

THÈSE

**POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE
SPÉCIALITÉ BIOLOGIE MÉDICALE**

Présentée et soutenue publiquement
par

Julie PERNIN-GRANDJEAN

Le 28 mars 2019

**MISE AU POINT D'UNE MÉTHODE DE SÉQUENÇAGE DE NOUVELLE GÉNÉRATION
POUR LE DIAGNOSTIC GÉNÉTIQUE DE LA DÉFICIENCE INTELLECTUELLE**

Directeur de thèse : Dr Eric BIETH

JURY :

Monsieur le Professeur **Patrick CALVAS**
Monsieur le Professeur **Gilles FAVRE**
Monsieur le Professeur **Yves CHAIX**
Madame le Professeur **Bettina COUDERC**
Monsieur le Docteur **Eric BIETH**
Madame le Docteur **Sophie JULIA**

Président
Assesseur
Assesseur
Assesseur
Assesseur
Assesseur

THÈSE

**POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE
SPÉCIALITÉ BIOLOGIE MÉDICALE**

Présentée et soutenue publiquement
par

Julie PERNIN-GRANDJEAN

Le 28 mars 2019

**MISE AU POINT D'UNE MÉTHODE DE SÉQUENÇAGE DE NOUVELLE GÉNÉRATION
POUR LE DIAGNOSTIC GÉNÉTIQUE DE LA DÉFICIENCE INTELLECTUELLE**

Directeur de thèse : Dr Eric BIETH

JURY :

Monsieur le Professeur **Patrick CALVAS**
Monsieur le Professeur **Gilles FAVRE**
Monsieur le Professeur **Yves CHAIX**
Madame le Professeur **Bettina COUDERC**
Monsieur le Docteur **Eric BIETH**
Madame le Docteur **Sophie JULIA**

Président
Assesseur
Assesseur
Assesseur
Assesseur
Assesseur

TABLEAU du PERSONNEL HU
des Facultés de Médecine de l'Université Paul Sabatier
au 1^{er} septembre 2018

Professeurs Honoraires

Doyen Honoraire	M. CHAP Hugues	Professeur Honoraire	M. MANSAT Michel
Doyen Honoraire	M. GUIRAUD-CHAUMEIL Bernard	Professeur Honoraire	M. MASSIP Patrice
Doyen Honoraire	M. LAZORTHES Yves	Professeur Honoraire	Mme MARTY Nicole
Doyen Honoraire	M. PUEL Pierre	Professeur Honoraire	M. MAZIERES Bernard
Doyen Honoraire	M. ROUGE Daniël	Professeur Honoraire	M. MONROZIES Xavier
Doyen Honoraire	M. VINEL Jean-Pierre	Professeur Honoraire	M. MOSCOVICI Jacques
Professeur Honoraire	M. ABBAL Michel	Professeur Honoraire	M. MURAT
Professeur Honoraire	M. ADER Jean-Louis	Professeur Honoraire	M. OLIVES Jean-Pierre
Professeur Honoraire	M. ALBAREDE Jean-Louis	Professeur Honoraire	M. PASCAL Jean-Pierre
Professeur Honoraire	M. ARBUS Louis	Professeur Honoraire	M. PESSEY Jean-Jacques
Professeur Honoraire	M. ARLET Jacques	Professeur Honoraire	M. PLANTE Pierre
Professeur Honoraire	M. ARLET Philippe	Professeur Honoraire	M. PONTONNIER Georges
Professeur Honoraire	M. ARLET-SUAU Elisabeth	Professeur Honoraire	M. POURRAT Jacques
Professeur Honoraire	M. ARNE Jean-Louis	Professeur Honoraire	M. PRADERE Bernard
Professeur Honoraire	M. BARRET André	Professeur Honoraire	M. PRIS Jacques
Professeur Honoraire	M. BARTHE Philippe	Professeur Honoraire	Mme PUEL Jacqueline
Professeur Honoraire	M. BAYARD Francis	Professeur Honoraire	M. PUEL Pierre
Professeur Honoraire	M. BOCCALON Henri	Professeur Honoraire	M. PUJOL Michel
Professeur Honoraire	M. BONAFE Jean-Louis	Professeur Honoraire	M. QUERREU Denis
Professeur Honoraire	M. BONEU Bernard	Professeur Honoraire	M. RAILHAC Jean-Jacques
Professeur Honoraire	M. BOUNHOURE Jean-Paul	Professeur Honoraire	M. REGIS Henri
Professeur Honoraire	M. BOUTAULT Franck	Professeur Honoraire	M. REGNIER Claude
Professeur Honoraire	M. BUGAT Roland	Professeur Honoraire	M. REME Jean-Michel
Professeur Honoraire	M. CAHUZAC Jean-Philippe	Professeur Honoraire	M. ROCHE Henri
Professeur Honoraire	M. CARATERO Claude	Professeur Honoraire	M. ROICCICOLI Pierre
Professeur Honoraire	M. CARLES Pierre	Professeur Honoraire	M. ROLLAND Michel
Professeur Honoraire	M. CARRIERE Jean-Paul	Professeur Honoraire	M. ROUJE-LATRILLE Christian
Professeur Honoraire	M. CARTON Michel	Professeur Honoraire	M. RUMEAU Jean-Louis
Professeur Honoraire	M. CATHALA Bernard	Professeur Honoraire	M. SALVADOR Michel
Professeur Honoraire	M. CHABANON Gérard	Professeur Honoraire	M. SALVAYRE Robert
Professeur Honoraire	M. CHAMONTIN Bernard	Professeur Honoraire	M. SARRAMON Jean-Pierre
Professeur Honoraire	M. CHAP Hugues	Professeur Honoraire	M. SIMON Jacques
Professeur Honoraire	M. CHAVOIN Jean-Pierre	Professeur Honoraire	M. SUC Jean-Michel
Professeur Honoraire	M. CLANET Michel	Professeur Honoraire	M. THOUVENOT Jean-Paul
Professeur Honoraire	M. CONTE Jean	Professeur Honoraire	M. TREMOULET Michel
Professeur Honoraire	M. COSTAGLIOLA Michel	Professeur Honoraire	M. VALDIGUIE Pierre
Professeur Honoraire	M. COTONAT Jean	Professeur Honoraire	M. VAYSSE Philippe
Professeur Honoraire	M. DABERNAT Henri	Professeur Honoraire	M. VIRENQUE Christian
Professeur Honoraire	M. DALOUS Antoine	Professeur Honoraire	M. VOIGT Jean-Jacques
Professeur Honoraire	M. DALY-SCHVEITZER Nicolas		
Professeur Honoraire	M. DAVID Jean-Frédéric		
Professeur Honoraire	M. DELSOL Georges		
Professeur Honoraire	Mme DELISLE Marie-Bernadette		
Professeur Honoraire	Mme DIDIER Jacqueline		
Professeur Honoraire	M. DUCOS Jean		
Professeur Honoraire	M. DUFFAUT Michel		
Professeur Honoraire	M. DUPRE M.		
Professeur Honoraire	M. DURAND Dominique		
Professeur Honoraire associé	M. DUTAU Guy		
Professeur Honoraire	M. ESCANDE Michel		
Professeur Honoraire	M. ESCHAPASSE Henri		
Professeur Honoraire	M. ESCOURROU Jean		
Professeur Honoraire	M. ESQUERRE J.P.		
Professeur Honoraire	M. FABIE Michel		
Professeur Honoraire	M. FABRE Jean		
Professeur Honoraire	M. FOURNIAL Gérard		
Professeur Honoraire	M. FOURNIE Bernard		
Professeur Honoraire	M. FOURTANIER Gilles		
Professeur Honoraire	M. FRAYSSE Bernard		
Professeur Honoraire	M. FREXINOS Jacques		
Professeur Honoraire	Mme GENESTAL Michèle		
Professeur Honoraire	M. GERAUD Gilles		
Professeur Honoraire	M. GHISOLFI Jacques		
Professeur Honoraire	M. GOUZI Jean-Louis		
Professeur Honoraire	M. GUIRAUD-CHAUMEIL Bernard		
Professeur Honoraire	M. HOFF Jean		
Professeur Honoraire	M. JOFFRE Francis		
Professeur Honoraire	M. LACOMME Yves		
Professeur Honoraire	M. LAGARRIGUE Jacques		
Professeur Honoraire	Mme LARENG Marie-Blanche		
Professeur Honoraire	M. LARENG Louis		
Professeur Honoraire	M. LAURENT Guy		
Professeur Honoraire	M. LAZORTHES Franck		
Professeur Honoraire	M. LAZORTHES Yves		
Professeur Honoraire	M. LEOPHONTE Paul		
Professeur Honoraire	M. MAGNAVAL Jean-François		
Professeur Honoraire	M. MANELFE Claude		

Professeurs Émérites

Professeur ADER Jean-Louis
Professeur ALBAREDE Jean-Louis
Professeur ARBUS Louis
Professeur ARLET-SUAU Elisabeth
Professeur BOCCALON Henri
Professeur BONEU Bernard
Professeur CARATERO Claude
Professeur CHAMONTIN Bernard
Professeur CHAP Hugues
Professeur CONTÉ Jean
Professeur COSTAGLIOLA Michel
Professeur DABERNAT Henri
Professeur FRAYSSE Bernard
Professeur DELISLE Marie-Bernadette
Professeur GUIRAUD-CHAUMEIL Bernard
Professeur JOFFRE Francis
Professeur LAGARRIGUE Jacques
Professeur LARENG Louis
Professeur LAURENT Guy
Professeur LAZORTHES Yves
Professeur MAGNAVAL Jean-François
Professeur MANELFE Claude
Professeur MASSIP Patrice
Professeur MAZIERES Bernard
Professeur MOSCOVICI Jacques
Professeur MURAT
Professeur ROQUES-LATRILLE Christian
Professeur SALVAYRE Robert
Professeur SARRAMON Jean-Pierre
Professeur SIMON Jacques

FACULTE DE MEDECINE TOULOUSE-PURPAN

37 allées Jules Guesde - 31062 TOULOUSE Cedex

Doyen : D. CARRIE

P.U. - P.H.

P.U. - P.H.

