)	Non-sens		0,003397	/	-N -N	? ;susceptibilité à l'autisme	De novo	En cours	
MF (2011)	RELN Chr. 7	-g.103143589_103143589insA -NM_005045.3 -c.8362del -p.(Tyr2788Ilefs*40)	Décalage du cadre de lecture	DI modérée, retard de langage, retard des acquisitions, pas d'acquisition de la propreté, troubles alimentaires de type pica avant 3 ans, retard de croissance staturo-pondérale, TSA, marche sur la pointe des pieds, particularités morphologiques	/	/	-N -N	-AR : lissencéphalie type Norman-Roberts -AD : épilepsie familiale du lobe temporal	De novo	En cours	Pas d'ADN du père.
	SHANK1 Chr. 19	-g.51169485T>C -/ -c.506A>G -p.(Tyr169Cys)	Faux sens		0,000519	/	-Dfc -N	/	De novo	En cours	Pas d'ADN du père.

TABLEAU III : Présentation des VUS mis en évidence lors de la réanalyse.

Les VUS nécessitent pour la plupart des analyses complémentaires. Dans certains cas, l'absence d'ADN paternel peut possiblement empêcher de conclure à la pathogénicité du variant.

AD : Autosomique Dominant ; AR : Autosomique Récessif ; Chr. : Chromosome ; D : Délétère ; Dfc : Délétère Faible Confiance ; N : Neutre ; PD : Probablement Délétère ; TSA : Troubles du Spectre Autistique ; / : non disponible.

Le variant causal du témoin positif est mis en évidence. Lors de la première analyse, ce variant a été classé selon un système propre au laboratoire. Après la nouvelle analyse, la classification de l'ACMG est appliquée. En voici le détail.

Il s'agit d'un enfant de 7 ans, adressé au médecin généticien pour une DI sans diagnostic. Il est issu d'un couple sans antécédents familiaux. La grossesse s'est déroulée sans particularité. Il est né à 40 semaines d'aménorrhée avec une bonne adaptation à la vie extra-utérine. Son poids, taille et périmètre crânien à la naissance sont, respectivement, 3660 g (+0,3 DS), 50 cm (+0 DS) et 34 cm (-0,7 DS). Il présente une DI sévère, une épilepsie partielle, une fontanelle antérieure très petite (malgré le périmètre crânien dans les normes), des particularités morphologiques, des anomalies des extrémités (camptodactylie de trois doigts de chaque main, orteils en flexion) et un micropénis. Il est noté également une protrusion de la langue avec bavage excessif, une bronchopathie chronique avec un encombrement bronchique important (toux et hypersalivation), une augmentation de l'écart inter-mamelonnaire, un angiome plan au niveau de l'hémi-abdomen droit (supérieur à 10 cm) et une constipation chronique. L'IRM (Imagerie par Résonance Magnétique) met en évidence une lissencéphalie, une agénésie du corps calleux, des troubles de la migration neuronale et une hypoplasie cérébelleuse. Le fond d'œil montre une atrophie optique, confirmée par IRM. Le caryotype et la CGH array ne sont pas contributifs. Une analyse par FISH (Fluorescent In Situ Hybridization) de la région 17p13 n'identifie aucune anomalie. L'analyse du panel de DI met en évidence un variant faux-sens *de novo* dans le quatrième et dernier exon du gène *TUBB2B* :

- g.3225929C>T ;
- NM_178012.4 :c.394G>A ;
- NP_821080.1 :p.Gly132Ser.

La classification de ce variant est effectuée en se basant sur les critères de l'ACMG rapportés en Annexes 3, 4, 5 et 6. Voici les critères de classification retenus :

- **Critère PS2** (Pathogenic Strong 2) : variant d'apparition *de novo* chez un individu sans antécédent familial, avec confirmation de la paternité et de la maternité. Il s'agit d'une forte preuve de pathogénicité. Dans notre cas, la paternité et la maternité de l'individu sont confirmés par QF-PCR (Quantitative Fluorescence-Polymerase Chain Reaction).
- **Critère PM1** (Pathogenic Moderate 1) : variant localisé dans un hot spot mutationnel et/ou dans un domaine critique et fonctionnel bien établi exempt de variant bénin. Il s'agit d'une preuve modérée de pathogénicité. Ce variant se situe sur un domaine protéique pour lequel de nombreux variants pathogènes sont rapportés dans la

littérature et exempt de variant bénin (Figure 21). Il s'agit du domaine GTPase de la famille Tubuline/FtsZ essentiel à la division cellulaire (Nogales *et al.* 1998).

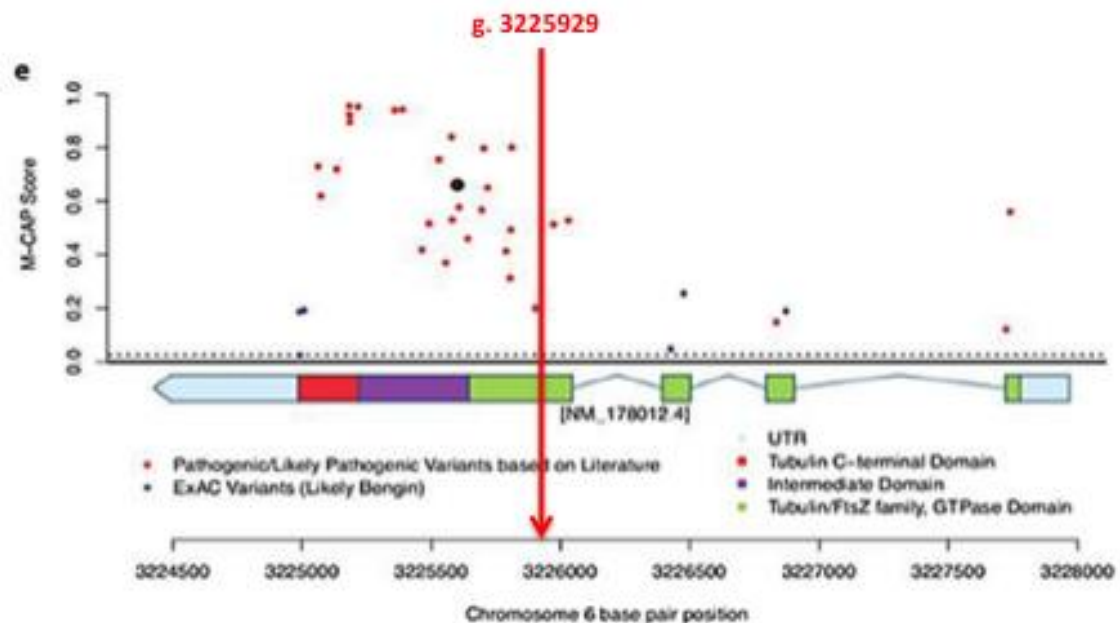


FIGURE 21 : Répartition des variants rapportés sur *TUBB2B* (modifié d'après Geiger *et al.* 2017).

Les positions génomiques des variants sont visibles sur l'axe horizontal en bas de la figure. La flèche rouge correspond au variant identifié au laboratoire de génétique du CHRU de Nancy par l'analyse du panel de 446 gènes de DI.

L'axe des ordonnées représente le M-CAP Score (Mendelian Clinically Applicable Pathogenicity Score). Il s'agit d'un score de prédiction de la pathogénicité des variants faux sens rares. Ce score utilise les scores de différents outils de prédiction *in silico* tels que SIFT, PolyPhen-2, CADD, MutationTaster, MutationAssessor ou FATHMM ainsi que les scores de conservation de différents outils tels que PhastCons et GERP. Un variant est prédit pathogène lorsque son score est supérieur ou égal à 0,025. La sensibilité du M-CAP est de 95 % (Jagadeesh *et al.* 2016).

- **Critère PM2** (Pathogenic Moderate 2) : le variant est absent des bases de données de population contrôle. Il s'agit d'une preuve modérée de pathogénicité. Ce variant est absent d'ExAC, de gnomAD et de 1000 Genomes (Birney et Soranzo 2015).
- **Critère PP3** (Pathogenic Supporting 3) : plusieurs preuves informatiques appuient un effet délétère sur le gène ou le produit génique (conservation, évolution, impact sur l'épissage, etc.). Il s'agit d'une preuve faible de pathogénicité. Ce variant est prédit délétère par de nombreux outils de prédiction *in silico* et de conservation (Tableau IV).

Outils de prédiction <i>in silico</i> et de conservation	Prédiction pour le variant NM_178012.4(TUBB2B):c.394G>A
SIFT	Délétère (faible confiance)
PolyPhen-2	Probablement délétère
PROVEAN	Délétère
MutationTaster	Pathogène
LRT	Délétère
MutationAssessor	Impact fonctionnel (Fort)
CADD phred	Probablement délétère (28)
FATHMM	Délétère
GERP ++ RS	Forte pression conservatrice (5,21)
Meta SVM	Délétère
PhastCons	Très conservé chez les mammifères

TABLEAU IV : Prédictions des différents outils de prédiction *in silico* consultés via VEP pour le variant présent dans TUBB2B.

Pour ce variant, la totalité des outils de prédiction et de conservation consultés sont en faveur du caractère délétère du variant.

- **Critère PP4** (Pathogenic Supporting 4) : le phénotype de l'individu ou l'histoire familiale est hautement spécifique de la pathologie avec une seule étiologie génétique. Il s'agit d'une preuve faible de pathogénicité. Les mutations faux-sens décrites dans ce gène sont associées à des anomalies cérébrales telles qu'une polymicrogyrie, une pachygyrie, une lissencéphalie, une agénésie du corps calleux et un défaut de migration neuronale. Les individus atteints présentent une microcéphalie, une épilepsie, un retard de développement psychomoteur et un retard mental (Fallet-Bianco *et al.* 2014; Guerrini *et al.* 2012; Jaglin *et al.* 2009; Laquerriere *et al.* 2016; Romaniello *et al.* 2012). Les anomalies cérébrales associées à ce gène sont spécifiques et notre individu présente une lissencéphalie, une agénésie du corps calleux et des troubles de la migration neuronale.

À noter que ce variant est rapporté dans la base de données de variants ClinVar (ID559768) comme VUS.

Au total nous avons un critère « *strong* », deux critères « *moderate* » et deux critères « *supporting* ». Ce variant est donc pathogène. Le système de classification de l'ACMG et le système propre au laboratoire aboutissent au même résultat.

III.2 Bilan des inclusions dans le projet INDEXES

Le projet INDEXES doit inclure dix individus et leurs deux parents.

III.2.1 Individus proposés par les cliniciens

À ce jour, quatre individus (CN, CR, SA et BPA) pour lesquels le panel de gènes DI n'a pas été réalisé sont proposés pour inclusion dans le projet par les généticiens cliniciens (Tableau V). L'inclusion de ces quatre individus dépendra du résultat de l'analyse du panel de 446 gènes de DI et des autres explorations manquantes.

III.2.2 Individus à présenter au comité scientifique

La réanalyse non contributive permet d'identifier deux nouveaux individus bons candidats (DL, ST) pour le projet INDEXES (Tableau V). Ils seront proposés au comité scientifique de génétique lors d'une prochaine réunion.

III.2.3 Individus à proposer pour inclusion en fonction des résultats du panel

Le sujet FE est porteur de deux VUS dans le gène *CPA6* dont la confirmation par séquençage de Sanger est en cours. Le sujet RD est porteur d'un VUS identifié dans *TCF4* nécessitant un dialogue clinico-biologique (Tableau V).

Les deux VUS dans *CPA6* concernent un patient (FE) âgé de 28 ans atteint d'une DI sévère avec TSA et absence de langage. Les données néonatales ne sont pas connues. Il présente une épilepsie, des stéréotypies vocales, des difficultés de coordination motrice, une hypotonie précoce ainsi que des apnées avec cyanose buccale d'origine non épileptique. Nous n'avons pas trouvé d'information concernant une éventuelle imagerie.

CPA6 est associé, d'après OMIM, à des épilepsies familiales du lobe temporal de transmission autosomique dominante ou récessive et à des épilepsies fébriles familiales de transmission autosomique récessive. Il n'y a pas d'autres troubles associés rapportés dans la littérature en dehors de quelques anomalies cérébrales (Belhedi et al. 2014; Salzmann et al. 2012; Sapio et al. 2012).