Classe Exceptionnelle et 1ère classe

2ème classe

M. ADOUE Daniel (C.E)	Médecine Interne, Gériatrie
M. AMAR Jacques	Thérapeutique
M. ATTAL Michel (C.E)	Hématologie
M. AVET-LOISEAU Hervé	Hématologie, transfusion
Mme BEYNE-RAUZY Odile	Médecine Interne
M. BIRMES Philippe	Psychiatrie
M. BLANCHER Antoine	Immunologie (option Biologique)
M. BONNEVILLE Paul (C.E)	Chirurgie Orthopédique et Traumatologie
M. BOSSAVY Jean-Pierre (C.E)	Chirurgie Vasculaire
M. BRASSAT David	Neurologie
M. BROUCHET Laurent	Chirurgie Thoracique et cardio-vascul
M. BROUSSET Pierre (C.E)	Anatomie pathologique
M. CALVAS Patrick (C.E)	Généraliste
M. CARRERE Nicolas	Chirurgie Générale
M. CARRIE Didier (C.E)	Cardiologie
M. CHAIX Yves	Pédiatrie
M. CHAUVEAU Dominique	Néphrologie
M. CHOLLET François (C.E)	Neurologie
M. DAHAN Marcel (C.E)	Chirurgie Thoracique et Cardiaque
M. DE BOISSEZON Xavier	Médecine Physique et Réadapt Fonct.
M. DEGUINE Olivier (C.E)	Oto-rhino-laryngologie
M. DUQUOMMUN Bernard	Cancérologie
M. FERRIERES Jean (C.E)	Epidémiologie, Santé Publique
M. FOURCADE Olivier	Anesthésiologie
M. FOURNIÉ Pierre	Ophthalmologie
M. GAME Xavier	Urologie
M. GEERAERTS Thomas	Anesthésiologie et réanimation
M. IZOPET Jacques (C.E)	Bactériologie-Virologie
Mme LAMANT Laurence (C.E)	Anatomie Pathologique
M. LANG Thierry (C.E)	Biostatistiques et Informatique Médicale
M. LANGIN Dominique (C.E)	Nutrition
M. LAUWERS Frédéric	Anatomie
M. LAUQUE Dominique (C.E)	Médecine Interne
M. LIBLAU Roland (C.E)	Immunologie
M. MALAUAUD Bernard	Urologie
M. MANSAT Pierre	Chirurgie Orthopédique
M. MARCHOU Bruno	Maladies Infectieuses
M. MAZIERES Julien	Pneumologie
M. MOLINIER Laurent	Epidémiologie, Santé Publique
M. MONTASTRUC Jean-Louis (C.E)	Pharmacologie
Mme MOYAL Elisabeth	Cancérologie
Mme NOURHASHEMI Fatemeh (C.E)	Gériatrie
M. OSWALD Eric	Bactériologie-Virologie
M. PARANT Olivier	Gynécologie Obstétrique
M. PARIENTE Jérémie	Neurologie
M. PARINAUD Jean (C.E)	Biol. Du Develop. et de la Reprod.
M. PAUL Carle	Dermatologie
M. PAYOUX Pierre	Biophysique
M. PAYRASTRE Bernard (C.E)	Hématologie
M. PERON Jean-Marie	Hépatogastro-Entérologie
M. PERRET Bertrand (C.E)	Biochimie
M. RASCOL Olivier (C.E)	Pharmacologie
M. RECHER Christian	Hématologie
M. RISCHMANN Pascal	Urologie
M. RIVIERE Daniel (C.E)	Physiologie
M. SALES DE GAUZY Jérôme	Chirurgie Infantile
M. SALLES Jean-Pierre (C.E)	Pédiatrie
M. SANS Nicolas	Radiologie
Mme SELVES Janick	Anatomie et cytologie pathologiques
M. SERRE Guy (C.E)	Biologie Cellulaire
M. TELMON Norbert (C.E)	Médecine Légale
M. VINEL Jean-Pierre (C.E)	Hépatogastro-Entérologie

P.U. Médecine générale

M. OUSTRIC Stéphane

Mme BONGARD Vanina	Epidémiologie
M. BONNEVILLE Nicolas	Chirurgie orthopédique et traumatologique
M. BUREAU Christophe	Hépatogastro-Entéro
Mme CASPER Charlotte	Pédiatrie
Mme CHARPENTIER Sandrine	Médecine d'urgence
M. COGNARD Christophe	Neuroradiologie
M. LAIREZ Olivier	Biophysique et médecine nucléaire
M. LAROCHE Michel	Rhumatologie
M. LEOBON Bertrand	Chirurgie Thoracique et Cardiaque
M. LOPEZ Raphael	Anatomie
M. MARTIN-BLONDEL Guillaume	Maladies infectieuses, maladies tropicales
M. MARX Mathieu	Oto-rhino-laryngologie
M. MAS Emmanuel	Pédiatrie
M. OLIVOT Jean-Marc	Neurologie
M. PORTIER Guillaume	Chirurgie Digestive
M. RONCALLI Jérôme	Cardiologie
Mme RUYSSSEN-WITRAND Adeline	Rhumatologie
Mme SAVAGNER Frédérique	Biochimie et biologie moléculaire
M. SOL Jean-Christophe	Neurochirurgie
Mme TREMOLLIERES Florence	Biologie du développement
Mme VEZZOSI Delphine	Endocrinologie

P.U. Médecine générale

M. MESTHÉ Pierre

Professeur Associé Médecine générale

M. ABITTEBOUL Yves

M. POUTRAIN Jean-Christophe

Professeur Associé en Neurologie

Mme PAVY-LE TRACON Annie

FACULTE DE MEDECINE TOULOUSE-RANGUEIL

133, route de Narbonne - 31062 TOULOUSE Cedex

Doyen : E. SERRANO

P.U. - P.H. Classe Exceptionnelle et 1ère classe		P.U. - P.H. 2ème classe	
M. ACAR Philippe	Pédiatrie	M. ARBUS Christophe	Psychiatrie
M. ACCADBLED Franck	Chirurgie Infantile	M. AUSSEIL Jérôme	Biochimie et biologie moléculaire
M. ALRIC Laurent (C.E)	Médecine Interne	M. BERRY Antoine	Parasitologie
Mme ANDRIEU Sandrine	Epidémiologie	M. BONNEVILLE Fabrice	Radiologie
M. ARNAL Jean-François	Physiologie	M. BOUNES Vincent	Médecine d'urgence
Mme BERRY Isabelle (C.E)	Biophysique	Mme BOURNET Barbara	Gastro-entérologie
M. BUJAN Louis (C. E)	Urologie-Andrologie	M. CHAPUT Benoit	Chirurgie plastique et des brûlés
Mme BURA-RIVIERE Alessandra	Médecine Vasculaire	M. CHAYNES Patrick	Anatomie
M. BUSCAIL Louis (C.E)	Hépat-Gastro-Entérologie	Mme DALENC Florence	Cancérologie
M. CANTAGREL Alain (C.E)	Rhumatologie	M. DECRAMER Stéphane	Pédiatrie
M. CARON Philippe (C.E)	Endocrinologie	M. DELOBEL Pierre	Maladies Infectieuses
M. CHAUFOUR Xavier	Chirurgie Vasculaire	M. FAGUER Stanislas	Néphrologie
M. CHIRON Philippe (C.E)	Chirurgie Orthopédique et Traumatologie	M. FRANCHITTO Nicolas	Addictologie
M. CONSTANTIN Arnaud	Rhumatologie	M. GARRIDO-STÔWHAS Ignacio	Chirurgie Plastique
M. COURBON Frédéric	Biophysique	Mme GOMEZ-BROUCHET Anne-Muriel	Anatomie Pathologique
Mme COURTADE SAIDI Monique	Histologie Embryologie	M. HUYGHE Eric	Urologie
M. DAMBRIN Camille	Chirurgie Thoracique et Cardiovasculaire	Mme LAPRIE Anne	Radiothérapie
M. DELABESSE Eric	Hématologie	M. LAURENT Camille	Anatomie Pathologique
M. DELORD Jean-Pierre	Cancérologie	M. MARCHEIX Bertrand	Chirurgie thoracique et cardiovasculaire
M. DIDIER Alain (C.E)	Pneumologie	M. MAURY Jean-Philippe	Cardiologie
Mme DULY-BOUHANICK Béatrice	Thérapeutique	M. MEYER Nicolas	Dermatologie
M. ELBAZ Meyer	Cardiologie	M. MUSCARI Fabrice	Chirurgie Digestive
M. GALINIER Michel (C.E)	Cardiologie	M. REINA Nicolas	Chirurgie orthopédique et traumatologique
M. GLOCK Yves (C.E)	Chirurgie Cardio-Vasculaire	M. SILVA SIFONTES Stein	Réanimation
M. GOURDY Pierre	Endocrinologie	M. SOLER Vincent	Ophthalmologie
M. GRAND Alain (C.E)	Epidémiologie. Eco, de la Santé et Prévention.	Mme SOMMET Agnès	Pharmacologie
M. GROLEAU RAOUX Jean-Louis	Chirurgie plastique	Mme SOTO-MARTIN Maria-Eugénia	Gériatrie et biologie du vieillissement
Mme GUIMBAUD Rosine	Cancérologie	M. TACK Ivan	Physiologie
Mme HANAIRE Hélène (C.E)	Endocrinologie	M. VERGEZ Sébastien	Oto-rhino-laryngologie
M. KAMAR Nassim	Néphrologie	M. YSEBAERT Loïc	Hématologie
M. LARRUE Vincent	Neurologie		
M. LEVADE Thierry (C.E)	Biochimie		
M. MALECAZE François (C.E)	Ophthalmologie	P.U. Médecine générale	
M. MARQUE Philippe	Médecine Physique et Réadaptation	Mme ROUGE-BUGAT Marie-Eve	
Mme MAZEREEUW Juliette	Dermatologie		
M. MINVILLE Vincent	Anesthésiologie Réanimation	Professeur Associé de Médecine Générale	
M. OTAL Philippe	Radiologie	M. BOYER Pierre	
M. RAYNAUD Jean-Philippa (C.E)	Psychiatrie Infantile		
M. RITZ Patrick	Nutrition	Professeur Associé en Pédiatrie	
M. ROLLAND Yves (C.E)	Gériatrie	Mme CLAUDET Isabelle	
M. ROUGE Daniel (C.E)	Médecine Légale		
M. ROUSSEAU Hervé (C.E)	Radiologie		
M. ROUX Franck-Emmanuel	Neurochirurgie		
M. SAILLER Laurent	Médecine Interne		
M. SCHMITT Laurent (C.E)	Psychiatrie		
M. SENARD Jean-Michel (C.E)	Pharmacologie		
M. SERRANO Elié (C.E)	Oto-rhino-laryngologie		
M. SOULAT Jean-Marc	Médecine du Travail		
M. SOULIE Michel (C.E)	Urologie		
M. SUG Bertrand	Chirurgie Digestive		
Mme TAUBER Marie-Thérèse (C.E)	Pédiatrie		
Mme URO-COSTÉ Emmanuelle	Anatomie Pathologique		
M. VAYSSIERE Christophe	Gynécologie Obstétrique		
M. VELLAS Bruno (C.E)	Gériatrie		
Professeur Associé de Médecine Générale			
M. STILLMUNKES André			

M.C.U. - P.H.