Ces deux variants dans *CPA6*, NM_020361.4:c.799G>A p.Gly267Arg et NM_020361.4:c.619C>G p.Gln207Glu, sont des variants faux-sens à l'état hétérozygote. Leurs fréquences respectives dans gnomAD sont 0,002045 et 0,001412. Sous l'hypothèse d'une transmission autosomique dominante, ces deux variants auraient été éliminés sur la base de leur fréquence. Cependant, une transmission récessive est possible. Leur impact sur les protéines est prédit neutre par PolyPhen-2 et délétère par SIFT. Leur interprétation est débattue dans ClinVar (pathogène, VUS ou probablement bénin). Le premier variant est

observé dans le pool des pères dans un nombre de reads inférieur au seuil à partir duquel l'on considère qu'un des pères est porteur hétérozygote (13 reads attendus contre 10 observés). Cette valeur étant relativement proche du seuil, la transmission paternelle n'est pas exclue et doit être recherchée par séquençage de Sanger. Le second variant apparaît comme étant hérité du père. Si le séquençage de Sanger indique que l'un ou les deux variants sont d'apparition *de novo*, ils pourraient être à l'origine de l'épilepsie de l'individu. Dans le cas contraire, si le séquençage de Sanger confirme que les deux variants sont hérités du père sain, ils pourront être écartés. Le panel sera alors considéré comme non contributif et le patient pourra être proposé à l'inclusion dans INDEXES.

Le VUS dans *TCF4* concerne un patient (FD) âgé de 6 ans, sans antécédents familiaux, atteint de DI sévère avec retard global du développement et des acquisitions. Les données néonatales ne sont pas connues. Il présente un retard de développement psychomoteur sévère avec un retard sévère de langage, une épilepsie, une microcéphalie, une dysplasie mineure cérébelleuse inférieure bilatérale, un corps calleux hypoplasique, un élargissement du système ventriculaire et une non visualisation de la commissure blanche antérieure. La recherche de la maladie de l'X fragile et le profil de méthylation de la région 15q11q13 ne mettent aucune anomalie en évidence. La CGH array identifie une délétion en 14q31.3q32.11 héritée de la mère et a priori sans lien avec les troubles de l'individu.

Le VUS présent chez cet enfant est un variant frameshift à l'état hétérozygote entraînant un codon stop prématuré, NM_001243235.1:c.308_309insT p.(Leu103Phefs*24), d'apparition *de novo* confirmée par séquençage de Sanger. Il se trouve dans l'exon 3/13 du gène (Figure 22). Il n'est pas dans gnomAD. *TCF4* est impliqué dans le syndrome de Pitt-Hopkins, de transmission autosomique dominante, se caractérisant par une DI et un retard de langage sévère, un retard moteur avec incoordination motrice et ataxie, une épilepsie, des stéréotypies, des anomalies des mains, des anomalies respiratoires de type hyperventilation et apnée et des particularités morphologiques caractéristiques (Marangi *et al.* 2011). La corrélation génotype-phénotype sera prochainement discutée avec le généticien clinicien prenant en charge le patient. Si le variant n'est pas retenu, l'inclusion dans INDEXES est possible. Si ce variant est retenu comme étant associé à la symptomatologie de l'enfant, la réanalyse aura permis d'apporter un diagnostic pour un des 44 cas étudiés.

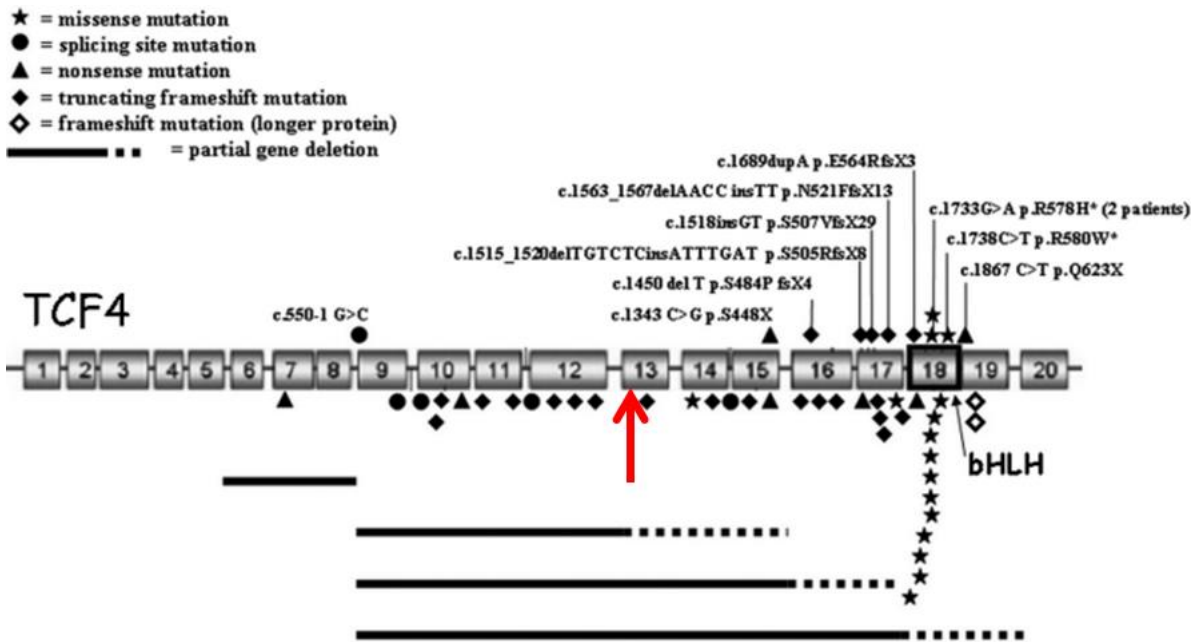


FIGURE 22 : Représentation des variants associés au syndrome de Pitt-Hopkins.

Les variants de la série de Marangi *et al.* 2011 se situent au-dessus de la représentation du gène *TCF4*. Les variants rapportés dans la littérature sont sous la représentation de *TCF4*.

La flèche rouge correspond au variant identifié au laboratoire de génétique du CHRU de Nancy par la réanalyse du panel de 446 gènes de DI (modifié d'après Marangi *et al.* 2011).

III.2.4 Individus à proposer pour inclusion en fonction du phénotype

Quatre individus (ON, ZI, AN et DM) avec une réanalyse non contributive sont de possibles candidats pour INDEXES mais les informations phénotypiques dont nous disposons pour le moment sont insuffisantes. Dans certains cas, la sévérité de la DI n'est pas connue. Chaque phénotype doit être discuté avec les cliniciens afin d'être réévalué (Tableau V). L'inclusion de ces quatre individus dans le projet INDEXES dépendra du tableau phénotypique et des résultats des examens nécessaires à l'inclusion.

Au total, il y a à ce jour, huit candidats pour l'inclusion dans le projet INDEXES (Tableau V) et quatre, pour lesquels le panel est non contributif, devant être discutés avec les cliniciens pour étayer le tableau clinique (Tableau V).

Patient (année de naissance)	Indication	Caryotype	X fragile	CGH array	Autres	Panel de gènes DI	Réserves pour inclusion
CN (2006)	DI sévère, TSA, microcéphalie -6 DS Seul cas familial	46,XY	Non réalisé	Non contributive	Étude de la région 15q11q13 et FISH 8p sans particularité Étude <i>UBE3A</i> en cours Panel microcéphalie primitive non contributif	Non réalisé	Panel de gènes DI non contributif
CR (2006)	Encéphalopathie épileptique sévère, régression des acquisitions à partir de l'âge de 2 ans, troubles du comportement (cris, vocalisations, agitation), hypertrophie gingivale localisée Frère de 30 ans en bonne santé avec projet parental Patient décédé subitement par arrêt cardio-respiratoire	À faire	Non réalisé	Non contributive	Panel de 71 gènes d'épilepsies génétiques non contributif	Non réalisé	Panel de gènes DI non contributif
SA (2007)	Encéphalopathie épileptique sévère, syndrome de West, hétérotopie périventriculaire, DI profonde, pas de langage, pas de marche, hypotonie	46,XY	Non réalisé	Délétion en 18q21.33 <i>de novo</i> , contenant <i>KIAA1468</i> gène peu décrit à l'heure actuelle Duplication interstitielle en 22q11.21, hérité du père, n'explique pas le phénotype	Bilan biochimique, FISH del1p36 et 5qter sans particularité Étude de <i>TSC1</i> , <i>ARX</i> , <i>MECP2</i> , <i>CDKL5</i> et <i>FGFR3</i> sans particularité, étude de <i>FNL1</i> en cours Étude de <i>TSC2</i> : variant à la position 1839+18G>A (hérité du père asymptomatique)	Non réalisé	Panel de gènes DI non contributif
BPA (2012)	Naissance à terme sur RCIU sévère, retard staturo-pondéral et psychomoteur, ne tient pas debout, ne parle pas, rachitique, anomalie cérébrale de type hypoplasie du mésencéphale, agénésie du corps calleux, malformation cardiaque	À faire	Non réalisé	Non contributive	Recherche d'unidisomies des chromosomes 7, 11 et 14 et recherche trisomie 8 en mosaïque sans particularité	Non réalisé	Panel de gènes DI non contributif
DL (2010)	Épilepsie, déficit sévère du langage (jamais acquis), retard des acquisitions, macrocrânie (sans macrocrânie familiale), dysplasie trigono-calleuse, retard de myélinisation, hypotonie, troubles vasomoteurs des extrémités, ombilication du mamelon, tâche café au lait de 1 cm sur la cuisse gauche, embonpoint abdominal en lien avec la prise de Micropakine.	46,XX (pas de chromosome 20 en anneau)	Normal	Non contributive	Méthylation 15q11q13, étude de <i>MECP2</i> et bilan biochimique sans particularité Étude <i>NSD1</i> en cours	Non contributif	/
ST (1992)	DI sévère, retard psychomoteur, malformations (rénales, cérébrales, cardiaques, osseuses), pectus excavatum, retard croissance staturo-pondérale, hernies, <i>fetal pads</i> , macrocéphalie, particularités morphologiques (face triangulaire allongée, enophtalmie, fentes palpébrales bas dehors, bout du nez large, racines hautes, oreilles courtes longues postérieures basses, hypertélorisme, palais ogival, micrognathie, philtrum profond, arc de cupidon marqué, ptosis discret), mal implantation des dents, hyperlaxité des mains, flessum des genoux, ectopie bilatérale, hydrocèle bilatérale, 2 hernies inguinales, deux tâches dépigmentées, (marche à 3 ans), myopie, hypermétropie. Parents apparentés au 1/16ème	46,XY	Normal	Délétion interstitielle 4q12 héritée de la mère, n'explique pas le phénotype	Recherche de la maladie de Steinert, FISH 12p, étude du gène <i>PTPN11</i> sans particularité	Non contributif	/

FE (1990)	DI sévère avec TSA et absence de langage, épilepsie, stéréotypies vocales, difficultés de coordination motrice, apnées avec cyanose buccale (d'origine non épileptique), hypotonie précoce	46,XY	Non réalisé	Non contributive	/	2 variants dans <i>CPA6</i> , hérités du père, Sanger en cours	Présentation si Sanger confirme la ségrégation des variants dans <i>CPA6</i>
RD (2012)	DI sévère, retard psychomoteur sévère, retard global de développement et des acquisitions, épilepsie, microcéphalie	À faire	Normal	Délétion 14q31.3q32.11, héritée de la mère, n'explique pas le phénotype	Méthylation 15q11q13 et bilan biochimique sans particularité	VUS <i>TCF4 de novo</i>	VUS dans <i>TCF4</i> peu probable mais à confirmer avec généticien clinique
ON (1991)	Retard psychomoteur, premiers mots vers 7 ans, troubles du comportement, IMC à + 3,41 DS, hirsutisme/hypertrichose, dyskinésies, crispation/douleurs des mains, strabisme, hallux valgus infraductus bilatéral	Non réalisé	Non réalisé	Non contributive	Étude de l'axe corticotrope sans particularité, absence d'hyperandrogénie	Non contributif	Discuter du phénotype
ZI (2005)	Retard de langage, TSA	46,XY	Normal	Délétion en 14q31.3 ne contenant aucun gène décrit à ce jour	/	Non contributif	Discuter du phénotype
AN (2006)	TSA, obésité, retard langage, difficultés praxiques et motrices, HTA précoce, oreilles mal ourlées, dermatose pileaire, otites à répétition	Non réalisé	Normal	Non contributive	/	Non contributif	Discuter du phénotype
DM (2012)	Retard de langage, TSA, hyperactivité, troubles de l'interaction et de la communication.	Non réalisé	Normal	Non contributive	Méthylation 15q11q13 sans particularité	Non contributif	Discuter du phénotype

TABLEAU V : Individus candidats pour le projet INDEXES.

Les individus CN, CR, SA et BPA sont des candidats proposés pour inclusion dans le projet par les généticiens cliniciens. Les individus DL et ST sont de bons candidats identifiés suite à la réanalyse non contributive. L'inclusion de l'individu FE dans le projet INDEXES nécessite la confirmation préalable par séquençage de Sanger de la ségrégation des deux variants dans *CPA6*. L'inclusion de l'individu RD dépend quant à elle de l'issue des discussions entre le clinicien et les biologistes concernant le variant dans *TCF4*. Les individus ON, ZI, AN et DM sont des candidats identifiés suite à la réanalyse non contributive mais nécessitant une discussion avec les cliniciens pour étayer le tableau clinique. Pour l'ensemble des individus, les résultats des explorations manquantes d'anomalies chromosomiques (caryotype), subchromosomiques (CGH array) et moléculaires (recherche de la maladie de l'X fragile et panel diagnostique de 446 gènes de DI) et leurs résultats conditionneront l'inclusion.

RCIU : Retard de Croissance Intra Utérin ; TSA : Troubles du Spectre Autistique ; HTA : HyperTension Artérielle ; IMC : Indice de Masse Corporelle ; ORL : Oto-Rhino-Laryngologique.

IV DISCUSSION ET PERSPECTIVES

IV.1 Discussion

IV.1.1 Évolution rapide des connaissances

Ces dernières années, les connaissances des gènes impliqués dans les maladies mendéliennes ont augmenté rapidement. De nouveaux gènes sont constamment découverts et les associations gène-pathologie et variant-pathologie se précisent. En octobre 2004, OMIM répertoriait 1636 troubles mendéliens (Wenger *et al.* 2017). En août 2018, OMIM en listait 8652, dont 5309 avec une association génotype-phénotype connue, ce qui illustre l'avancée rapide des connaissances génétiques.

Dans la plupart des études traitant d'une réanalyse de données de NGS, la résolution de la majorité des cas est consécutive à des publications récentes (Bowling *et al.* 2017; Wenger *et al.* 2017; Wright *et al.* 2018). Dans l'étude de Smith *et al.* réalisée en 2017, une seconde analyse des données de séquençage exomique permet de reclasser 96 cas, dont 73 % grâce aux découvertes de nouveaux gènes publiées dans la littérature (Smith *et al.* 2017).

Wenger *et al.* illustre bien l'intérêt d'une réanalyse des données de NGS en décrivant un individu pour lequel la réanalyse des données d'exome permet d'identifier un variant faux-sens *de novo* rare dans *KMT2A*. La première analyse conclut à un résultat non informatif en juillet 2012, date à laquelle le gène est très peu connu. Deux semaines plus tard, la première publication sur *KMT2A* apparaît et, en juin 2015, le même variant est identifié comme probablement pathogène chez un autre individu. Sans une nouvelle analyse, cet individu n'aurait toujours pas de diagnostic moléculaire.

Le panel de gènes de DI ayant fait l'objet de notre réanalyse comporte un sous-ensemble de gènes dont l'association génotype-phénotype était mal connue au moment de la première interprétation du panel. Depuis lors, de nouvelles connaissances acquises sur l'ensemble des gènes du panel – en particulier sur les gènes du domaine de la recherche – auraient pu permettre de résoudre de nouveaux cas.

IV.1.2 Difficultés et erreurs d'interprétation

Amendola *et al.* démontre en 2016 une discordance intra-laboratoire de 21 % dans l'interprétation de variants en comparant deux méthodes d'interprétations : celle propre à chaque laboratoire et celle recommandée par l'ACMG. Vingt-quatre pourcent des variants classés différemment étaient susceptibles d'entraîner des différences de prise en charge médicale, c'est-à-dire que les variants étaient classés comme pathogènes ou probablement pathogènes par une méthode et VUS, probablement bénins ou bénins par l'autre méthode.

En inter-laboratoire, sur 97 variants analysés, la discordance est passée de 66 % à 29 % après des efforts de consensus (Amendola *et al.* 2016). Ces discordances reflètent la difficulté d'interprétation des variants, même en suivant les recommandations de l'ACMG. En effet, la pathogénicité des variants est souvent inconnue et il est nécessaire d'évaluer un grand nombre de critères hétérogènes pour leur interprétation. Cela peut conduire à une variabilité de l'interprétation en fonction de la personne. Des erreurs d'interprétation sont possibles comme illustré par Nambot *et al.* et par Al-Mushedi *et al.*, en 2018. Nambot *et al.* obtient, lors d'une réanalyse de données de 156 exomes, quatre diagnostics positifs supplémentaires découlant d'une mauvaise interprétation initiale (Nambot *et al.* 2018). Al-Mushedi *et al.* présente deux variants interprétés comme pathogènes après réanalyse dans deux gènes déjà connus au moment de l'analyse primaire. Il conclut donc à une interprétation initiale inappropriée.

Puisque les recommandations de l'ACMG n'étaient encore pas utilisées au laboratoire lors de la première analyse et que la seconde a été effectuée par des biologistes différents, le rendement diagnostique du panel de gènes de DI aurait pu être augmenté par la réinterprétation des données.

IV.1.3 Utilisation du partage de données international

Dans la publication de Nambot *et al.*, en 2018, 50 % des diagnostics nouvellement positifs sont résolus lors de la réanalyse de données de WES grâce à une collaboration avec une autre équipe de scientifiques par l'intermédiaire d'une des plateformes internationales de partage de données (Nambot *et al.* 2018). Ces dernières peuvent aider à résoudre des cas complexes comme par exemple des cas de digénisme ou des cas de phénotypes atypiques permettant de constater que le spectre phénotypique associé à un gène est plus large que celui qui est connu.

Dans notre cas, des variants dans des gènes peu décrits ont été soumis dans GeneMatcher. Malheureusement, aucun nouveau cas n'a été résolu grâce à cet outil.

Plusieurs hypothèses peuvent expliquer l'absence de résolution de cas supplémentaires lors de la réanalyse.

IV.1.4 Utilisation du même pipeline

Le pipeline utilisé n'a pas été changé ou mis à jour. Les données utilisées pour la réanalyse sont celles générées par le pipeline maison auxquelles certaines annotations sont ajoutées. Cet ajout a permis une mise à jour des fréquences de populations et de données provenant

de ClinVar et d'OMIM mais semble insuffisant (Hiatt *et al.* 2018). Selon les outils utilisés lors des étapes importantes du processus bioinformatique – l'alignement, l'identification des variants et l'annotation – il existe des différences sur les variants identifiés, ce qui témoigne de la sensibilité variable de la technique (Cornish et Guda 2015; McCarthy *et al.* 2014). En 2018, Wright *et al.* augmente son rendement diagnostique de 13 points après réanalyse – celui-ci passant de 27 % à 40 % – et affirme que 23 % des cas nouvellement résolus font suite à une amélioration du pipeline d'analyse (Wright *et al.* 2018). Ewans *et al.*, décrit en 2018 un variant dans *PRPF31* – gène impliqué dans les rétinites pigmentaires – qui était absent du VCF lors de l'analyse initiale des données de WES et présent sur le VCF de la réanalyse un an plus tard. Cette différence s'explique par une amélioration du pipeline bioinformatique utilisé, avec une sensibilité accrue lors de l'étape de détection des variants. Dans notre cas, les données brutes étaient archivées et n'ont pu être récupérées qu'à la fin de ce travail. C'est pourquoi, il n'était pas possible d'utiliser un autre pipeline ou de mettre à jour celui existant.

IV.1.5 Limites des paramètres utilisés

IV.1.5.1.1 *Filtres appliqués*

Les paramètres utilisés par le pipeline maison présentent des limites pouvant être responsables de faux positifs ou de faux négatifs.

Les variants avec une fréquence dans la population supérieure à 1 % ne sont pas conservés. Ce filtrage est réalisé en utilisant ExAC et peut être une source de faux négatifs. En effet, Li *et al.* ont montré que la déficience intellectuelle et la schizophrénie partageaient des gènes et voies moléculaires communes (Li *et al.* 2016). Plus de 20 % des données de cette base concernent des individus atteints de schizophrénie. Certains variants pathogènes pourraient se révéler présents dans cette base de données et être éliminés à tort.

L'utilisation d'un nouveau pipeline avec des outils différents de ceux du pipeline maison aurait été judicieuse pour la réanalyse des données de séquençage du panel de gènes de DI.

IV.1.5.1.2 *Pool des parents*

L'utilisation du pool des parents est efficace si l'on considère l'hypothèse selon laquelle les données de séquençage issues de chaque parent y sont présentes en quantités équivalentes. Il est cependant possible que les séquences de certains parents soient sous-représentées avec un risque de conclure à tort qu'un variant est d'apparition *de novo* chez le

cas probant. La confirmation par séquençage de Sanger permet de confirmer la ségrégation familiale mais est limitée aux variants candidats.

IV.1.5.1.3 Inactivation de l'X

Pour les troubles à transmission dominante liée à l'X, seuls les variants d'apparition *de novo* sont retenus lors de la réanalyse si les deux parents sont sains. Cependant, le cas d'une transmission maternelle d'un chromosome X inactivé chez la mère et porteur d'un variant délétère ne peut être exclue.

IV.1.5.1.4 Pathologies à transmission récessive liée à l'X

Pour les troubles à transmission récessive liée à l'X, seuls les variants chez les patients de sexe masculin et d'apparition *de novo* ou hérités de la mère sont retenus lors de la réanalyse. Cependant, un individu de sexe féminin exprimant une symptomatologie mineure par rapport aux individus de sexe masculin ne peut être exclue.

IV.1.6 Mosaïque

La présence d'une mosaïque chez l'individu analysé peut être à l'origine de faux négatifs. Un variant en mosaïque survient à un stade post-zygotique. Il est présent dans certaines cellules seulement. Les cellules sanguines peuvent ne pas être porteuses de ce variant, auquel cas l'ADN séquencé – extrait à partir d'un échantillon sanguin – n'est pas porteur du variant. Un tel variant ne peut pas être détecté dans cette étude et serait un faux négatif s'il jouait un rôle dans la DI de l'individu testé.

De la même manière, un variant peut être identifié au niveau des lymphocytes et être absent des cellules cérébrales entraînant un faux positif.

Une mosaïque peut également être présente chez les parents et entraîner des faux négatifs, dans le cas d'un parent asymptomatique porteur d'un variant au niveau des lymphocytes mais pas au niveau cérébral par exemple, ou des faux positifs dans le cas contraire.

IV.1.7 La réanalyse des panels dans la littérature

À ce jour, il n'existe pas de données de réanalyse de panel de gènes dans la littérature mais plusieurs publications récentes concernant des séquençages d'exomes et de génomes sont disponibles (Alfares *et al.* 2018; Costain *et al.* 2018). Contrairement aux panels – rapidement obsolètes avec la découverte rapide de nouveaux gènes – le WES analyse simultanément tous les gènes codants des protéines, incluant ainsi les gènes impliqués dans la DI encore inconnus à ce jour. Dans un panel, le gène responsable de la DI peut ne pas se trouver parmi les gènes inclus. Dans ce cas de figure, une réanalyse ne permettra pas d'identifier la cause moléculaire de la DI de l'individu. Une nouvelle analyse des données de WES a une

probabilité plus importante de résoudre de nouveaux cas. Dans la DI, le taux diagnostique d'une analyse en panel de gènes et en séquençage d'exome est comparable et oscille entre 25 et 30 % (Hamdan *et al.* 2014; Monroe *et al.* 2016; Redin *et al.* 2014; Vissers, Gilissen, and Veltman 2016; Wright *et al.* 2018). Les études de réanalyse de WES présentent un rendement diagnostique de 10 à 15 % des cas réanalysés ce qui porte le rendement total entre 30 et 40 % (Al-Nabhani *et al.* 2018; Ewans *et al.* 2018; Nambot *et al.* 2018; Wenger *et al.* 2017; Wright *et al.* 2018). L'approche par séquençage d'exome est plus coûteuse que le panel, mais permet de réutiliser ses données au fil des années et d'augmenter ainsi son rendement diagnostique. Une explication majeure de l'absence de résolution de nouveaux cas à la suite d'une nouvelle analyse des données du panel de gènes de DI pourrait être le nombre limité de gènes qu'il contient.