M. ABBO Olivier	Chirurgie infantile
M. APOIL Pol Andra	Immunologie
Mme ARNAUD Catherine	Epidémiologie
Mme BERTOLI Sarah	Hématologie, transfusion
M. BIETH Eric	Génétique
Mme CASPAR BAUGUIL Sylvie	Nutrition
Mme CASSAGNE Myriam	Ophthalmologie
Mme CASSAING Sophie	Parasitologie
M. CAVAIGNAC Etienne	Chirurgie orthopédique et traumatologie
Mme CHANTALAT Elodie	Anatomie
M. CONGY Nicolas	Immunologie
Mme COURBON Christine	Pharmacologie
Mme DAMASE Christine	Pharmacologie
Mme de GLISEZENSKY Isabelle	Physiologie
Mme DE MAS Véronique	Hématologie
M. DUBOIS Damien	Bactériologie Virologie Hygiène
Mme FILLAUX Judith	Parasitologie
M. GANTET Pierre	Biophysique
Mme GENNERO Isabelle	Biochimie
Mme GENOUX Annelise	Biochimie et biologie moléculaire
M. HAMDJ Salouane	Biochimie
Mme HITZEL Anne	Biophysique
M. IRIART Xavier	Parasitologie et mycologie
Mme JONCA Nathalie	Biologie cellulaire
M. KIRZIN Sylvain	Chirurgie générale
Mme LAPEYRE-MESTRE Maryse	Pharmacologie
M. LHERMUSIER Thibault	Cardiologie
M. LHOMME Sébastien	Bactériologie-virologie
Mme MONTASTIER Emilie	Nutrition
Mme MOREAU Marion	Physiologie
Mme NOGUEIRA M.L.	Biologie Cellulaire
M. PILLARD Fabien	Physiologie
Mme PUISSANT Bénédicte	Immunologie
Mme RAYMOND Stéphanie	Bactériologie Virologie Hygiène
Mme SABOURDY Frédérique	Biochimie
Mme SAUNE Karine	Bactériologie Virologie
M. TAFANI Jean-André	Biophysique
M. TREINER Emmanuel	Immunologie
Mme VAYSSE Charlotte	Cancérologie
M. VIDAL Fabien	Gynécologie obstétrique

M.C.U. Médecine générale

M. BRILLAC Thierry
Mme DUPOUY Julie

M.C.U. - P.H.

Mme ABRAVANEL Florence	Bactériologie Virologie Hygiène
Mme BASSET Céline	Cytologie et histologie
Mme CAMARE Caroline	Biochimie et biologie moléculaire
M. CAMBUS Jean-Pierre	Hématologie
Mme CANTERO Anne-Valérie	Biochimie
Mme CARFAGNA Luana	Pédiatrie
Mme CASSOL Emmanuelle	Biophysique
Mme CAUSSE Elizabeth	Biochimie
M. CHASSAING Nicolas	Génétique
M. CLAVEL Cyril	Biologie Cellulaire
Mme COLOMBAT Magali	Anatomie et cytologie pathologiques
Mme CORRE Jill	Hématologie
M. DE BONNECAZE Guillaume	Anatomie
M. DEDOUIT Fabrice	Médecine Légale
M. DELPLA Pierre-André	Médecine Légale
M. DESPAS Fabien	Pharmacologie
M. EDOUARD Thomas	Pédiatrie
Mme ESQUIROL Yolande	Médecine du travail
Mme EVRARD Solène	Histologie, embryologie et cytologie
Mme GALINIER Anne	Nutrition
Mme GALLINI Adeline	Epidémiologie
Mme GARDETTE Virginie	Epidémiologie
M. GASQ David	Physiologie
M. GATIMEL Nicolas	Médecine de la reproduction
Mme GRARE Marion	Bactériologie Virologie Hygiène
Mme GUILBEAU-FRUGIER Céline	Anatomie Pathologique
M. GUILLEMINAULT Laurent	Pneumologie
Mme GUYONNET Sophie	Nutrition
M. HERIN Fabrice	Médecine et santé au travail
Mme INGUENEAU Cécile	Biochimie
M. LEANDRI Roger	Biologie du dével et de la reproduction
M. LEPAGE Benoit	Biostatistiques et Informatique médicale
Mme MAUPAS SCHWALM Françoise	Biochimie
M. MIEUSSET Roger	Biologie du dével et de la reproduction
M. MOLLIS Guillaume	Médecine interne
Mme NASR Nathalie	Neurologie
M. RIMAILHO Jacques	Anatomie et Chirurgie Générale
M. RONGIERES Michel	Anatomie - Chirurgie orthopédique
Mme VALLET Marion	Physiologie
M. VERGEZ François	Hématologie

M.C.U. Médecine générale

M. BISMUTH Michel
Mme ESCOURROU Brigitte

Maitres de Conférences Associés de Médecine Générale

Dr FREYENS Anne
Dr IRI-DELAHAYE Motoko
Dr CHICOULEA Bruno

Dr BIREBENT Jordan
Dr BOURGEOIS Odile
Dr LATROUS Leila

Dédicaces aux membres du jury

Je remercie **Monsieur le Professeur Patrick Calvas**,

Président du Jury,

Professeur des Universités, Praticien Hospitalier,

Service de Génétique médicale, Hôpital Toulouse Purpan,

qui m'a accueillie pendant mon internat au sein de son service. Grâce à lui j'ai pu poursuivre ma formation en génétique en conciliant avec bonheur recherche théorique et appliquée.

Vous me faites l'honneur de présider ce Jury de thèse après m'avoir conseillé, guidé et enseigné votre savoir pendant tout mon internat.

Je remercie **Monsieur le Professeur Gilles FAVRE**,

Professeur des Universités, Praticien Hospitalier,

Directeur du Cancéropôle Grand Sud-Ouest,

Directeur du CRCT (Centre de Recherches en Cancérologie de Toulouse), Oncopôle de Toulouse,

Chef de service du laboratoire de Biologie Médicale Oncologique, IUC-T (Institut Universitaire du Cancer de Toulouse),

qui a bien voulu accepter ma présence pendant un mois au sein de son laboratoire à l'IUC-T afin que je puisse approfondir ma formation sur le séquençage haut débit. Ce stage m'a aidée de façon considérable à avoir un regard critique sur la stratégie de développement du test DI117, au cœur de ce projet de thèse.

Je vous remercie de faire partie de mon Jury de thèse afin d'apporter votre expertise avérée sur les techniques de séquençage haut débit.

Je remercie **Monsieur le Professeur Yves CHAIX**,

Professeur des Universités, Praticien Hospitalier,

Service de Neurologie Pédiatrique, Hôpital Toulouse Purpan,

d'avoir bien voulu accepter de faire partie de mon Jury de thèse afin d'y apporter sa connaissance et son expérience sur la neuro-pédiatrie et plus précisément sur la déficience intellectuelle, objet de mon projet de recherche.

Je remercie **Madame le Professeur Bettina COUDERC**,

Professeur des Universités, Faculté de Pharmacie de Toulouse,

Chercheur au CRCT (Centre de Recherches en Cancérologie de Toulouse), Oncopôle de Toulouse,

pour sa présence au sein de mon Jury de thèse qui nous apporte une approche de la recherche scientifique hors du circuit hospitalier et donc tout à fait intéressante et instructive.

Je remercie **Monsieur le Docteur Eric BIETH**,

Directeur de thèse,

Maître de Conférences des Universités, Praticien Hospitalier,

Service de Génétique médicale, Hôpital Toulouse Purpan,

pour la qualité de son encadrement sur l'élaboration de cette thèse.

Merci pour vos conseils, pour le temps que vous m'avez consacré et pour votre implication. J'espère que ce travail sera à la hauteur de vos espérances, et de la confiance que vous m'avez accordée.

Je remercie **Madame le Docteur Sophie JULIA**,

Praticien hospitalier,

Service de Génétique médicale, Hôpital Toulouse Purpan,

de m'avoir permis d'assister à quelques consultations de génétique, un autre angle plus concret de notre profession où l'humain prend toute sa place, et nous conforte sur l'importance de l'aboutissement de nos recherches.

Dédicaces personnelles

Je tiens à remercier très chaleureusement Madame Véronique GASTON pour sa grande disponibilité, pour son aide tout au long de ce projet, et pour m'avoir fait partager son savoir-faire technique avec beaucoup de pédagogie.

Je remercie également tout particulièrement le Docteur Nicolas CHASSAING, tuteur pour la spécialisation de mon DES en génétique moléculaire, pour sa gentillesse et sa bienveillance, ainsi que pour les nombreux conseils qu'il m'a prodigués au cours de ces années d'internat.

Je remercie énormément Christophe HABIB, notre bio-informaticien, pour le temps précieux qu'il m'a consacré et pour son implication remarquable dans ce travail.

Je tiens aussi à témoigner ma reconnaissance à toutes les personnes qui m'ont aidées, de près ou de loin, à réaliser ce projet.

Je remercie l'équipe du service de Génétique médicale et l'équipe du Laboratoire de Génétique de m'avoir accueillie, et de l'aide qu'ils m'ont apportée à un moment ou à un autre au cours de mon internat.

A Quentin, ma tendre moitié, mon rayon de soleil, qui a su me réconforter et me remotiver dans les moments difficiles, et dont l'aide pour parfaire ce manuscrit a été plus que précieuse. Sans toi, ta bienveillance, et tes astucieux conseils en informatique, je ne serais pas en train d'écrire ces quelques lignes aujourd'hui... Encore merci milles fois

A mes parents, Martine et Dominique,
qui m'ont toujours soutenue et encouragée,
qui ont cru en moi et m'ont épaulée tout du long de mes études, notamment lors des concours de PCEM1 et de l'ECN,
relecteurs assidus de mes écritures,
je ne serais pas là où j'en suis sans eux aujourd'hui...

A mes grands-parents,
Mamie chat et Papy chat, Mamie Jo et Papy Hubert, Mamie Geneviève et Papy Maurice,
qui ont toujours été là lorsque j'avais besoin d'eux, depuis que j'ai pointé le bout de mon nez et encore aujourd'hui, je ne les remercierai jamais assez pour tout cela...

A mon fidèle ami à quatre pattes, mon confident, adepte de mes répétitions orales,
même si la génétique doit être un sujet encore un peu flou pour lui...

Je dédie cette thèse

à Quentin,

à mes parents,

à mes grands-parents,

et à mon Gucci.

Table des matières

Dédicaces aux membres du jury.....	7
Dédicaces personnelles	9
Table des matières.....	12
Liste des documents en Annexe	15
Table des Illustrations.....	16
Figures	16
Tableaux.....	20
Liste des abréviations	22
Lexique.....	25
Introduction.....	32
I. La déficience intellectuelle (DI)	32
1. Prévalence et définitions de la DI	33
2. Hétérogénéité des présentations cliniques de la Déficience Intellectuelle	35
3. Hétérogénéité des étiologies des DI.....	36
4. Classifications des gènes impliqués dans la DI	39
5. Bilan génétique en 2019 pour le diagnostic étiologique de la DI	44
II. Le séquençage de nouvelle génération (NGS)	49
1. Evolution et principes du séquençage de l'ADN.....	49
2. Principes méthodologiques du séquençage NGS de 2 ^{ème} génération : les différentes phases techniques, bases bio-informatiques de l'analyse des données et terminologie	76
Matériel et méthodes.....	81
I. Sélection d'une cohorte de patients témoins	81
II. Recueil des données de séquençage (fichier .bam, .bai, et .vcf) de certains patients établies dans d'autres laboratoires	84