IV.1.8 Délai entre les deux analyses

Le rythme d'apparition de nouvelles publications concernant la déficience intellectuelle est rapide avec, en moyenne, 250 nouvelles associations génotype-phénotype et 9200 associations variants-phénotype rapportées chaque année (Wenger *et al.* 2017). Au fil du temps, la communauté scientifique apporte de nouvelles informations. Ces dernières sont à l'origine de la plupart des cas solutionnés lors de réinterprétations, justifiant un délai minimal entre deux analyses (Bowling *et al.* 2017; Wenger *et al.* 2017). Dans l'étude de Wright *et al.*, 2018, les données issues de la littérature ont permis de résoudre plus des deux tiers des cas réanalysés (Wright *et al.* 2018). Il n'existe pas de consensus sur le délai d'attente entre deux analyses. Cependant, les réanalyses de données d'exomes dans la littérature s'effectuent, la plupart du temps, après un délai de 12 mois minimum (Nambot *et al.* 2018; Wenger *et al.* 2017; Wright *et al.* 2018). Hiatt *et al.* confirme en 2018 que plus l'écart de temps entre deux analyses est long, plus le nombre de variants pathogènes ou probablement pathogènes identifiés est important. Les réanalyses effectuées dans la même année que l'analyse initiale n'identifient que 1 % de nouveaux cas par rapport à l'ensemble des cas réanalysés. Celles effectuées 12 à 23 mois, 24 à 35 mois ou à plus de 36 mois de la première identifient respectivement 2,6 %, 13,2 % et 21,7 % de nouveaux cas par rapport à l'ensemble des cas réanalysés (Hiatt *et al.* 2018). La nouvelle analyse des données du panel de gènes de DI est effectuée un à deux ans après la première. Ce délai était peut-être insuffisant pour que de nouvelles informations sur les gènes du panel soient disponibles dans la littérature et les bases de données. Un intervalle plus important entre les deux analyses est peut-être nécessaire pour résoudre de nouveaux cas.

IV.1.9 Importance des données phénotypiques

Pour des analyses de NGS dans le cadre de la DI, si le laboratoire ne dispose pas du phénotype de l'individu analysé, les chances de prioriser le variant causal sont réduites. La participation des cliniciens s'avère primordiale. Ce fait est souligné par Baldrige *et al.*, en 2017, qui indique que le partenariat du clinicien avec le laboratoire peut augmenter le rendement diagnostique de WES de 7 % (Baldrige *et al.* 2017). En effet, le généticien clinique a une connaissance directe et détaillée du phénotype des individus qu'il a reçus en consultation. Il est au fait des examens complémentaires effectués et de la ségrégation du trouble dans la famille. Ces informations ne sont pas toujours communiquées en totalité au laboratoire ou facilement accessibles. Par exemple, l'appréciation des particularités morphologiques par le biologiste est complexe en l'absence de photographies de l'individu concerné. Le dialogue clinico-biologique est donc essentiel pour la résolution de nombreux cas. Pour la réanalyse du panel de gènes de DI, les données phénotypiques insuffisantes ont été complétées par les informations trouvées dans le dossier du patient. Cependant, les données phénotypiques disponibles ne sont peut-être pas exhaustives, notamment dans le cas de symptômes d'apparition tardive, pouvant conduire à des faux négatifs. Une mise à jour des données phénotypiques avec l'aide du généticien clinique avant la réanalyse aurait pu aider à la priorisation des variants.

Le phénotypage inversé est souvent utilisé lorsqu'un variant candidat est identifié dans un gène pour lequel le phénotype associé ne correspond pas entièrement à celui de l'individu analysé. Il consiste à réexaminer l'individu pour comparer son phénotype à ceux décrits dans la littérature et nécessite un partenariat avec le généticien. Voici l'exemple d'un variant mis en évidence lors de la réanalyse des panels de DI, illustrant l'importance du dialogue clinico-biologique pour la classification des variants. L'individu concerné est un homme de 25 ans, né de parents consanguins au 1/16e, sans antécédent familial notable. Les données néonatales ne sont pas connues. Il présente un retard de développement psychomoteur avec une absence de lecture, un retard de langage, un retard de la marche (acquise vers 3 ans) et aucune autonomie pour l'habillement et la toilette. Il a des *fetal pads*, une cardiomyopathie hypertrophique biventriculaire, des particularités morphologiques cérébrales discrètes, une pathologie kystique rénale, une ectopie testiculaire bilatérale avec hydrocèle, deux hernies inguinales, une hernie ombilicale, un pectus excavatum modéré, des atteintes osseuses à type de chondrodysplasie avec polykystose et un retard de croissance staturo-pondérale. Des particularités morphologiques sont identifiées : une face triangulaire allongée, des oreilles postérieures basses, une enophtalmie avec hypertélorisme, des fentes palpébrales orientées en bas et en dehors, un discret ptosis, un nez à bout large et à racines

hautes, un philtrum court, une micrognathie, une mauvaise implantation des dents, une mauvaise occlusion dentaire et un palais ogival. Il a des fausses routes, un défaut de mastication, une hyperlaxité des mains, un flessum des genoux, une vessie hyperactive, deux tâches dépigmentées, une myopie. Le caryotype, la recherche de la maladie de l'X fragile, la recherche de la maladie de Steinert et l'analyse du gène *PTPN11* n'ont pas mis en évidence d'anomalie. La CGH array identifie une délétion interstitielle en 4q12, héritée de la mère et sans lien apparent avec les troubles de l'individu.

Lors de la réanalyse, un variant faux sens *de novo* (NM_001079846.1:c.5150G>A p.(Ser1717Asn)) dans le gène *CREBBP*, à l'état hétérozygote est retenu (Figure 23). Sa fréquence dans gnomAD est de 0,000030 (soit 1 personne porteuse du variant sur 33 000 individus). Il n'est pas décrit dans ClinVar. Il est prédit comme neutre ou bénin par SIFT et PolyPhen-2 respectivement. Il est prédit comme délétère et probablement délétère par d'autres outils de prédiction *in silico* comme MutationTaster et LoFtool. Ce variant se situe sur un domaine protéique doigt de zinc de type ZZ. *CREBBP* régule la transcription de certains gènes sensibles à l'AMPc (Adénosine MonoPhosphate cyclique) grâce à l'activation de certains facteurs de transcription (Kwok *et al.* 1994; Nordheim 1994). Il est impliqué dans le syndrome de Rubinstein-Taybi qui se transmet sur un mode autosomique dominant. Ce syndrome associe de multiples anomalies congénitales telles qu'une déficience intellectuelle avec retard de langage, un retard de croissance postnatal, une microcéphalie, des pouces et gros orteils larges et hyperlaxes, diverses malformations cardiaques, des *fetal pads*, des anomalies dentaires, des problèmes articulaires, des infections respiratoires fréquentes, des déformations du sternum et des anomalies oculaires (ptosis, glaucome, anomalies de la réfraction). Le visage des individus atteints présente des particularités morphologiques comme des sourcils arqués, de longs cils, des fentes palpébrales obliques en bas et en dehors, un hypertélorisme, une arête nasale large, un palais ogival et une légère micrognathie. Ces individus ont souvent une expression faciale caractéristique. Il existe avec ce syndrome un risque accru de développer des tumeurs bénignes et malignes, principalement des tumeurs cérébrales et des leucémies (Hennekam 2006; Rubinstein et Taybi 1963; Stevens, Pouncey, et Knowles 2011). Compte-tenu de la corrélation génotype-phénotype, ce variant est retenu comme variant candidat et discuté avec le généticien clinicien en charge du patient. La confrontation des données clinico-biologiques permet d'écarter ce variant au regard de la spécificité phénotypique du syndrome de Rubinstein-Taybi. Le panel de cet individu reste non contributif et le phénotype sévère de cet individu en fait un excellent candidat pour le projet INDEXES.

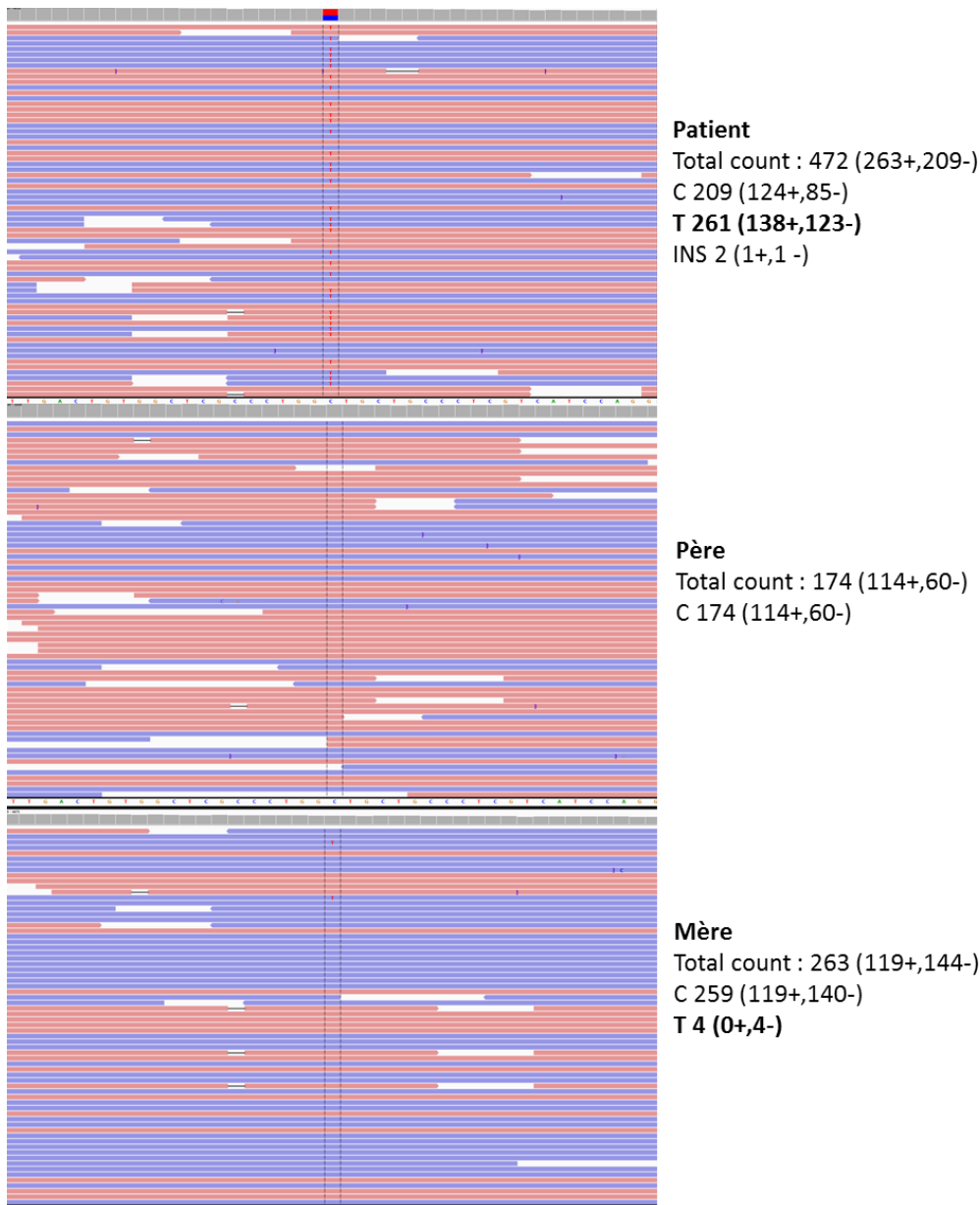


FIGURE 23 : Visualisation du variant dans *CREBBP* avec IGV.

L'individu est porteur hétérozygote de ce variant qui est identifié sur 55,3 % des reads. Ce variant est totalement absent du pool des pères : le père de l'individu n'est a priori pas porteur du variant. Le nombre de reads porteurs du variant dans le pool des mères est sous le seuil à partir duquel l'on considère qu'une des mères est porteuse hétérozygote (six reads attendus contre quatre observés). Bien que le syndrome de Rubinstein-Taybi n'affecte généralement qu'un seul individu d'une même famille (survenue *de novo*), il est décrit, dans certains cas, des parents d'intelligence normale porteurs d'une faible mosaïque somatique d'un variant pathogène dans *CREBBP*. Le faible taux de mosaïcisme et l'expressivité variable expliquent que ces individus ne soient pas affectés (Bartsch *et al.* 2010; Chiang *et al.* 2009; Tajir *et al.* 2013). L'hypothèse d'une éventuelle mosaïque chez la mère de l'individu pourrait expliquer les 4 reads porteurs du variant dans le pool des mères.

IV.2 Perspectives : bénéfice diagnostic du projet INDEXES

L'approche par séquençage d'exome permet une analyse des régions codantes des gènes et des sites d'épissage. Bien qu'il ne représente qu'environ 1 % du génome, l'exome contient à lui seul 85 % des mutations impliquées dans les maladies génétiques mendéliennes (Choi *et al.* 2009).

La stratégie *de novo*, basée sur le séquençage d'exome en trio et utilisée dans le projet INDEXES, est la stratégie de choix pour la recherche des causes génétiques dans les maladies sporadiques (Gilissen *et al.* 2014). Bien que les transmissions liées à l'X et les transmissions autosomiques récessives ne puissent pas être exclues, une proportion importante de DI est causée par des mutations *de novo* dans les gènes codants des protéines. Gilissen *et al.*, montre en 2012 que cette stratégie peut permettre d'élucider des causes génétiques de DI non syndromique jusqu'à 6 cas sur 10 étudiés (Gilissen *et al.* 2012). Comme discuté précédemment, le rendement peut être amélioré par des réanalyses régulières des résultats non contributifs. Cette stratégie s'applique donc particulièrement bien à la recherche de mutations responsables de DI.

INDEXES offre une possibilité supplémentaire d'identifier la cause moléculaire des troubles de ces individus et de mettre fin à leur errance diagnostique.

V CONCLUSION

Cette réinterprétation n'a pas permis de résoudre de nouveaux cas. Plusieurs hypothèses peuvent expliquer l'absence de nouveaux diagnostics, mais il semble que la réinterprétation des variants en sortie d'un même pipeline bioinformatique, ainsi que le nombre limité de gènes inclus dans le panel constituent les principales limites pour une réanalyse efficace. Les cas restant non informatifs pourraient bénéficier d'une réanalyse après une révision du pipeline bioinformatique, voire d'un séquençage d'exome.

La réinterprétation a toutefois permis d'identifier huit candidats à l'inclusion pour le projet INDEXES. Parmi eux, quatre ont un phénotype sévère. La réanalyse est non contributive pour deux d'entre eux (DL et ST) et des VUS sont mis en évidence chez les deux autres (FE et RD), nécessitant pour l'un une confirmation par séquençage de Sanger et pour l'autre un dialogue clinico-biologique. Pour les quatre autres (ON, ZI, AN et DM), les informations phénotypiques doivent être mises à jour par les cliniciens avant d'envisager une inclusion dans INDEXES.

Pour conclure, même en l'absence de coût matériel supplémentaire, compte-tenu du temps nécessaire et du rendement nul, la valeur d'une réinterprétation à un an de variants identifiés par séquençage d'un panel de gènes de DI, à la seule lumière de la littérature actualisée n'est pas démontrée.

BIBLIOGRAPHIE

- Adzhubei, I., D. M. Jordan, and S. R. Sunyaev. 2013. 'Predicting functional effect of human missense mutations using PolyPhen-2', *Curr Protoc Hum Genet*, Chapter 7: Unit7 20.
- Al-Nabhani, M., S. Al-Rashdi, F. Al-Murshedi, A. Al-Kindi, K. Al-Thihli, A. Al-Saegh, A. Al-Futaisi, W. Al-Mamari, F. Zadjali, and A. Al-Maawali. 2018. 'Re-analysis of Exome Sequencing Data of Intellectual Disability Samples: Yields and Benefits', *Clin Genet*.
- Alfares, A., T. Aloraini, L. A. Subaie, A. Alissa, A. A. Qudsi, A. Alahmad, F. A. Mutairi, A. Alswaid, A. Alothaim, W. Eyaid, M. Albalwi, S. Alturki, and M. Alfadhel. 2018. 'Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing', *Genet Med*.
- Agilent Technologies. SureSelect target Enrichment System. Protocol. Agilent Technologies, Inc. Décembre 2009.
- Amberger, J. S., C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh. 2015. 'OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders', *Nucleic Acids Res*, 43: D789-98.
- Amendola, L. M., G. P. Jarvik, M. C. Leo, H. M. McLaughlin, Y. Akkari, M. D. Amaral, J. S. Berg, S. Biswas, K. M. Bowling, L. K. Conlin, G. M. Cooper, M. O. Dorschner, M. C. Dulik, A. A. Ghazani, R. Ghosh, R. C. Green, R. Hart, C. Horton, J. J. Johnston, M. S. Lebo, A. Milosavljevic, J. Ou, C. M. Pak, R. Y. Patel, S. Punj, C. S. Richards, J. Salama, N. T. Strande, Y. Yang, S. E. Plon, L. G. Biesecker, and H. L. Rehm. 2016. 'Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium', *Am J Hum Genet*, 98: 1067-76.
- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5)* (American Psychiatric Pub).
- Baldrige, D., J. Heeley, M. Vineyard, L. Manwaring, T. L. Toler, E. Fassi, E. Fiala, S. Brown, C. W. Goss, M. Willing, D. K. Grange, B. A. Kozel, and M. Shinawi. 2017. 'The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results', *Genet Med*, 19: 1040-48.
- Bartsch, O., W. Kress, O. Kempf, S. Lechno, T. Haaf, and U. Zechner. 2010. 'Inheritance and variable expression in Rubinstein-Taybi syndrome', *Am J Med Genet A*, 152A: 2254-61.
- Beckman Coulter. SPRIselect User Guide. SPRI based Size Selection. Beckman Coulter, Inc. Octobre 2012.
- Belhedi, N., N. Perroud, F. Karege, M. Vessaz, A. Malafosse, and A. Salzmann. 2014. 'Increased CPA6 promoter methylation in focal epilepsy and in febrile seizures', *Epilepsy Res*, 108: 144-8.
- Birney, E., and N. Soranzo. 2015. 'Human genomics: The end of the start for population sequencing', *Nature*, 526: 52-3.
- Bowling, K. M., M. L. Thompson, M. D. Amaral, C. R. Finnila, S. M. Hiatt, K. L. Engel, J. N. Cochran, K. B. Brothers, K. M. East, D. E. Gray, W. V. Kelley, N. E. Lamb, E. J. Lose, C. A. Rich, S. Simmons, J. S. Whittle, B. T. Weaver, A. S. Nesmith, R. M. Myers, G. S. Barsh, E. M. Bebin, and G. M. Cooper. 2017. 'Genomic diagnosis for children with intellectual disability and/or developmental delay', *Genome Med*, 9: 43.
- Broad Institute. <https://www.broadinstitute.org>. Broad – Sample Preparation (Tracking, Shearing, Diversity, Size Selection, SPRI), de Broad Institute, [en ligne], disponible sur <https://www.broadinstitute.org/genome-sequencing/broadillumina-genome-analyzer-boot-camp> (page consultée le 05/07/2018)
- Buermans, H. P., and J. T. den Dunnen. 2014. 'Next generation sequencing technology: Advances and applications', *Biochim Biophys Acta*, 1842: 1932-41.
- Chelly, J., M. Khelifaoui, F. Francis, B. Cherif, and T. Bienvenu. 2006. 'Genetics and pathophysiology of mental retardation', *Eur J Hum Genet*, 14: 701-13.

- Chiang, P. W., N. C. Lee, N. Chien, W. L. Hwu, E. Spector, and A. C. Tsai. 2009. 'Somatic and germ-line mosaicism in Rubinstein-Taybi syndrome', *Am J Med Genet A*, 149A: 1463-7.
- Choi, M., U. I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloglu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane, and R. P. Lifton. 2009. 'Genetic diagnosis by whole exome capture and massively parallel DNA sequencing', *Proc Natl Acad Sci U S A*, 106: 19096-101.
- Choi, Y., G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan. 2012. 'Predicting the functional effect of amino acid substitutions and indels', *PLoS One*, 7: e46688.
- Chun, S., and J. C. Fay. 2009. 'Identification of deleterious mutations within three human genomes', *Genome Res*, 19: 1553-61.
- Collège National des Enseignants et Praticiens de Génétique Médicale. 2016. *GÉNÉTIQUE MÉDICALE* (Elsevier Masson).
- Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental Disorders; Board on the Health of Select Populations; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine. 2015. *Mental Disorders and Disabilities Among Low-Income Children* (The National Academies Press: Washington, DC).
- Cooper, G. M., E. A. Stone, G. Asimenos, Nisc Comparative Sequencing Program, E. D. Green, S. Batzoglou, and A. Sidow. 2005. 'Distribution and intensity of constraint in mammalian genomic sequence', *Genome Res*, 15: 901-13.
- Cornish, A., and C. Guda. 2015. 'A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference', *Biomed Res Int*, 2015: 456479.
- Costain, G., R. Jobling, S. Walker, M. S. Reuter, M. Snell, S. Bowdin, R. D. Cohn, L. Dupuis, S. Hewson, S. Mercimek-Andrews, C. Shuman, N. Sondheimer, R. Weksberg, G. Yoon, M. S. Meyn, D. J. Stavropoulos, S. W. Scherer, R. Mendoza-Londono, and C. R. Marshall. 2018. 'Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing', *Eur J Hum Genet*, 26: 740-44.
- de Ligt, J., M. H. Willemsen, B. W. van Bon, T. Kleefstra, H. G. Yntema, T. Kroes, A. T. Vulto-van Silfhout, D. A. Koolen, P. de Vries, C. Gilissen, M. del Rosario, A. Hoischen, H. Scheffer, B. B. de Vries, H. G. Brunner, J. A. Veltman, and L. E. Vissers. 2012. 'Diagnostic exome sequencing in persons with severe intellectual disability', *N Engl J Med*, 367: 1921-9.
- Dong, C., P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, and X. Liu. 2015. 'Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies', *Hum Mol Genet*, 24: 2125-37.
- Eilbeck, K., S. E. Lewis, C. J. Mungall, M. Yandell, L. Stein, R. Durbin, and M. Ashburner. 2005. 'The Sequence Ontology: a tool for the unification of genome annotations', *Genome Biol*, 6: R44.
- Ellison, J. W., J. A. Rosenfeld, and L. G. Shaffer. 2013. 'Genetic basis of intellectual disability', *Annu Rev Med*, 64: 441-50.
- Ewans, L. J., D. Schofield, R. Shrestha, Y. Zhu, V. Gayevskiy, K. Ying, C. Walsh, E. Lee, E. P. Kirk, A. Colley, C. Ellaway, A. Turner, D. Mowat, L. Worgan, M. L. Freckmann, M. Lipke, R. Sachdev, D. Miller, M. Field, M. E. Dinger, M. F. Buckley, M. J. Cowley, and T. Roscioli. 2018. 'Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders', *Genet Med*.
- Fallet-Bianco, C., A. Laquerriere, K. Poirier, F. Razavi, F. Guimiot, P. Dias, L. Loeuillet, K. Lascelles, C. Beldjord, N. Carion, A. Toussaint, N. Revencu, M. C. Addor, B. Lhermitte, M. Gonzales, J. Martinovich, B. Bessieres, M. Marcy-Bonniere, F. Jossic, P. Marcorelles, P. Loget, J. Chelly, and N. Bahi-Buisson. 2014. 'Mutations in tubulin genes are frequent causes of various foetal malformations of cortical development including microlissencephaly', *Acta Neuropathol Commun*, 2: 69.