III.	Extraction des ADN	85
IV.	Mise en pratique expérimentale du test DI44.....	85
1.	Caractéristiques techniques du test DI44.....	85
2.	Préparation de la librairie jusqu'au séquençage sur le PGM™ Ion Torrent™	86
V.	Conception des sondes de capture du test DI117 pour un séquençage sur technologie Illumina	95
VI.	Préparation de la librairie du test DI117 en capture et séquençage sur le MiSeq Illumina	96
1.	Préparation de la librairie	96
2.	Séquençage par le MiSeq Illumina avec la génération de données par le logiciel associé au séquenceur	101
VII.	Traitement bio-informatique à partir des fichiers FASTQ	102
	Résultats	103
	Préambule.....	103
I.	Développement d'une méthode de séquençage de nouvelle génération de 117 gènes pour le diagnostic de la déficience intellectuelle (DI)	104
1.	Choix de la technologie utilisée	104
2.	Sélection des gènes cibles.....	106
3.	Conception des sondes de capture.....	107
II.	Comparaison des performances du test DI117 (Illumina) avec le test DI44 (Ion Torrent).....	109
1.	Performances théoriques des tests NGS DI44 versus DI117 pour les 44 gènes communs.....	109
2.	Performances réelles déduites de la pratique expérimentale	123
3.	Coût, temps technique	176
III.	Résultats du test NGS DI117 appliqué à 15 patients atteints de DI	177
1.	Profondeur de lecture du séquençage réalisé.....	178
2.	Couverture réelle des régions ciblées (CDS+ /- 25)	180

3.	Détection des variations « vraies positives »	184
IV.	Analyse comparative des résultats du test DI117 face à d'autres tests NGS dédiés à la DI 187	
1.	Comparaison des variants détectés par le test DI117 avec ceux détectés par les autres tests.....	190
2.	Comparaison des variants détectés avec le pipeline du test DI117 par rapport au pipeline des autres tests	196
3.	Hypothèses expliquant la différence de variants observés entre le test DI117 et les tests des autres CHU	200
	Discussion	212
	Pourquoi un panel de 117 gènes ?	214
	Le test DI117 est-il au moins aussi performant que le test DI44 ?.....	216
	Le test DI117 satisfait-il pleinement aux exigences de qualité requises pour une application à visée diagnostique ?.....	221
	Le test DI117 est-il perfectible ?	224
	Quelles sont les limites d'un test NGS de 2ème génération ?	225
	Quelles sont les perspectives pour le diagnostic génétique de la DI ?.....	227
	Conclusion.....	229
	Références bibliographiques	230
	Annexes	234

Liste des documents en Annexe

Annexe 1 : Tableau récapitulatif des gènes inclus dans les différents tests NGS de DI....	235
Annexe 2 : Feuille de travail récapitulant le protocole utilisé pour le test DI44.....	239
Annexe 3 : Feuille de travail récapitulant le protocole utilisé pour le test DI117.....	245
Annexe 4 : Tableau récapitulant les 117 gènes inclus dans le test DI117.....	258
Annexe 5 : Estimation du rendement diagnostique des tests DI44 et DI117	268
Annexe 6 : Tableau précisant le nombre de "Sanger" nécessaires pour compléter la couverture théorique sur CDS \pm 6 pb des tests DI44 et DI117	269
Annexe 7 : Tableau détaillant la couverture déduite de la pratique expérimentale des tests DI44 et DI117 sur les régions CDS \pm 6 pb et CDS \pm 25 pb	271
Annexe 8 : Tableau des taux de couverture déduits de la pratique expérimentale des tests DI44 et DI117 sur les régions CDS \pm 6 pb et CDS \pm 25 pb	277
Annexe 9 : Tableau des profondeurs de lecture moyennes par gène, patient par patient, sur les 44 gènes (CDS \pm 25), avec calcul de la profondeur de lecture moyenne par gène pour l'ensemble des 6 patients, et avec indication de la profondeur de lecture moyenne pour l'ensemble des 44 gènes par patient (PLM) : Test DI44 comparé au Test DI117.....	278
Annexe 10 : Nombre de variants retenus selon les différents filtres bio-informatiques, dont nombre de variants <i>frameshift</i> (FS), sur les 44 gènes et sur les 6 patients, pour les deux tests DI44 et DI117	279
Annexe 11 : Etude du coût des réactifs et du temps technique nécessaires pour le test DI44.....	280
Annexe 12 : Etude du coût des réactifs et du temps technique nécessaires pour le test DI117.....	281
Annexe 13 : Tableau des profondeurs de lecture moyennes par gène, patient par patient, sur les 117 gènes (CDS \pm 25) ; avec calcul de la profondeur de lecture moyenne par gène pour l'ensemble des 15 patients, et avec indication de la profondeur de lecture moyenne pour l'ensemble des 117 gènes par patient (PLM) pour le Test DI117.....	282
Annexe 14 : Tableau des taux de couverture $>30X$ sur les 117 gènes (CDS \pm 25) du test DI117,détaillés par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 15 patients.....	285

Table des Illustrations

Figures

Figure 1 : Algorithme décisionnel proposé par l'ANPGM (Association Nationale des Praticiens de Génétique Moléculaire) pour guider les investigations du bilan génétique dans le cadre d'un diagnostic clinique de déficience intellectuelle.....	48
Figure 2 : Profil de migration après électrophorèse capillaire.....	51
Figure 3 : Principe du séquençage par la méthode de Sanger.....	51
Figure 4 : Photographie du séquenceur ABI 3500XL présent au laboratoire.....	52
Figure 5 : Principe de la PCR en émulsion.....	56
Figure 6 : Principe du pyroséquençage.....	57
Figure 7 : Encodage des bases pour le séquençage SOLiD.....	59
Figure 8 : Les cycles de séquençage de la technologie SOLiD.....	60
Figure 9 : Amplification par pont (<i>bridge PCR</i>) de la technologie <i>Illumina</i> (images extraites des documents <i>Illumina</i>).....	62
Figure 10 : Principe du séquençage de l'ADN avec la technologie <i>Solexa/Illumina</i>	66
Figure 11 : Principe du séquence par mesure du pH de la technologie Ion Torrent™.....	68
Figure 12 : Représentation des variations de pH mesurées lors du séquençage par la technologie Ion Torrent™.....	68
Figure 13 : Principales caractéristiques des séquenceurs NGS de 2 ^{ème} génération.....	70
Figure 14 : Principe du séquençage de la technologie de <i>Pacific Biosciences (PacBio)</i>	74
Figure 15 : Principe du séquençage par nanopore d' <i>Oxford Nanopore Technologies</i>	74
Figure 16 : Préparation de la librairie avec la technologie <i>Ampliseq™</i> de <i>Thermo Fisher Scientific</i>	87
Figure 17 : Photographie du Bioanalyseur 2100 (Agilent) et de la puce utilisée, et visualisation d'un exemple de résultats obtenus pour un échantillon donné.....	88
Figure 18 : Préparation de la matrice de séquençage : étape d'amplification clonale par PCR en émulsion (emPCR) réalisée sur le système Ion OneTouch™ 2 (OT2).....	90
Figure 19 : Configuration de microréacteur souhaitée parmi les différentes possibilités..	90

Figure 20 : Préparation de la matrice de séquençage : étape d'enrichissement en ISP contenant des fragments d'ADN réalisée sur le module Ion OneTouch™ ES (ES)	91
Figure 21 : Etape de séquençage réalisée par le PGM™ Ion Torrent™	92
Figure 22 : Photographie du séquenceur PGM™ Ion Torrent™	93
Figure 23 : Génération des données de séquençage par le logiciel du PGM™ Ion Torrent™	94
Figure 24 : Préparation de la librairie avec la technologie SureSelectQXT d'Agilent.....	98
Figure 25 : Combinaison des librairies et préparation pour le séquençage	100
Figure 26 : Photographie du séquenceur MiSeq Illumina	101
Figure 27 : Graphique représentant la couverture théorique des tests DI44 et DI117 pour les 44 gènes (CDS ± 25 pb).....	115
Figure 28 : Graphique représentant la différence (en %) entre les couvertures théoriques pour les 44 gènes (CDS ± 25 pb) : Test DI117 comparé au Test DI44	116
Figure 29 : Graphique représentant la couverture théorique des tests DI44 et DI117 pour les 44 gènes (CDS ± 6 pb).....	119
Figure 30 : Graphique représentant la différence (en %) entre les couvertures théoriques pour les 44 gènes (CDS ± 6 pb) : Test DI117 comparé au Test DI44	120
Figure 31 : Graphique représentant le nombre de "Sanger" nécessaire(s) pour compléter la couverture théorique des 44 gènes (sur CDS ± 6) : Test DI44 versus Test DI117	122
Figure 32 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le <i>run</i> de séquençage du Test DI44 réalisé sur le séquenceur PGM™ Ion Torrent™	125
Figure 33 : Graphique représentant la fraction couverte des bases ciblées (avec la région cible limitée aux 44 gènes communs avec le Test DI44) en fonction de la profondeur de lecture pour le 1er <i>run</i> de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina	126
Figure 34 : Graphique représentant la fraction couverte des bases ciblées (avec la région cible limitée aux 44 gènes communs avec le Test DI44) en fonction de la profondeur de lecture pour le 2 ^{ème} <i>run</i> de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina	127
Figure 35 : Graphique représentant la couverture minimale en pratique des tests DI44 et DI117 pour les 44 gènes (CDS ± 25 pb)	130

Figure 36 : Graphique représentant la différence (en %) entre les couvertures minimales en pratique pour les 44 gènes (CDS \pm 25 pb) : Test DI117 comparé au Test DI44.....	131
Figure 37 : Graphique représentant le nombre de "Sanger" nécessaires pour compléter la couverture réelle des 44 gènes (sur CDS \pm 6) : Test DI44 comparé au Test DI117	138
Figure 38 : Profondeurs de lecture moyennes par gène pour les 44 gènes (CDS \pm 25) : Test DI44 comparé au Test DI117 (n = 6 patients).....	142
Figure 39 : Nombre de variants retenus sur les 44 gènes avec le Filtre 1 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%) retrouvés par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117	146
Figure 40 : Nombre de variants retenus sur les 44 gènes, en excluant les variants <i>frameshift</i> , avec le Filtre 1 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%) par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117 ...	148
Figure 41 : Nombre de variants retenus sur les 44 gènes, avec le Filtre 2 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%, et fréquence < 4 / run), retrouvés par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117	150
Figure 42 : Visualisation sur le logiciel IGV du variant d' <i>ARID1B</i> (NM_020732.3:c.228delC) retenu par le Filtre 2 avec 2 VC pour le test DI44, et non retrouvé avec le test DI117	154
Figure 43 : Représentation sur IGV du variant de <i>GRIN2B</i> (NM_000834.3:c.1341C>T) retenu par le Filtre 2 avec 2 VC pour le test DI117	155
Figure 44 : Visualisation sur IGV du variant de <i>SMARCA2</i> (NM_003070.4:c.683A>C) pour les 2 tests, bien qu'étant dans un défaut de design du test DI117.....	158
Figure 45 : Représentation sur IGV du variant d' <i>ANKRD11</i> (NM_001256182.1:c.6787C>T) retenu par le Filtre 2 avec 2 VC pour le test DI44 et par le Filtre 2 avec 1 VC pour le test DI117.....	159
Figure 46 : Représentation sur IGV du variant de <i>RAI1</i> (NM_030665.3:c.864_872delGCAGCAGCA) retenu par le Filtre 2 avec 2 VC pour le test DI44, et retrouvé mal annoté dans le Filtre 2 avec 1 VC pour le test DI117.....	160
Figure 47 : Représentation sur IGV du variant de <i>GRIN1</i> (NM_007327.3:c.1045G>T) retenu par le Filtre 2 avec 2 VC pour le test DI117, et décelé par le test DI44 mais éliminé par le Filtre 1 à cause d'une VAF < 35%.....	163
Figure 48 : Représentation sur IGV du variant d' <i>EP300</i> (NM_001429.3:c.2242-4delT) retenu par le Filtre 2 avec 2 VC pour le test DI117	164