- Fromer, M., and S. M. Purcell. 2014. 'Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data', *Curr Protoc Hum Genet*, 81: 7 23 1-21.
- Geiger, J. T., A. B. Schindler, C. Blauwendraat, H. S. Singer, and S. W. Scholz. 2017. 'TUBB2B Mutation in an Adult Patient with Myoclonus-Dystonia', *Case Rep in Neuro*, 9: 216-21.
- Gilissen, C., J. Y. Hehir-Kwa, D. T. Thung, M. van de Vorst, B. W. van Bon, M. H. Willemsen, M. Kwint, I. M. Janssen, A. Hoischen, A. Schenck, R. Leach, R. Klein, R. Tearle, T. Bo, R. Pfundt, H. G. Yntema, B. B. de Vries, T. Kleefstra, H. G. Brunner, L. E. Vissers, and J. A. Veltman. 2014. 'Genome sequencing identifies major causes of severe intellectual disability', *Nature*, 511: 344-7.
- Gilissen, C., A. Hoischen, H. G. Brunner, and J. A. Veltman. 2012. 'Disease gene identification strategies for exome sequencing', *Eur J Hum Genet*, 20: 490-7.
- Goodwin, S., J. D. McPherson, and W. R. McCombie. 2016. 'Coming of age: ten years of next-generation sequencing technologies', *Nat Rev Genet*, 17: 333-51.
- Guerrini, R., D. Mei, D. M. Cordelli, D. Pucatti, E. Franzoni, and E. Parrini. 2012. 'Symmetric polymicrogyria and pachygyria associated with TUBB2B gene mutations', *Eur J Hum Genet*, 20: 995-8.
- Guo, Y., J. Li, C. I. Li, J. Long, D. C. Samuels, and Y. Shyr. 2012. 'The effect of strand bias in Illumina short-read sequencing data', *BMC Genomics*, 13: 666.
- Hamdan, F. F., M. Srour, J. M. Capo-Chichi, H. Daoud, C. Nassif, L. Patry, C. Massicotte, A. Ambalavanan, D. Spiegelman, O. Diallo, E. Henrion, A. Dionne-Laporte, A. Fougerat, A. V. Pshezhetsky, S. Venkateswaran, G. A. Rouleau, and J. L. Michaud. 2014. 'De novo mutations in moderate or severe intellectual disability', *PLoS Genet*, 10: e1004772.
- Harripaul, R., A. Noor, M. Ayub, and J. B. Vincent. 2017. 'The Use of Next-Generation Sequencing for Research and Diagnostics for Intellectual Disability', *Cold Spring Harb Perspect Med*, 7.
- Hennekam, R. C. 2006. 'Rubinstein-Taybi syndrome', *Eur J Hum Genet*, 14: 981-5.
- Hiatt, S. M., M. D. Amaral, K. M. Bowling, C. R. Finnila, M. L. Thompson, D. E. Gray, J. M. J. Lawlor, J. N. Cochran, E. M. Bebin, K. B. Brothers, K. M. East, W. V. Kelley, N. E. Lamb, S. E. Levy, E. J. Lose, M. B. Neu, C. A. Rich, S. Simmons, R. M. Myers, G. S. Barsh, and G. M. Cooper. 2018. 'Systematic reanalysis of genomic data improves quality of variant interpretation', *Clin Genet*, 94: 174-78.
- Hill, M. K., A. D. Archibald, J. Cohen, and S. A. Metcalfe. 2010. 'A systematic review of population screening for fragile X syndrome', *Genet Med*, 12: 396-410.
- Jaglin, X. H., K. Poirier, Y. Saillour, E. Buhler, G. Tian, N. Bahi-Buisson, C. Fallet-Bianco, F. Phan-Dinh-Tuy, X. P. Kong, P. Bomont, L. Castelnau-Ptakhine, S. Odent, P. Loget, M. Kossorotoff, I. Snoeck, G. Plessis, P. Parent, C. Beldjord, C. Cardoso, A. Represa, J. Flint, D. A. Keays, N. J. Cowan, and J. Chelly. 2009. 'Mutations in the beta-tubulin gene TUBB2B result in asymmetrical polymicrogyria', *Nat Genet*, 41: 746-52.
- Jagadeesh, K. A., A. M. Wenger, M. J. Berger, H. Guturu, P. D. Stenson, D. N. Cooper, J. A. Bernstein, and G. Bejerano. 2016. 'M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity', *Nat Genet*, 48: 1581.
- Kaufman, L., M. Ayub, and J. B. Vincent. 2010. 'The genetic basis of non-syndromic intellectual disability: a review', *J Neurodev Disord*, 2: 182-209.
- Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, and J. Shendure. 2014. 'A general framework for estimating the relative pathogenicity of human genetic variants', *Nat Genet*, 46: 310-5.
- Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson. 2012. 'VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing', *Genome Res*, 22: 568-76.
- Kwok, R. P., J. R. Lundblad, J. C. Chrivia, J. P. Richards, H. P. Bachinger, R. G. Brennan, S. G. Roberts, M. R. Green, and R. H. Goodman. 1994. 'Nuclear protein CBP is a coactivator for the transcription factor CREB', *Nature*, 370: 223-6.

- Lacoste, C., A. Fabre, C. Pecheux, N. Levy, M. Krahn, P. Malzac, N. Bonello-Palot, C. Badens, and P. Bourgeois. 2017. '[Next-generation DNA sequencing in clinical diagnostics]', *Arch Pediatr*, 24: 373-83.
- Landrum, M. J., J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, and D. R. Maglott. 2016. 'ClinVar: public archive of interpretations of clinically relevant variants', *Nucleic Acids Res*, 44: D862-8.
- Landrum, M. J., J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein, D. M. Church, and D. R. Maglott. 2014. 'ClinVar: public archive of relationships among sequence variation and human phenotype', *Nucleic Acids Res*, 42: D980-5.
- Laquerriere, A., M. Gonzales, Y. Saillour, M. Cavallin, N. Joye, C. Quelin, L. Bidat, M. Dommergues, G. Plessis, F. Encha-Razavi, J. Chelly, N. Bahi-Buisson, and K. Poirier. 2016. 'De novo TUBB2B mutation causes fetal akinesia deformation sequence with microlissencephaly: An unusual presentation of tubulinopathy', *Eur J Med Genet*, 59: 249-56.
- Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, and Consortium Exome Aggregation. 2016. 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature*, 536: 285-91.
- Lelieveld, S. H., M. R. Reijnders, R. Pfundt, H. G. Yntema, E. J. Kamsteeg, P. de Vries, B. B. de Vries, M. H. Willemsen, T. Kleefstra, K. Lohner, M. Vreeburg, S. J. Stevens, I. van der Burgt, E. M. Bongers, A. P. Stegmann, P. Rump, T. Rinne, M. R. Nelen, J. A. Veltman, L. E. Vissers, H. G. Brunner, and C. Gilissen. 2016. 'Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability', *Nat Neurosci*, 19: 1194-6.
- Leonard, H., and X. Wen. 2002. 'The epidemiology of mental retardation: challenges and opportunities in the new millennium', *Ment Retard Dev Disabil Res Rev*, 8: 117-34.
- Li, H., and R. Durbin. 2009. 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25: 1754-60.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.
- Li, J., T. Cai, Y. Jiang, H. Chen, X. He, C. Chen, X. Li, Q. Shao, X. Ran, Z. Li, K. Xia, C. Liu, Z. S. Sun, and J. Wu. 2016. 'Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database', *Mol Psychiatry*, 21: 290-7.
- Marangi, G., S. Ricciardi, D. Orteschi, S. Lattante, M. Murdolo, B. Dallapiccola, C. Biscione, R. Lecce, P. Chiurazzi, C. Romano, D. Greco, R. Pettinato, G. Sorge, C. Pantaleoni, E. Alfei, I. Toldo, C. Magnani, P. Bonanni, F. Martinez, G. Serra, D. Battaglia, D. Lettori, G. Vasco, A. Baroncini, C. Daolio, and M. Zollino. 2011. 'The Pitt-Hopkins syndrome : report of 16 new patients and clinical diagnostic criteria', *Am J Med Genet A*, 7: 1536-45.

- Maulik, P. K., M. N. Mascarenhas, C. D. Mathers, T. Dua, and S. Saxena. 2011. 'Prevalence of intellectual disability: a meta-analysis of population-based studies', *Res Dev Disabil*, 32: 419-36.
- McCarthy, D. J., P. Humburg, A. Kanapin, M. A. Rivas, K. Gaulton, J. B. Cazier, and P. Donnelly. 2014. 'Choice of transcripts and software has a large effect on variant annotation', *Genome Med*, 6: 26.
- McClellan, J., and M. C. King. 2010a. 'Genetic heterogeneity in human disease', *Cell*, 141: 210-7.
- . 2010b. 'Genomic analysis of mental illness: a changing landscape', *JAMA*, 303: 2523-4.
- McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. 'The Ensembl Variant Effect Predictor', *Genome Biol*, 17: 122.
- Metzker, M. L. 2010. 'Sequencing technologies - the next generation', *Nat Rev Genet*, 11: 31-46.
- Monroe, G. R., G. W. Frederix, S. M. Savelberg, T. I. de Vries, K. J. Duran, J. J. van der Smagt, P. A. Terhal, P. M. van Hasselt, H. Y. Kroes, N. M. Verhoeven-Duif, I. J. Nijman, E. C. Carbo, K. L. van Gassen, N. V. Knoers, A. M. Hovels, M. M. van Haelst, G. Visser, and G. van Haften. 2016. 'Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability', *Genet Med*, 18: 949-56.
- Nambot, S., J. Thevenon, P. Kuentz, Y. Duffourd, E. Tisserant, A. L. Bruel, A. L. Mosca-Boidron, A. Masurel-Paulet, D. Lehalle, N. Jean-Marcais, M. Lefebvre, P. Vabres, S. El Chehadeh-Djebbar, C. Philippe, F. Tran Mau-Them, J. St-Onge, T. Jouan, M. Chevarin, C. Poe, V. Carmignac, A. Vitobello, P. Callier, J. B. Riviere, L. Faivre, C. Thauvin-Robinet, and Group Orphanomix Physicians. 2018. '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*, 20: 645-54.
- Ng, P. C., and S. Henikoff. 2003. 'SIFT: Predicting amino acid changes that affect protein function', *Nucleic Acids Res*, 31: 3812-4.
- Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure, and M. J. Bamshad. 2010. 'Exome sequencing identifies the cause of a mendelian disorder', *Nat Genet*, 42: 30-5.
- Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D. A. Nickerson, and J. Shendure. 2009. 'Targeted capture and massively parallel sequencing of 12 human exomes', *Nature*, 461: 272-6.
- Nogales, E., K. H. Downing, L. A. Amos, and J. Löwe. 1998. 'Tubulin and FtsZ form a distinct family of GTPases', *Nat Struct Mol Biol*, 5: 451-8.
- Nordheim, A. 1994. 'Transcription factors. CREB takes CBP to tango', *Nature*, 370: 177-8.
- ORPHANET. <https://www.orpha.net/>. Le portail des maladies rares et des médicaments orphelins. Rechercher un test diagnostique. [en ligne], disponible sur : [https://www.orpha.net/consor/cgi-bin/ClinicalLabs_Search.php?lng=FR&type_list=clinicalLabs_search_simple_shd&data_id=3256&Maladie\(s\)/groupes%20de%20maladies=Deficience-intellectuelle-non-syndromique-liee-a-l-X&title=Deficience-intellectuelle-non-syndromique-liee-a-l-X&search=ClinicalLabs_Search_Simple&ChdId=3256&ClinicalLabs_ClinicalLabs_Search_diseaseGroup=deficience-intellectuelle-liee-a-l-X&ClinicalLabs_ClinicalLabs_Search_diseaseType=Pat&ClinicalLabs_ClinicalLabs_Search_GeographicType=null&ClinicalLabs_ClinicalLabs_Search_ClinicalLabsList_Sort=QoS](https://www.orpha.net/consor/cgi-bin/ClinicalLabs_Search.php?lng=FR&type_list=clinicalLabs_search_simple_shd&data_id=3256&Maladie(s)/groupes%20de%20maladies=Deficience-intellectuelle-non-syndromique-liee-a-l-X&title=Deficience-intellectuelle-non-syndromique-liee-a-l-X&search=ClinicalLabs_Search_Simple&ChdId=3256&ClinicalLabs_ClinicalLabs_Search_diseaseGroup=deficience-intellectuelle-liee-a-l-X&ClinicalLabs_ClinicalLabs_Search_diseaseType=Pat&ClinicalLabs_ClinicalLabs_Search_GeographicType=null&ClinicalLabs_ClinicalLabs_Search_ClinicalLabsList_Sort=QoS)
- Petrovski, S., Q. Wang, E. L. Heinzen, A. S. Allen, and D. B. Goldstein. 2013. 'Genic intolerance to functional variation and the interpretation of personal genomes', *PLoS Genet*, 9: e1003709.