Figure 49 : Représentation sur IGV du variant de <i>GRIN2A</i> (NM_000833.4:c.1275G>A) retenu par le Filtre 2 avec 2 VC pour le test DI117, et présent dans une région de <i>Gap 30X</i> pour le test DI44	167
Figure 50 : Représentation sur IGV du variant d' <i>ANKRD11</i> (NM_001256182.1:c.7128C>G) retenu par le Filtre 2 avec 2 VC pour le test DI117 et par le Filtre 2 avec 1 VC pour le test DI44.....	170
Figure 51 : Représentation sur IGV du variant de <i>GRIN2A</i> (NM_000833.4:c.1275G>A) retenu par le Filtre 2 avec 2 VC pour le test DI117, et présent dans une région de <i>Gap 30X</i> pour le test DI44	173
Figure 52 : Profondeurs de lecture moyennes par gène du test DI117 pour les 117 gènes (CDS ± 25) (n = 15 patients)	179
Figure 53 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le 1 ^{er} <i>run</i> de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina	181
Figure 54 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le 2 ^{ème} <i>run</i> de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina	182
Figure 55 : Schéma représentant le moment d'obtention du fichier BAM et du fichier VCF lors d'un test NGS.....	189
Figure 56 : Schéma expliquant les analyses et les filtrages réalisés lors de la comparaison des variants détectés par le test DI117 avec ceux décelés par les autres tests NGS de DI	191
Figure 57 : Diagrammes de Venn représentant la comparaison entre les variants détectés par le test DI117 et ceux décelés par le test extérieur (selon les filtrages décrits)	193
Figure 58 : Diagrammes de Venn représentant la comparaison entre les variants non synonymes situés dans les régions codantes et les variants localisés dans les régions d'épissage détectés par le test DI117 et ceux décelés par le test extérieur (selon les filtrages décrits).....	195
Figure 59 : Schéma expliquant les analyses et les filtrages réalisés lors de la comparaison des variants détectés avec le pipeline du test DI117 par rapport à ceux décelés par le pipeline des autres tests.....	197
Figure 60 : Diagrammes de Venn représentant la comparaison entre les variants non synonymes situés dans les régions codantes et les variants localisés dans les régions	

d'épissage détectés par le pipeline du test DI117 et ceux décelés par le pipeline du test extérieur (selon les filtrages décrits)	199
Figure 61 : Représentation sur IGV du variant du gène <i>SLC6A8</i> (NM_005629.3:c.-5A>G) du patient 3 situé dans une région ayant une profondeur de lecture < à 30X dans le test extérieur et une profondeur de lecture > à 100X avec le test DI117	207
Figure 62 : Visualisation sur IGV du variant retrouvé chez le patient 10 dans le gène <i>KMT2D</i> (NM_003482.3 :c.13885A>C) avec le test extérieur, mais non décelé avec le test DI117.....	208
Figure 63 : Visualisation sur IGV du variant retrouvé chez le patient 10 dans le gène <i>DLG3</i> (NM_021120.3 :c.202A>C) avec le test extérieur, mais non décelé avec le test DI117 ...	209
Figure 64 : Représentation sur IGV du variant du gène <i>SCN8A</i> (NM_014191.3:c.2545-7dupT) du patient 3, probable variant faux-positif dû à une erreur de séquençage, pour le test extérieur et pour le test DI117	210
Figure 65 : Visualisation sur IGV du variant du gène <i>ARID1B</i> (NM_020732.3 :c.954_956delAGG) du patient 4 pour le test extérieur et pour le test DI117.....	211

Tableaux

Tableau 1 : Principales caractéristiques des séquenceurs NGS de 3 ^{ème} génération.....	75
Tableau 2 : Présentation des 16 patients témoins de notre cohorte	82
Tableau 3 : Tableau récapitulatif des caractéristiques techniques des différents tests	84
Tableau 4 : Couverture théorique des tests DI44 et DI117 sur les régions CDS ± 6 pb et CDS ± 25 pb	111
Tableau 5 : Représentation des taux de couverture > 30X par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 6 patients, sur les 44 gènes (CDS ± 25) pour le Test DI44	133
Tableau 6 : Représentation des taux de couverture > 30X par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 6 patients, sur les 44 gènes (CDS ± 25) pour le Test DI117	134

Tableau 7 : Représentation des taux de couverture réelle minimale et maximale par gène sur les régions CDS \pm 6 pb et CDS \pm 25 pb des tests DI44 et DI117, avec indication du nombre de "Sanger" nécessaires pour compléter la couverture sur CDS \pm 6 pb	136
Tableau 8 : Tableau récapitulant la PLM (profondeur de lecture moyenne sur les régions CDS \pm 25 de l'ensemble des 44 gènes) de chaque patient pour chacun des 2 tests, DI44 et DI117.....	141
Tableau 9 : Analyse du variant confirmé en séquençage Sanger causal de chaque patient	144
Tableau 10 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 1.....	153
Tableau 11 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 2.....	157
Tableau 12 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 3.....	162
Tableau 13 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 4.....	166
Tableau 14 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 5.....	169
Tableau 15 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 6.....	172
Tableau 16 : Tableau récapitulant les variants retenus par le « Filtre 2 - 2 VC » qui diffèrent entre les 2 tests, Test DI44 et Test DI117, et les hypothèses émises pour expliquer ces écarts.....	175
Tableau 17 : Tableau récapitulant la PLM (profondeur de lecture moyenne sur les régions CDS \pm 25 de l'ensemble des 117 gènes) de chaque patient pour le Test DI117	178
Tableau 18 : Tableau des moyennes des taux de couverture > 30X du Test DI117 pour les 117 gènes (CDS \pm 25) (n = 15 patients).....	183
Tableau 19 : Analyse du variant confirmé causal de chaque patient et des variants non causaux détectés en séquençage Sanger	186
Tableau 20 : Tableau récapitulant les caractéristiques techniques des différents tests NGS réalisés chez les patients étudiés	188
Tableau 21 : Tableau récapitulant les variants décelés qui diffèrent et les hypothèses émises pour expliquer ces écarts de détection.....	202

Liste des abréviations

AAIDD : *American Association on Intellectual and Developmental Disabilities*

ACPA (ou *CGH array*) : Analyse chromosomique sur puce à ADN

AD : Autosomique dominant

ANPGM : Association nationale des praticiens de génétique moléculaire

APA : *American Psychiatric Association*

AR : Autosomique récessif

ADN : Acide désoxyribonucléique

ARN : Acide ribo-nucléique

ARNm : ARN messenger

BAM : *Binary Alignment/Map*

CCD (*Charge Coupled Device*) : Dispositif à transfert de charge

CDS (*Coding DNA Sequence*) : Séquence codante de l'ADN

CIM : Classification internationale des maladies

CHU : Centre hospitalo-universitaire

CMV : Cytomégalovirus

CNV (*Copy Number Variation*) : Variation du nombre de copies

ddNTP : Didésoxyribonucléotide tri-phosphate

dNTP : Désoxyribonucléotide tri-phosphate

DI : Déficience intellectuelle (substantif) ou atteint de déficience intellectuelle (adjectif)

DI-S : Déficience intellectuelle syndromique

DI-NS : Déficience intellectuelle non syndromique

DPN : Diagnostic prénatal

DS : Déviation standard

DSM : *Diagnostic and Statistical Manual of Mental Disorders*

EDTA : acide éthylène-diamine-tétraacétique

FC : Flow Cell

FISH (*Fluorescent In-Situ Hybridization*) : Hybridation *in situ* fluorescente

FS (*Frameshift*) : Variant décalant le cadre de lecture

HGVS : *Human Genome Variation Society*

IGV : *Integrative Genomics Viewer*

Indel : petite Insertion / Délétion (aussi appelée delins)

ISP : *Ion Sphere Particules*

Kb : Kilobases (1 000 pb)

LBM : Laboratoire de biologie médicale

LX : Lié à l'X

Mb : Mégabases (1 000 000 pb)

NGS (*Next Generation Sequencing*) : Séquençage de nouvelle génération ou séquençage massif en parallèle (MPS)

OMIM® : *Online Mendelian Inheritance in Man*®

OMS : Organisation Mondiale de la Santé

Pb : Paire de bases

PCR (*Polymerase Chain Reaction*) : Réaction de polymérisation en chaîne

pH : Potentiel hydrogène

PLM : Profondeur de lecture moyenne

QI : Quotient intellectuel

SAM : *Sequence Alignment/Map*

TSA : Troubles du Spectre Autistique

« Sanger » : Amplicon séquencé par la technique de Sanger

SNP (*Single Nucleotide Polymorphism*) : Polymorphisme

SNV (*Single Nucleotide Variant*) : Variation d'un nucléotide

SV (*Structural Variation*) : Variations de structure

VAF (*Variant Allele Frequency*) : Fréquence allélique du variant

VC : *Variant Caller*

VCF : *Variant Call Format*

WES (*Whole Exome Sequencing*) : Séquençage de l'exome entier

WGS (*Whole Genome Sequencing*) : Séquençage du génome entier

ZMW : *Zero-Mode Waveguide*

Lexique

Assemblage de novo des génomes : assemblage de données de séquences d'un génome inconnu afin d'obtenir la séquence d'un organisme pour lequel il n'existe pas une séquence de référence dans les bases de données

Allèle : variante d'un gène, résultant d'une variation de séquence

Amorce : courte séquence d'ARN ou d'ADN, complémentaire du début d'une matrice, servant de point de départ à la synthèse du brin complémentaire de cette matrice par une ADN polymérase

Amplicon : fragment d'ADN amplifié en un grand nombre (plusieurs millions) de copies par une réaction par PCR

Alignement des reads : détermination par un logiciel bio-informatique de la position des *reads* sur un génome de référence

Annotation des variants : documentation des variants par un logiciel bio-informatique à partir d'informations recueillies dans des bases de données

Base-calling : algorithme informatique affectant une base (A,C,G ou T) à un signal généré par un séquenceur automatique

CDS (*Coding DNA Sequence*) ou séquence codante de l'ADN : ensemble des parties des exons qui seront traduites en protéines. Dans un gène, elle commence au niveau du triplet initiateur de la traduction et se termine au niveau du codon stop, terminateur de la traduction.