- Rauch, A., J. Hoyer, S. Guth, C. Zweier, C. Kraus, C. Becker, M. Zenker, U. Huffmeier, C. Thiel, F. Ruschendorf, P. Nurnberg, A. Reis, and U. Trautmann. 2006. 'Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation', *Am J Med Genet A*, 140: 2063-74.
- Redin, C., B. Gerard, J. Lauer, Y. Herenger, J. Muller, A. Quartier, A. Masurel-Paulet, M. Willems, G. Lesca, S. El-Chehadeh, S. Le Gras, S. Vicaire, M. Philipps, M. Dumas, V. Geoffroy, C. Feger, N. Haumesser, Y. Alembik, M. Barth, D. Bonneau, E. Colin, H. Dollfus, B. Doray, M. A. Delrue, V. Drouin-Garraud, E. Flori, M. Fradin, C. Francannet, A. Goldenberg, S. Lumbroso, M. Mathieu-Dramard, D. Martin-Coignard, D. Lacombe, G. Morin, A. Polge, S. Sukno, C. Thauvin-Robinet, J. Thevenon, M. Doco-Fenzy, D. Genevieve, P. Sarda, P. Edery, B. Isidor, B. Jost, L. Olivier-Faivre, J. L. Mandel, and A. Piton. 2014. 'Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing', *J Med Genet*, 51: 724-36.
- Reva, B., Y. Antipin, and C. Sander. 2011. 'Predicting the functional impact of protein mutations: application to cancer genomics', *Nucleic Acids Res*, 39: e118.
- Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm, and Acmg Laboratory Quality Assurance Committee. 2015. '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*, 17: 405-24.
- Robertson, J., C. Hatton, E. Emerson, and S. Baines. 2015. 'Prevalence of epilepsy among people with intellectual disabilities: A systematic review', *Seizure*, 29: 46-62.
- Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. 'Integrative genomics viewer', *Nat Biotechnol*, 29: 24-6.
- Romaniello, R., A. Tonelli, F. Arrigoni, C. Baschiroto, F. Triulzi, N. Bresolin, M. T. Bassi, and R. Borgatti. 2012. 'A novel mutation in the beta-tubulin gene TUBB2B associated with complex malformation of cortical development and deficits in axonal guidance', *Dev Med Child Neurol*, 54: 765-9.
- Rubinstein, J. H., and H. Taybi. 1963. 'Broad thumbs and toes and facial abnormalities. A possible mental retardation syndrome', *Am J Dis Child*, 105: 588-608.
- Salzmann, A., M. Guipponi, P. J. Lyons, L. D. Fricker, M. Sapio, C. Lamercy, C. Buresi, B. Ouled Amar Bencheikh, F. Lahjouji, R. Ouazzani, A. Crespel, D. Chaigne, and A. Malafosse. 2012. 'Carboxypeptidase A6 gene (CPA6) mutations in a recessive familial form of febrile seizures and temporal lobe epilepsy and in sporadic temporal lobe epilepsy', *Hum Mutat*, 33: 124-35.
- Sapio, M. R., A. Salzmann, M. Vessaz, A. Crespel, P. J. Lyons, A. Malafosse, and L. D. Fricker. 2012. 'Naturally occurring carboxypeptidase A6 mutations: effect on enzyme function and association with epilepsy', *J Biol Chem*, 287: 42900-9.
- Schwarz, J. M., D. N. Cooper, M. Schuelke, and D. Seelow. 2014. 'MutationTaster2: mutation prediction for the deep-sequencing age', *Nat Methods*, 11: 361-2.
- Shihab, H. A., J. Gough, D. N. Cooper, P. D. Stenson, G. L. Barker, K. J. Edwards, I. N. Day, and T. R. Gaunt. 2013. 'Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models', *Hum Mutat*, 34: 57-65.
- Sian Ellard, Helen Lindsay, Nick Camm, Chris Watson, Steve Abbs, Yvonne Wallis, Chris Mattocks, Graham R Taylor and Ruth Charlton. 2014. 'Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation.', *Association for Clinical Genetic Science*.
- Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. 'Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes', *Genome Res*, 15: 1034-50.

- Smith, E. D., K. Radtke, M. Rossi, D. N. Shinde, S. Darabi, D. El-Khechen, Z. Powis, K. Helbig, K. Waller, D. K. Grange, S. Tang, and K. D. Farwell Hagman. 2017. 'Classification of Genes: Standardized Clinical Validity Assessment of Gene-Disease Associations Aids Diagnostic Exome Analysis and Reclassifications', *Hum Mutat*, 38: 600-08.
- Sobreira, N., F. Schiettecatte, D. Valle, and A. Hamosh. 2015. 'GeneMatcher: a matching tool for connecting investigators with an interest in the same gene', *Hum Mutat*, 36: 928-30.
- Stevens, C. A., J. Pouncey, and D. Knowles. 2011. 'Adults with Rubinstein-Taybi syndrome', *Am J Med Genet A*, 155A: 1680-4.
- Tajir, M., P. Fergelot, G. Lancelot, S. C. Elalaoui, B. Arveiler, D. Lacombe, and A. Sefiani. 2013. 'Germline mosaicism in Rubinstein-Taybi syndrome', *Gene*, 518: 476-8.
- Veltman, J. A., and H. G. Brunner. 2012. 'De novo mutations in human genetic disease', *Nat Rev Genet*, 13: 565-75.
- Visser, L. E., J. de Ligt, C. Gilissen, I. Janssen, M. Steehouwer, P. de Vries, B. van Lier, P. Arts, N. Wieskamp, M. del Rosario, B. W. van Bon, A. Hoischen, B. B. de Vries, H. G. Brunner, and J. A. Veltman. 2010. 'A de novo paradigm for mental retardation', *Nat Genet*, 42: 1109-12.
- Visser, L. E., C. Gilissen, and J. A. Veltman. 2016. 'Genetic studies in intellectual disability and related disorders', *Nat Rev Genet*, 17: 9-18.
- Wenger, A. M., H. Gudur, J. A. Bernstein, and G. Bejerano. 2017. 'Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers', *Genet Med*, 19: 209-14.
- Wildeman, M., E. van Ophuizen, J. T. den Dunnen, and P. E. Taschner. 2008. 'Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker', *Hum Mutat*, 29: 6-13.
- Wright, C. F., T. W. Fitzgerald, W. D. Jones, S. Clayton, J. F. McRae, M. van Kogelenberg, D. A. King, K. Ambridge, D. M. Barrett, T. Bayzatinova, A. P. Bevan, E. Bragin, E. A. Chatzimichali, S. Gribble, P. Jones, N. Krishnappa, L. E. Mason, R. Miller, K. I. Morley, V. Parthiban, E. Prigmore, D. Rajan, A. Sifrim, G. J. Swaminathan, A. R. Tivey, A. Middleton, M. Parker, N. P. Carter, J. C. Barrett, M. E. Hurles, D. R. FitzPatrick, H. V. Firth, and D. D. D. study. 2015. 'Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data', *Lancet*, 385: 1305-14.
- Wright, C. F., J. F. McRae, S. Clayton, G. Gallone, S. Aitken, T. W. FitzGerald, P. Jones, E. Prigmore, D. Rajan, J. Lord, A. Sifrim, R. Kellsell, M. J. Parker, J. C. Barrett, M. E. Hurles, D. R. FitzPatrick, and H. V. Firth. 2018. 'Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders', *Genet Med*.
- Xue, Y., A. Ankala, W. R. Wilcox, and M. R. Hegde. 2015. 'Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing', *Genet Med*, 17: 444-51.
- Zerbino, D. R., P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Giron, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, and P. Flicek. 2018. 'Ensembl 2018', *Nucleic Acids Res*, 46: D754-D61.
- Zhu, X., S. Petrovski, P. Xie, E. K. Ruzzo, Y. F. Lu, K. M. McSweeney, B. Ben-Zeev, A. Nissenkorn, Y. Anikster, D. Oz-Levi, R. S. Dhindsa, Y. Hitomi, K. Schoch, R. C. Spillmann, G. Heimer, D. Marek-Yagel, M. Tzadok, Y. Han, G. Worley, J. Goldstein,

Y. H. Jiang, D. Lancet, E. Pras, V. Shashi, D. McHale, A. C. Need, and D. B. Goldstein. 2015. 'Whole-exome sequencing in undiagnosed genetic diseases: interpreting 119 trios', *Genet Med*, 17: 774-81.

ANNEXES

Outil	Ressource internet	Référence
Base de données		
1000 Genomes	http://1000genomes.org	(Birney and Soranzo 2015)
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	(Landrum <i>et al.</i> 2016; Landrum <i>et al.</i> 2014)
dbSNP	https://www.ncbi.nlm.nih.gov/snp	/
Ensembl	http://www.ensembl.org	(Zerbino <i>et al.</i> 2018)
ExAC	http://exac.broadinstitute.org	(Lek <i>et al.</i> 2016)
gnomAD	http://gnomad.broadinstitute.org	(Lek <i>et al.</i> 2016)
OMIM	https://www.omim.org	(Amberger <i>et al.</i> 2015)
RVIS	http://genic-intolerance.org	(Petrovski <i>et al.</i> 2013)
Sequence Ontology	https://www.ensembl.org/Tools/VEP	(Eilbeck <i>et al.</i> 2005)
Outil bioinformatique		
BWA	http://bio-bwa.sourceforge.net	(Li and Durbin 2009)
CADD	http://cadd.gs.washington.edu/	(Kircher <i>et al.</i> 2014)
FATHMM	http://fathmm.biocompute.org.uk	(Shihab <i>et al.</i> 2013)
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	/
Galaxy	https://galaxyproject.org	/
GERP ++	http://mendel.stanford.edu/SidowLab/downloads/gerp/	(Cooper <i>et al.</i> 2005)
IGV	http://igv.org/	(Robinson <i>et al.</i> 2011)
LRT	http://www.genetics.wustl.edu/jflab/lrt_query.html	(Chun and Fay 2009)
MetaSVM	http://annovar.openbioinformatics.org/	(Dong <i>et al.</i> 2015)
Mutalyzer	https://www.mutalyzer.nl/	(Wildeman <i>et al.</i> 2008)
MutationTaster	http://www.mutationtaster.org/	(Schwarz <i>et al.</i> 2014)
MutationAssesor	http://mutationassessor.org/	(Reva, Antipin, and Sander 2011)
PhastCons	http://compugen.cshl.edu/phast/	(Siepel <i>et al.</i> 2005)
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	(Adzhubei, Jordan, and Sunyaev 2013)
PROVEAN	http://provean.jcvi.org	(Choi <i>et al.</i> 2012)
SAMtools	http://samtools.sourceforge.net	(Li <i>et al.</i> 2009)
SeattleSeq	http://snp.gs.washington.edu/SeattleSeqAnnotation138/	(Ng <i>et al.</i> 2009)
SIFT	http://sift.jcvi.org/	(Ng and Henikoff 2003)
VarScan2	http://varscan.sourceforge.net	(Koboldt <i>et al.</i> 2012)
VarSome	https://varsome.com/	/
VEP	https://www.ensembl.org/Tools/VEP	(McLaren <i>et al.</i> 2016)
XHMM	https://atgu.mgh.harvard.edu/xhmm/	(Fromer and Purcell 2014)

ANNEXE 1 : Bases de données et outils principalement utilisés pour l'alignement, l'identification, l'annotation et la classification des variants.