Cluster : Groupe du même fragment d'ADN issu de la PCR avec formation de ponts

CNV (*Copy Number Variant*) : variation du nombre de copie d'une séquence supérieure ou égale à 0,1 Kb

Couverture d'une région cible : pourcentage de la région ciblée pour lequel la profondeur de séquençage dépasse une valeur minimale prédéfinie ; elle s'exprime en %. Le pourcentage moyen de couverture à une profondeur « n » est donc le nombre de bases séquencées au moins n fois par rapport au nombre total de bases ciblées. Par exemple : « 90% à 20X » signifie que 90% des bases ciblées sont au minimum couvertes par 20 lectures (*reads*) indépendantes

Couverture théorique : couverture calculée à partir de la localisation des amorces ou des sondes conçues dans le cadre du développement d'un test

Déficience intellectuelle légère (selon l'OMS) : DI avec un score de QI compris entre 50 et 69

Déficience intellectuelle modérée (selon l'OMS) : DI avec un score de QI compris entre 35 et 49

Déficience intellectuelle sévère (selon l'OMS) : DI avec un score de QI compris entre 20 et 34

Déficience intellectuelle profonde (selon l'OMS) : DI avec un score de QI inférieur à 20

DéfiScience : filière nationale de santé pour les maladies rares du développement cérébral et pour la déficience intellectuelle

Démultiplexage : traitement informatique des *reads* visant à les attribuer au bon échantillon, grâce aux séquences code-barres ou Index

Dessin ou Design d'un test : travail de conception de l'ensemble des amorces ou des sondes pour cibler les régions étudiées par ce test

Enrichissement par amplification : sélection des fragments d'ADN d'intérêt à l'aide d'une amplification sélective par PCR

Enrichissement par capture : sélection des fragments d'ADN d'intérêt par des procédés d'hybridation

Epigénétique : mécanisme modifiant l'expression des gènes sans en changer la séquence nucléotidique (ADN), de manière réversible, transmissible (lors des divisions cellulaires) et adaptative

Epissage : processus par lequel les ARN transcrits à partir de l'ADN génomique peuvent subir des étapes de coupure et ligature qui conduisent à l'élimination de certaines régions dans l'ARN messager mature ; les segments conservés s'appellent des exons et ceux qui sont éliminés s'appellent des introns

Exome : région codante de tous les gènes

Exome clinique : région codante des gènes dont la signification clinique est connue

Exons codants : exons qui seront traduits en protéine ; tous les exons d'un gène ne sont pas codants.

Fichier BAM : version binaire et compressée du fichier SAM ; il peut être accompagné d'un fichier d'index (extension .bai)

Fichier BCL : fichier binaire contenant les bases détectées, leur qualité pour chaque cycle ; il constitue le format brut pour la plateforme Illumina

Fichier DAT : fichier contenant la conversion des mesures brutes de pH en mesures numériques des voltages

Fichier FASTQ : fichier texte standard contenant les séquences associées à la valeur de qualité de leurs nucléotides

Fichier SAM : fichier texte représentant des séquences alignées issus des programmes d'alignement

Fichier VCF : fichier texte répertoriant les variations de séquence observées chez un ou plusieurs patients suite à l'étape informatique de *variant-calling*

Fichier WELLS : fichier d'entrée pour la détection des bases ; il constitue le format brut pour la plateforme Ion Proton et PGM™ Ion Torrent™

Flow Cell (FC) : support solide utilisé pour la réaction de PCR avec formation de ponts et pour la réaction de séquençage par la technologie *Illumina*

Génome : ensemble du matériel génétique d'une espèce

Génome de référence : base de données de séquences nucléiques consensus pour une espèce. Il représente l'assemblage du génome de différents donneurs, et non d'un

individu en particulier. Ainsi, il existe des différences entre le génome de référence et celui d'un individu donné, tels que les SNP.

Haplotype : L'ensemble des gènes situés sur un même chromosome et dont les allèles ségrégent ensemble lors de la méiose constituent un haplotype

Hétérozygote pour un gène : possédant deux allèles différents de ce gène

Homopolymère : région de l'ADN constituée par la répétition d'une même base nucléotidique plusieurs fois d'affilée

IGV (*Integrative Genomics Viewer*) : logiciel permettant de visualiser les séquences obtenues dans le cadre d'un test NGS après alignement

Indel : insertion et délétion de 1 à 99 pb (aussi appelée delins)

Indexage : étape visant à attribuer à chaque échantillon une séquence code-barres ou Index permettant l'identification de l'échantillon

Librairie : banque de fragments d'ADN cible que l'on veut séquencer (une librairie par patient)

Longueur de lecture : nombre de bases individuelles composant une lecture (*read*) donnée

Mutation *de novo* : mutation présente chez un individu, alors qu'aucun des deux parents ne la possède dans son patrimoine génétique

Mutation en mosaïque : une mutation est dite « en mosaïque » lorsqu'elle n'est observée que dans une partie des cellules d'un individu. Le mosaïcisme s'explique par l'apparition de la mutation au cours du développement de l'embryon : plus la mutation survient tôt dans son développement, plus la proportion de cellules mutées est importante.

Oligonucléotide : courts segments d'ARN ou d'ADN de quelques dizaines de nucléotides

Panel de gènes : ensemble des gènes étudiés lors d'une analyse par un séquençage NGS

Panel DI44 : panel ciblé sur les 44 gènes les plus fréquemment impliqués dans la DI en France ; son rendement diagnostique est estimé à environ 10 à 12% dans le cadre de DI sans orientation clinique évidente

PCR (*Polymerase Chain Reaction*) : réaction de polymérisation en chaîne ; technique d'amplification enzymatique par une ADN polymérase qui permet à partir d'un fragment d'ADN, d'obtenir un grand nombre (plusieurs millions) de copies identiques de ce même fragment.

PCR en émulsion : Chaque goutte d'émulsion est un microréacteur dans lequel l'unique fragment d'ADN fixé sur une microbille va pouvoir être amplifié par PCR

PCR avec formation de ponts (*bridge PCR*) : PCR sur phase solide ayant des cycles d'amplification avec formation de ponts

PCR multiplexe : amplification de plus d'un amplicon à la fois, par l'utilisation d'au moins trois amorces pour la réaction de PCR

Pipeline d'analyse bio-informatique : ensemble des logiciels utilisés pour analyser les données générées par le NGS ; il comprend notamment des outils pour l'alignement des séquences, la détection et l'annotation des variants

Point de branchement d'épissage : séquence située à l'intérieur d'un intron comportant une adénosine qui joue un rôle central dans le processus d'épissage

Polymorphisme (SNP, *single-nucleotide polymorphism*) : variation d'1 pb du génome, en comparaison au génome de référence, présente chez au moins 1% des individus de la population étudiée

Profondeur de lecture : nombre de lectures (*reads*) couvrant une zone déterminée, exprimée en X. Pour une base donnée, elle correspond au nombre de fois où cette base est séquencée. On peut également fixer une « profondeur de lecture minimale » à atteindre, comme un seuil de qualité pour valider le résultat sur une base donnée. Par exemple, une région à 100X signifie qu'il y a 100 lectures (*reads*) différentes couvrant chaque base de cette région.

Profondeur de lecture moyenne (PLM) : moyenne des profondeurs de lecture observées sur l'ensemble d'une région donnée (par exemple, sur l'ensemble des régions ciblées par un panel de gènes)

Read ou lecture : succession de bases lues par le séquenceur au niveau de chaque cluster, ou puits, selon la technologie utilisée ; il correspond à la séquence lue d'un des fragments d'ADN de la librairie

Rendement diagnostique : rapport du nombre de patients avec un ou plusieurs variant(s) pathogène(s) décelé(s) sur le nombre de patients étudiés

Robustesse d'un test : un test est robuste si des changements mineurs dans la procédure n'ont pas d'influence sur ses résultats

Run : un *run* correspond à un passage sur un séquenceur. Il regroupe l'ensemble des échantillons ayant été séquencés en même temps.

Sensibilité d'un test : capacité à donner un résultat positif lorsqu'une hypothèse est vérifiée ; c'est-à-dire, dans le cadre d'un test diagnostique, à détecter tous les malades (vrai positifs) et à générer le moins de faux-négatifs

Spécificité d'un test : capacité à donner un résultat négatif lorsque l'hypothèse n'est pas vérifiée ; c'est-à-dire, dans le cadre d'un test diagnostique, à ne détecter que les malades (avoir le plus de vrai négatifs et le moins de faux-positifs)

Séquençage de nouvelle génération (NGS) ou Séquençage massif en parallèle (MPS) : regroupent les technologies de séquençage haut-débit de 2^{ème} et de 3^{ème} génération

Séquençage de 2^{ème} génération : séquençage d'un ensemble de molécules nucléotidiques à l'aide de techniques de "*wash-and-scan*" (ou cycles)

Séquençage de 3^{ème} génération : procédé de séquençage de molécules uniques ne nécessitant pas de "*wash-and-scan*"

Séquence code-barres ou Index : séquence oligonucléotidique, ajoutée artificiellement à chaque fragment d'ADN d'un patient, spécifique à chaque patient au cours d'un *run* de séquençage NGS et permettant d'attribuer ultérieurement les séquences propres à chaque patient lors de l'analyse bio-informatique

SNV (Single Nucleotide Variant) : variation de 1 pb, par rapport au génome de référence, de type substitution

Système Streptavidine/Biotine : système qui repose sur l'affinité exceptionnellement haute de la streptavidine pour la biotine à l'origine d'une liaison non covalente

Trimming des reads : nettoyage des *reads* par un logiciel informatique qui enlève les adaptateurs et les portions de *read* de mauvaise qualité

Variant : variation de séquence nucléotidique par rapport au génome de référence

Variant de découverte fortuite (*incidental finding*) : variant dont l'information génétique n'a pas de lien avec l'indication ayant conduit à la prescription de l'examen

Variant allele frequency (VAF) : fréquence à laquelle se trouve le variant ; c'est le rapport entre le nombre de lectures contenant le variant et nombre de lectures totales à cette position donnée

Variant-calling : algorithme informatique visant à détecter des variations de séquence par rapport au génome de référence sur les fragments séquencés

Variant Caller : logiciel utilisé pour l'étape de *variant-calling*

Variant faux-négatif : variant non trouvé par le test alors qu'il a été mis en évidence par une méthode de référence

Variant faux-positif : variant détecté par le test alors qu'il a été infirmé par une méthode de référence

Variant vrai négatif : variant infirmé par une méthode de référence non décelé par le test

Variant vrai positif : variant confirmé par une méthode de référence retrouvé par le test

Variant frameshift : variant entraînant un décalage du cadre de lecture, conduisant soit à un codon stop prématuré et donc à la synthèse d'une protéine tronquée en général peu ou pas fonctionnelle (cas le plus fréquent), soit à la destruction de l'ARN messenger avant sa traduction et donc à l'absence de synthèse de protéine

"Wash-and-scan" : technique basée sur des ADN polymérases et des réactifs qui doivent être enlevés à chaque cycle après l'incorporation des bases à lire

Introduction

I. La déficience intellectuelle (DI)

La DI est un problème majeur de santé publique, puisqu'elle touche 1 à 2% environ de la population mondiale (4, 5). Elle est responsable d'un handicap parfois majeur nécessitant une prise en charge spécialisée tout au long de la vie de la personne. En dehors des aspects moraux, l'impact économique de la DI pour la société est considérable et justifie aussi des mesures pour la prévention et pour un diagnostic précoce. Ce dernier est rendu difficile du fait de la très grande hétérogénéité de l'étiologie et il est très souvent obtenu au prix d'une longue errance, ce qui retarde la prise en charge adaptée des patients. Les causes génétiques constituent dans les pays développés la majeure partie des DI. Leur diagnostic a considérablement progressé ces dix dernières années grâce aux biotechnologies faisant appel aux puces à ADN, celles pour l'analyse chromosomique comparative sur puce à ADN (ACPA appelée aussi *CGH array*) et celles plus récentes pour le séquençage de nouvelle génération* (NGS, *Next Generation Sequencing*). Si les premières sont accessibles en routine depuis plusieurs années dans notre CHU, en revanche le diagnostic de la DI par NGS n'est toujours pas opérationnel.

Or, près de 300 nouveaux cas de DI sont diagnostiqués cliniquement chaque année dans les 8 départements rattachés à l'ex région Midi-Pyrénées dont les trois-quarts demeurent sans diagnostic étiologique après un bilan initial incluant l'analyse chromosomique.

Le laboratoire de génétique médicale s'est fixé comme objectif majeur le développement d'un test moléculaire performant reposant sur le séquençage haut-débit (NGS) d'un panel de gènes* impliqués dans la DI, que nous avons nommé test DI117. C'est dans ce contexte que s'inscrit mon travail de thèse, visant à développer ce test DI117 et à évaluer sa performance diagnostique. Dans cette partie introductive, je commencerai par faire une présentation synthétique de la DI et de son diagnostic puis je présenterai la méthode de séquençage de nouvelle génération que j'ai évaluée dans ce travail.

1. Prévalence et définitions de la DI

La prévalence de la DI est difficile à déterminer avec précision car elle varie selon la méthodologie de l'étude et la définition utilisée. En France, elle est estimée pour la Déficience Intellectuelle Légère (DIL ; QI=50-69), entre 10 et 20 pour 1000, et, pour la Déficience Intellectuelle Sévère (DIS ; QI < 50) entre 3 et 4 pour 1000. Ces chiffres sont semblables à ceux retrouvés dans les autres pays développés. Toutes les études révèlent une proportion plus importante de garçons avec un sex-ratio entre 1,2 et 1,9 (6-8). Cela s'explique en grande partie par le nombre élevé d'étiologies génétiques liées au chromosome X.

La terminologie utilisée pour désigner les personnes présentant un déficit des capacités intellectuelles a beaucoup évolué au fil du temps. Depuis une dizaine d'année, le terme « retard mental » (RM), devenu désuet, a été remplacé par « déficience intellectuelle » (DI). A noter que, l'OMS, dans la CIM-11 (Classification internationale des maladies), en cours de rédaction, envisage la dénomination « troubles du développement intellectuel » (« *disorders of intellectual development* »), considérée comme une sous-classe de « troubles neurodéveloppementaux » (« *neurodevelopmental disorders* »).

Plusieurs définitions de la DI ont été proposées par les autorités internationales, telles que l'Organisation Mondiale de la Santé (OMS), l'*American Association on Intellectual and Developmental Disabilities* (AAIDD), et l'*American Psychiatric Association* (APA).

La plus utilisée dans la littérature scientifique est celle de l'APA que l'on trouve dans le DSM-5 (*Diagnostic and Statistical Manual of Mental Disorders*). La DI y est définie comme une diminution significative des facultés intellectuelles entraînant un déficit du fonctionnement adaptatif dans 3 domaines (conceptuel, social et pratique), débutant pendant la période du développement. La DI se distingue ainsi du déclin neurocognitif des affections neurodégénératives débutant en général à l'âge adulte.

Le fonctionnement intellectuel global est évalué par le score de quotient intellectuel (QI), obtenu à l'aide de tests psychométriques standardisés et étalonnés sur une population de référence de même tranche d'âge chronologique, de même culture et dépourvue de handicap sensoriel. Le score de QI ambitionne d'être un reflet du niveau d'intelligence selon une échelle gaussienne basée sur des tests normés. La moyenne de ces tests est un score de 100 et la déviation standard (DS) est de 15.

Les tests de QI portent sur le raisonnement, la résolution de problèmes, la pensée abstraite, la faculté de jugement, l'apprentissage académique, et l'apprentissage par l'expérience.

Les plus utilisés sont le WISC-V (*Wechsler Intelligence Scale for Children*), pour les enfants entre 6 et 16 ans, et le WAIS-IV (*Wechsler Adult Intelligence Scale*), utilisable à partir de 16 ans.

La déficience intellectuelle est définie par un score de QI inférieur à 70.

Les compétences adaptatives sont mesurées par d'autres échelles, la plus connue étant celle de Vineland (*Vineland Adaptive Behavior Scales, VABS*).

Différents critères permettent d'évaluer le degré de sévérité de la DI.

L'OMS (CIM-10) utilise le score de QI reflétant le niveau intellectuel. Ainsi, la DI est qualifiée de légère pour un score de QI entre 69 et 50, de modérée entre 49 et 35, de sévère entre 34 et 20, et de profonde en dessous de 20. Cependant, cette classification de la sévérité reposant uniquement sur le fonctionnement intellectuel global est de moins en moins utilisée.

Ainsi, l'APA et l'AAIDD prennent en compte le retentissement sur le fonctionnement adaptatif. Depuis le DSM-5, l'APA définit ainsi la gravité de la DI en fonction de l'importance du déficit des capacités adaptatives dans les 3 domaines conceptuel, social et pratique. L'AAIDD préfère une classification sur la base de l'intensité de soutien requis.

De ces 3 approches émanent 3 classifications, définissant chacune 4 niveaux de gravité. En général, il existe une correspondance entre les catégories de ces classifications.

2. Hétérogénéité des présentations cliniques de la Déficience Intellectuelle

La DI peut survenir de façon isolée, ou être associée à divers troubles physiques ou mentaux. La présentation clinique des patients atteints de DI est donc très variée. L'association fréquente de certains symptômes permet de caractériser des syndromes particuliers et de regrouper les individus atteints de DI dans des groupes phénotypiquement homogènes.

On différencie ainsi les DI isolées dites non syndromiques (DI-NS) des DI syndromiques (DI-S); les DI associées uniquement à des troubles neuropsychiatriques sont habituellement considérées comme des DI-NS.

De nombreuses comorbidités peuvent être associées à la DI et compliquer sa prise en charge.

La plus fréquente d'entre elles est la présence de malformations congénitales. Elle concerne environ 28% des cas de DI légères à modérées, et plus de la moitié des cas de DI sévères à profondes (9). La trisomie 21 qui représente 8% des cas de DI est par exemple associée à une malformation cardiaque dans un peu plus d'un cas sur deux (10). En revanche, chez les patients ayant une DI d'origine non chromosomique avec malformation congénitale, ce sont surtout les malformations cérébrales, musculo-squelettiques et génito-urinaires qui prédominent (9).

Les troubles du spectre autistique (TSA) constituent un autre facteur important de comorbidité. Ainsi, 28% des individus atteints de DI auraient des TSA associés (11). Le syndrome de l'X fragile est un exemple connu de cette association. Réciproquement, 67% des patients TSA seraient atteints de DI (12). L'autisme et la DI partagent par conséquent

des mécanismes physiopathologiques communs. Les étiologies génétiques identifiées dans les TSA sont en général également impliquées dans la DI, parfois sans TSA.

Parmi les autres comorbidités notables, il convient de citer l'épilepsie (22% des patients DI), les troubles moteurs d'origine centrale (20% des patients DI), les troubles neuropsychiatriques (hyperactivité, déficit de l'attention, agressivité, dépression) ainsi que des déficits sensoriels (auditifs, visuels) (13).

Le taux de mortalité des patients présentant une DI légère isolée est comparable à celui de la population générale. En revanche il est plus élevé lorsque l'intensité de la déficience augmente ou lorsqu'il existe des comorbidités associées (14, 15).

3. Hétérogénéité des étiologies des DI

Une DI est toujours la conséquence d'un trouble pré ou post natal du développement cérébral. Les causes sont diverses et multiples.

Elles sont schématiquement réparties en 2 groupes : les causes environnementales et les causes génétiques.

Dans les pays développés, c'est 15 à 20% des cas de DI qui peuvent être attribués à une origine environnementale (5). Les formes génétiques sont donc très probablement les formes les plus fréquentes sachant que dans environ 35 à 40% des cas, aucune étiologie n'est identifiée (5). De façon intéressante, ces proportions varient selon la sévérité de la DI : plus la déficience est sévère, plus la part génétique est importante et identifiable.

Les étiologies environnementales (acquises) peuvent être subdivisées en 4 grands groupes :

- les causes toxiques, comprenant l'exposition *in utero* à certaines substances neurotoxiques comme l'alcool, responsable du syndrome d'alcoolisme fœtal, certaines drogues (comme la cocaïne) ou médicaments (notamment certains antiépileptiques) (5), ou autres produits (plomb, mercure, arsenic, etc.) (5, 16), et l'intoxication postnatale au plomb ou au mercure (17, 18)
- les causes infectieuses, telles que les encéphalites anté ou postnatales (CMV, toxoplasme, virus de l'herpès, de la rubéole, etc.) (5, 19, 20)
- les causes traumatiques, avec notamment les hémorragies intracérébrales et l'ischémie cérébrale (secondaire à un accident vasculaire cérébral ou après une noyade) (5)
- les complications liées à la grande prématurité représentent à elles seules près de 5% des cas de DI. Elles résultent le plus souvent de lésions cérébrales ou de maladie pulmonaire chronique. Leur prévalence est corrélée à l'importance de la prématurité (10).

a. Etiologies génétiques

On estime qu'environ 1/3 des 22 à 23 000 gènes humains sont exprimés au niveau du cerveau, participant à son développement et à son fonctionnement (5).

Par conséquent, les causes génétiques sont extrêmement nombreuses. En dresser une liste exhaustive est un défi (21), d'autant plus que leur nombre ne cesse d'augmenter avec la découverte de nouveaux gènes en cause.

Ainsi, la recherche en ligne avec le mot clé « déficience intellectuelle » conduit sur Orphanet (www.orpha.net) à 358 syndromes distincts, tandis que sur OMIM (www.omim.org) les mots clés « intellectual disability or mental retardation » débouchent sur 2905 entrées. Le site internet Phenomizer (<http://www.compbio.charite.de/phenomizer/>) indique, quant à lui, 1242 syndromes et la

base de données London medical databases (LMD) (22) 1445 syndromes avec le mot clé «intellectual disability» (HP :0001249) (23, 24)

Tous les modes de transmission génétique ont été rapportés qu'ils soient mendéliens (autosomique ou lié au chromosome X, dominant ou récessif) ou non mendéliens (c'est-à-dire sans altération de la séquence du génome nucléaire, tel que le phénomène d'empreinte parentale).

On peut diviser les causes génétiques en 2 groupes : chromosomique et génique.

b. Anomalies chromosomiques

Les anomalies chromosomiques représentent l'étiologie la plus fréquente de DI (environ 15% des DI) (5).

Il s'agit soit de modifications du nombre de chromosomes (aneuploidie), soit d'anomalies de structure. Les anomalies de structure résultent de cassures chromosomiques suivies par un ou plusieurs recollements anormaux. Elles peuvent être équilibrées, sans perte ou gain de matériel chromosomique, ou déséquilibrées, avec perte (délétion) ou gain (duplication). Parmi les anomalies de structure déséquilibrées, on distingue les CNV* (*Copy Number Variations*), délétions ou duplications de grande taille, supérieures à 1 Kb (25), et les *indels**, insertions et/ou délétions de petite taille, inférieures à 1 Kb (26). A noter que les remaniements équilibrés n'ont en général pas de conséquence phénotypique, sauf si la cassure chromosomique interrompt un gène ou altère son expression.