ACAN	ATXN3L	CRBN	GABBR2	IKBKG	MID1	PHF21A	SHROOM4	TBX1
ACAT1	AUTS2	CREBBP	GABRA1	IL1RAPL1	KMT2A	PHF6	SKI	TCF4
ACSL4	BCKDHA	CSDE1	GABRB3	IQSEC2	MMACHC	PHF8	SLC13A5	TECR
ACTB	BCKDHB	CTNNB1	GABRD	IVD	MSL3	PHGDH	SLC16A2	TIMM8A
ACTG1	BCKDK	CTSF	GABRG2	KANSL1	MTHFR	PIGV	SLC17A5	TMLHE
ACY1	BCOR	CUL4B	GALT	KANSL2	MTR	PLCB1	SLC1A1	TNK2
ADA	BMP3	CUX2	GAMT	KAT6A	MUT	PLP1	SLC25A13	TPP1
ADNP	BRD3	CYFIP1	GAP43	KAT6B	MYCN	PMM2	SLC25A15	TRAPPC9
ADSL	BRPF1	CYP11A1	GATAD2B	KATNAL2	MYT1L	PNKP	SLC25A22	TRIO
AFF2	BRWD3	CYP27A1	GATM	KCNA1	NAA10	POGZ	SLC2A1	TRIP12
AGA	BTD	CYP2U1	GBA	KCNB1	NAGLU	POMGNT1	SLC35A2	TSC1
AGO1	C12ORF57	D2HGDH	GCDH	KCNJ10	NDE1	PORCN	SLC35A3	TSC2
AHI1	CA8	DBT	GCH1	KCNK9	NDST1	POU1F1	SLC35C1	TSPAN7
ALDH18A1	CACNA1C	DCX	GCSH	KCNMA1	NDUFA1	PPP2R5D	SLC46A1	TUBA1A
ALDH3A2	CACNA1D	DDHD2	GD1	KCNQ2	NEDD4L	PPT1	SLC6A1	TUBB2B
ALDH5A1	CACNA1F	DDX3X	GFPT2	KCNQ3	NF1	PQBP1	SLC6A4	TUSC3
ALDH7A1	CACNG2	DEAF1	GK	KCNT1	NFIX	PRKRA	SLC6A8	UBE2A
ALDOB	CAMTA1	DEPDC5	GLDC	KCTD13	NGFR	PRODH	SLC9A6	UBE3A
ALG1	CASK	DHCR7	GLUD1	KDM5C	NHS	PRPS1	SMARCA2	UBE3B
ALG12	CASR	DISC1	GNAO1	KDM6A	NIPBL	PRRT2	SMARCA4	UBR1
ALG13	CBS	DKC1	GPC3	KIAA2022	NLGN3	PRSS12	SMARCB1	UPB1
ALG3	CC2D1A	DLG1	GPRASP1	KIF1A	NLGN4X	PSAP	SMARCE1	UPF3B
ALG6	CCDC22	DLG3	GRIA3	KLF8	NPC1	PTCHD1	SMC1A	UROC1
AMT	CDH10	DMD	GRID2	KLHL15	NPC2	PTEN	SMC3	USP27X
ANK2	CDH9	DNM1	GRIK2	KMT2D	NRXN1	PTPN11	SMS	USP9X
ANK3	CDK5RAP2	DNM3	GRIN1	L1CAM	NRXN2	PTS	SOBP	VCP
ANKH	CDK6	DPAGT1	GRIN2A	LAMC3	NSD1	PURA	SOX11	VLDLR
ANKRD11	CDKL5	DPM1	GRIN2B	LAMP2	NSDHL	QARS	SOX3	VPS13B
AP1S2	CENPJ	DSCAM	GRIP1	LAS1L	NSUN2	RAB39B	SOX5	WDR45
AP4B1	CEP152	DYNC1H1	GTF2I	LGI1	NTNG1	RAI1	SOX7	WDR62
AP4E1	CHD2	DYRK1A	GUSB	LIMK1	OCRL	RBM10	SPTAN1	WHSC1
AP4M1	CHD7	EHMT1	HCCS	LRP2	OFD1	RELN	SRCAP	WNK3
AP4S1	CHD8	EIF2S3	HCFC1	MACROD2	OPHN1	RLIM	SRD5A3	WWOX
ARFGEF2	CHRNA2	EIF4E	HCN1	MAGI2	OTC	RNF135	SRGAP3	YWHAE
ARHGEF9	CHRNA4	ELP2	HDAC4	MAN1B1	OXTR	RORA	ST3GAL3	ZBTB20
ARID1A	CHRNA7	EN2	HDAC8	MAN2B1	PACS1	RPL10	ST3GAL5	ZC3H14
ARID1B	CHRN2	EP300	HEPACAM	MANBA	PAFAH1B1	RPS6KA3	STAG1	ZC4H2
ARIH1	CLCN2	EPB41L1	HEXA	MAOA	PAH	SATB2	STIL	ZDHHHC9
ARSA	CLCN4	ERLIN2	HGSNAT	MAST1	PAK3	SCN1A	STX1A	ZEB2
ARX	CLIC2	FGD1	HIVEP2	MBD5	PAX5	SCN1B	STXBP1	ZFHX4
ASAH1	CLN3	FLNA	HMGA2	MBTPS2	PCCA	SCN2A	SYN1	ZMYM3
ASL	CLN5	FMR1	HOXA1	MCPH1	PCCB	SCN8A	SYNE1	ZMYND11
ASPA	CLN6	FOLR1	HPRT1	MECP2	PCDH19	SEMA5A	SYNGAP1	ZNF292
ASPM	CLN8	FOXG1	HRAS	MED12	PCDH8	SETBP1	SYP	ZNF526
ASS1	CNKSR2	FOXP1	HSD17B10	MED13L	PDHA1	SETD2	SZT2	ZNF711
ATP6AP2	CNOT3	FOXP2	HUWE1	MED17	PDHX	SETD5	TAF1	ZNF81
ATP7A	CNTN4	FRMPD4	IDS	MED23	PEX1	SGSH	TAKO2	
ATP7B	CNTNAP2	FTCD	IDUA	MEF2C	PEX12	SHANK1	TBC1D24	
ATR	CPA6	FTSJ1	IGF1	MET	PEX6	SHANK2	TBL1XR1	
ATRX	CPS1	FUCA1	IGF1R	MFSD8	PEX7	SHANK3	TBR1	

ANNEXE 2 : Liste des 446 analysés, version n°2 du panel.

Les 7 gènes supplémentaires par rapport à la version 1 apparaissent en gris.

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

ANNEXE 3 : Récapitulatif des critères de l'ACMG permettant de classer les variants en 5 catégories : bénin, probablement bénin, VUS, probablement pathogène et pathogène (d'après Richards et al., 2015).

Les lignes organisent les critères en différents types de paramètres d'interprétation.
Les colonnes organisent les critères en fonction de leur force pour une affirmation bénigne (à gauche) ou pathogène (à droite) :

BS : *Benign Strong* (en violet)

BP : *Benign Supporting* (en bleu)

PM : *Pathogenic Moderate* (en vert)

PP : *Pathogenic Supporting* (en orange)

PS : *Pathogenic Strong* (en rouge)

PVS : *Pathogenic Very Strong* (en rouge foncé)

FH : Family History ; LOF : Loss Of Function ; MAF : Minor Allele Frequency.

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	<p>BS1 Allele frequency is greater than expected for disorder (see Table 6)</p> <p>BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age</p> <p>BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing</p> <p>BS4 Lack of segregation in affected members of a family</p> <p>Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.</p>
Supporting	<p>BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease</p> <p>BP2 Observed in <i>trans</i> with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in <i>cis</i> with a pathogenic variant in any inheritance pattern</p> <p>BP3 In-frame deletions/insertions in a repetitive region without a known function</p> <p>BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)</p> <p>Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.</p> <p>BP5 Variant found in a case with an alternate molecular basis for disease</p> <p>BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation</p> <p>BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved</p>

ANNEXE 4 : Récapitulatif du poids de chaque critère de classification des variants bénins et de leur précautions d'interprétation (selon Richards et al., 2015).

Il détaille chacun des critères utilisés pour l'interprétation d'un variant. Ils sont regroupés en fonction de leur preuve de pathogénicité.

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical \pm1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> • Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>) • Use caution interpreting LOF variants at the extreme 3' end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case-control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p> <p>PM2 Absent from controls (or at extremely low frequency if recessive) (Table 6) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.</p> <p>PM3 For recessive disorders, detected in <i>trans</i> with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase.</p> <p>PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants</p> <p>PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.</p> <p>PM6 Assumed de novo, but without confirmation of paternity and maternity</p>
Supporting	<p>PP1 cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data</p> <p>PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease</p> <p>PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.</p> <p>PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology</p> <p>PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation</p>

ANNEXE 5 : Récapitulatif du poids de chaque critère de classification des variants pathogènes et de leurs précautions d'interprétation (selon Richards *et al.*, 2015).

Il détaille chacun des critères utilisés pour l'interprétation d'un variant. Ils sont regroupés en fonction de leur preuve de pathogénicité.






LOF : Loss Of Function ; OR : Odd Ratio.

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

ANNEXE 6 : Règles de combinaison des différents critères permettant de classer un variant de séquence en variant bénin (classe 1), variant probablement bénin (classe 2), VUS (classe 3), variant probablement pathogène (classe 4) ou variant pathogène (classe 5) (selon Richards et al., 2015).

DEMANDE D'IMPRIMATUR

Date de soutenance : 5 octobre 2018

<p align="center">DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE</p> <p>présenté par : Charlotte FELICI</p> <p><u>Sujet :</u> Valeur ajoutée d'une réanalyse de données de séquençage d'un panel de gènes de déficience intellectuelle.</p> <p><u>Jury :</u> Président : M. Jean-Louis MERLIN, Professeur Directeur : Mme Anne-Sophie DENOMME-PICHON, Médecin Juges : M. Philippe JONVEAUX, Professeur Mme Céline BONNET, Maître de Conférences M. Marc MULLER, Pharmacien</p>	<p align="center">Vu, Nancy, le 17 août 2018</p> <p>Le Président du Jury Directeur de Thèse</p> <p>M. Jean-Louis MERLIN Mme Anne-Sophie DENOMME-PICHON</p> <p align="center"> </p>
<p align="center">Vu et approuvé, Nancy, le 7.09.2018</p> <p align="center">Doyen de la Faculté de Pharmacie de l'Université de Lorraine,</p> <p align="center"> Pr. Raphaël DUVAL</p> <p align="center"></p>	<p align="center">Vu, Nancy, le 13 SEP. 2018</p> <p align="center">Le Président de l'Université de Lorraine,</p> <p align="center"> Pierre MUTZENHARDT</p> <p align="center">N° d'enregistrement : 10425</p>

N° d'identification : 10425

TITRE

Valeur ajoutée d'une réanalyse de données de séquençage d'un panel de gènes de déficience intellectuelle.

Thèse soutenue le 5 octobre 2018

Par Charlotte FELICI

RESUME

La DI est un trouble neuro-développemental hétérogène dont une large proportion est due à des variants rares. Aujourd'hui, près de 30 % des DI restent sans diagnostic moléculaire.

L'analyse d'un panel de 446 gènes de DI a été effectuée chez 66 individus au laboratoire de génétique du CHRU de Nancy et a permis de résoudre 22 cas (33 %). Les 44 individus restant sans diagnostic étiologique font l'objet d'une réanalyse qui est le sujet de ce travail. Il s'agit de réinterpréter les variants identifiés précédemment en tenant compte de la littérature actuelle. L'objectif est triple : évaluer la valeur ajoutée d'une réanalyse de variants, résoudre de nouveaux cas et sélectionner parmi les individus restant sans diagnostic moléculaire des candidats pour l'inclusion dans le projet INDEXES (INtellectual Disability EXploration by Exome Sequencing) visant à explorer 10 sujets atteints de DI sévère par séquençage exomique en trio.

Aucun nouveau cas n'est résolu avec cette réinterprétation mais quatre individus sans diagnostic moléculaire sont identifiés comme de très bons candidats pour une inclusion dans INDEXES.

De nombreuses hypothèses expliquent l'absence de nouveaux diagnostics lors de la réanalyse mais il semble que la réinterprétation des variants en sortie d'un même pipeline bioinformatique, ainsi que le nombre de gènes limités inclus dans le panel constituent les principales limites pour une réanalyse efficace.

En conclusion, compte-tenu du temps nécessaire et du rendement nul, la valeur d'une réinterprétation de variants identifiés par séquençage d'un panel de gènes de DI, à la seule lumière de la littérature actualisée, à un an environ, n'est pas démontrée.

MOTS CLES : Panel de gènes – Déficience intellectuelle – Séquençage de nouvelle génération – Réinterprétation de variants

Directeur de thèse	Intitulé du laboratoire	Nature
Dr Anne-Sophie DENOMME-PICHON	Laboratoire de génétique CHRU de Nancy	Expérimentale <input checked="" type="checkbox"/>
Pr. Jean-Louis MERLIN	Service de biopathologie ICL Nancy	Bibliographique <input type="checkbox"/>
		Thème <input checked="" type="radio"/>

THEMES

1 – Sciences fondamentales
3 – Médicament
⑤ – Biologie

2 – Hygiène/Environnement
4 – Alimentation – Nutrition
6 – Pratique professionnelle