Les anomalies chromosomiques sont diagnostiquées par des techniques de cytogénétique : le caryotype, l'hybridation in situ (FISH pour *Fluorescence in situ hybridation*) et l'ACPA (analyse comparative sur puce à ADN ou *CGH array*).

Parmi elles, on retrouve la trisomie 21, qui est à l'origine de 7 à 11% des cas (13, 27), ce qui en fait la première cause de DI. Le rôle des anomalies de structure est souvent sous-estimé dans la DI, mais les progrès techniques apportés par l'ACPA dans les années 2000

puis par le séquençage à haut-débit (cf. partie Introduction, paragraphe II.1.b, page 53) permettent de mettre en évidence des anomalies de structure méconnues jusqu'alors (28, 29). Les CNV représenteraient 20% des cas de DI en excluant les causes environnementales (5, 30).

c. Anomalies géniques

Environ 20 à 30% des cas de DI ont une étiologie génique identifiable aujourd'hui (5) sous la forme de petites variations de l'ADN (mutations ponctuelles) de type substitution nucléotidique (*Single Nucleotide Variation, SNV*) * ou de type petite insertion ou petite délétion nucléotidique (*Indels*).

Parmi les causes monogéniques il convient de citer, du fait de sa fréquence, le syndrome de l'X fragile, première cause de DI héréditaire, dominant lié à l'X. Bien qu'il ne représente que 2% des cas de DI, le syndrome de l'X fragile serait la deuxième cause la plus fréquente de DI (5).

Le diagnostic des anomalies géniques repose sur des techniques de génétique moléculaire, essentiellement le séquençage de l'ADN.

4. Classifications des gènes impliqués dans la DI

Les nombreux gènes impliqués dans la DI peuvent être classifiés de différentes manières. Plus de 700 gènes sont impliqués dans les cas de DI (29). Compte-tenu de ce nombre considérable, une classification est en effet indispensable pour orienter les examens génétiques lors de la démarche diagnostique. Ainsi, en fonction de la présentation clinique et de l'histoire familiale, certains gènes seront préférentiellement analysés.

Tout d'abord, en se basant sur le phénotype, on peut opposer les gènes responsables de DI-NS à ceux responsables des différentes DI-S. Cette classification est très utile en

pratique clinique, car elle permet d'orienter les examens diagnostiques en fonction de la présentation clinique. Cependant, de nombreux gènes sont à l'origine à la fois de formes de DI-NS et de DI-S. C'est notamment le cas des gènes *ARID1B* (syndrome de Coffin-Siris), *ATRX* (α -thalassémie et retard mental lié à l'X), et *RPS6KA3* (syndrome de Coffin-Lowry) (31-33). Cette variabilité phénotypique de certains gènes peut ainsi compliquer cette classification pourtant utile en pratique.

Une autre façon de classifier les gènes de DI est de se référer au mode de transmission de leurs mutations : autosomique dominant (AD), autosomique récessif (AR), lié à l'X, ou non mendélien. Cette classification aide également à cibler les examens diagnostiques, notamment dans le cadre de formes familiales, et à interpréter leurs résultats. Par exemple, si une DI sur plusieurs générations successives est mise en évidence, les examens seront ciblés en premier lieu sur les gènes pouvant avoir une transmission autosomique dominante.

Enfin, les gènes de DI peuvent aussi être classés selon le mécanisme physiopathologique dans lequel ils sont impliqués. Cependant, dans de très nombreux cas, le lien entre la fonction de la protéine codée par le gène et la survenue de la DI est méconnu ou encore hypothétique. Il est toutefois possible de dégager quelques principaux mécanismes biologiques dans lesquels les gènes de DI semblent préférentiellement jouer un rôle. D'une manière générale, les gènes ayant un rôle dans le contrôle de l'expression génique sont plutôt responsables de DI-S. Ceux codant pour des protéines de la synapse dont les mutations affectent la transmission synaptique ou la plasticité synaptique sont plutôt retrouvés dans les DI-NS et les TSA (34). Cette classification est de plus en plus utilisée dans les articles scientifiques car elle liste les gènes selon l'impact phénotypique engendré par les mutations. Ainsi, les mutations de gènes impliqués dans une même voie de signalisation peuvent être à l'origine de tableaux cliniques divers mais dans lesquels certains symptômes sont communs et ont valeur d'orientation pour le clinicien (par exemple, les RASopathies).

Il me paraît important de donner quelques exemples de cette classification ci-dessous.

a. Gènes codant pour des enzymes impliquées dans les voies métaboliques

Les DI liées à des dysfonctionnements de voies métaboliques représentent 1 à 5% des DI, et ont le plus souvent une transmission autosomique récessive (27, 35). La DI est secondaire soit à une accumulation d'un substrat toxique, soit à un déficit d'un substrat essentiel au développement cérébral, soit encore à un déficit énergétique lors d'une étape critique du développement cérébral.

La plus connue d'entre elles est la phénylcétonurie. Elle est liée à des mutations bi-alléliques inactivatrices du gène de la PAH (phénylalanine-4-hydroxylase), une enzyme hépatique, conduisant à une accumulation de son substrat, la phénylalanine, toxique pour le cerveau.

b. Gènes de la neurogenèse

Les gènes identifiés dans la DI comme ayant un impact sur la neurogenèse contrôlent la division cellulaire des neuroblastes. Leur déficit peut entraîner une réduction du nombre de neurones qui conditionnent le volume cérébral, et conduire alors à une DI associée à une microcéphalie (petite tête).

C'est le cas par exemple du gène *MCPH1* impliqué dans le syndrome de microcéphalie primaire autosomique récessive (MCPH). *MCPH1* code pour la microcéphaline, une protéine impliquée dans la régulation du cycle cellulaire des neuroblastes. Une inactivation du gène *MCPH1* entraîne une entrée prématurée en phase de mitose de ces cellules ce qui perturbe la neurogenèse, conduisant ainsi à une microcéphalie associée à une DI (36).

c. Gènes de la migration neuronale

Durant le développement, les cellules nerveuses migrent des zones ventriculaires et sous-ventriculaires jusqu'à leur localisation définitive. De nombreuses protéines jouent un rôle dans ce processus migratoire. Une dysfonction de l'une d'entre elles conduit en général à une DI associée à des anomalies de giration cérébrale.

Par exemple, le gène *L1CAM* code pour la molécule d'adhésion cellulaire transmembranaire L1, exprimée principalement dans le système nerveux en développement. Elle joue un rôle crucial dans la migration radiale des neurones corticaux (37). Sa dysfonction entraîne le syndrome L1 caractérisé entre-autres par une hydrocéphalie et une DI.

d. Gènes de la fonction synaptique

Les gènes de la fonction synaptique sont souvent impliqués dans l'autisme, mais ils peuvent aussi être impliqués dans la DI.

La grande majorité des synapses du système nerveux sont chimiques, c'est-à-dire qu'elles répondent à un stimulus en libérant un neurotransmetteur dans la fente synaptique.

A l'état basal, des vésicules contenant les neurotransmetteurs sont présentes dans la terminaison présynaptique. Après stimulation, ces vésicules sont transloquées et arrimées à la membrane présynaptique. Ce trafic des vésicules est, entre autre, contrôlé par des GTPases, les protéines Rab. Puis, en réponse à un influx intracellulaire de calcium, ces vésicules subissent une exocytose entraînant la libération du neurotransmetteur dans la fente synaptique. De nombreux gènes responsables de DI sont impliqués dans ces différents mécanismes.

Le gène *STXBP1* (aussi appelé *Munc18-1*) en est un exemple. Il code pour une protéine qui interagit avec la syntaxine et permet l'arrimage des vésicules à la membrane présynaptique (38). Des mutations de ce gène sont à l'origine d'encéphalopathies épileptiques avec DI.

La densité post-synaptique est la région intracellulaire immédiatement sous-membranaire, contenant la partie intracytoplasmique des récepteurs des neurotransmetteurs, des canaux ioniques, ainsi que le cytosquelette sous-membranaire. Elle joue un rôle majeur dans la transduction post-synaptique du signal. La mutation des gènes de la densité post-synaptique peut provoquer une DI.

Par exemple, le gène *GRIA3* code pour une sous-unité d'un récepteur ionotrope au glutamate (récepteur AMPA). Une anomalie de ce gène provoque une disruption de la transduction du signal pour les neurones excitateurs glutaminergiques, résultant en une DI (39).

La régulation de la densité protéique post-synaptique est très importante pour la plasticité synaptique. Deux mécanismes permettent cette régulation : d'une part, l'augmentation de l'expression de certaines protéines sous l'action de facteurs de transcription, d'autre part la dégradation de ces protéines par la protéolyse. Une dysfonction des gènes ayant un rôle dans cette régulation peut avoir pour conséquence une DI.

Un exemple est le gène *UBE3A*, qui code pour une ubiquitine ligase de la protéolyse cellulaire. Cette protéine est nécessaire à la dégradation (et donc au recyclage) d'autres protéines de la cellule. Une dysfonction de ce processus est responsable du syndrome d'Angelman (40). Le syndrome d'Angelman se caractérise, entre autres, par une DI sévère, une déficience motrice sévère et une épilepsie.

L'épine dendritique a la capacité de changer de taille et de forme, grâce à la polymérisation/dépolymérisation des filaments d'actine et de microtubules, phénomène contrôlé par la voie de signalisation Rho-GTPase. Une dysfonction des gènes régulant ces mécanismes peut être à l'origine d'une DI.

Un exemple de gène activateur de la voie de signalisation Rho-GTPase est le gène *OPHN1*. Ce gène contrôle ainsi la structure, la fonction, et la plasticité des synapses excitatrices glutaminergiques. Sa dysfonction implique un développement altéré de ces synapses résultant en une DI (41).

e. Gènes appartenant à des cascades de signalisation intracellulaire

L'activation des récepteurs des neurotransmetteurs entraîne des cascades de signalisation intracellulaire postsynaptique, en particulier de la voie Ras-MAPK-ERK et PI3K-AKT-mTOR. Un certain nombre de gènes connus de la DI ont un rôle dans ces voies de signalisation.

Le gène *PTPN11*, par exemple, code pour une phosphatase de la voie de signalisation Ras-MAPK-ERK, la protéine SHP2. Le syndrome de Noonan, responsable d'une DI de sévérité variable, est causé par des mutations activatrices de SHP2 (42).

f. Gènes ayant un rôle dans la régulation épigénétique de la transcription

La régulation de la transcription génique est essentielle au bon fonctionnement cérébral, comme le démontre l'implication dans la DI de nombreux gènes ayant un rôle dans l'expression génique.

Le gène *MECP2*, par exemple, réprime la transcription génique en se liant à l'ADN méthylé. Des mutations de ce gène sont responsables du syndrome de Rett, une affection neurologique sévère associée à des troubles du spectre autistique (43).

5. Bilan génétique en 2019 pour le diagnostic étiologique de la DI

Le diagnostic d'une cause génétique revêt une importance cruciale pour la prise en charge des patients atteints de DI et de leurs apparentés. Il est notamment indispensable pour évaluer le risque de récurrence et délivrer un conseil génétique précis à la famille. Un diagnostic prénatal (DPN) peut être discuté dans la plupart des cas. En plus de l'intérêt pronostique, l'identification précoce d'une cause génétique conduit à la mise en place d'un suivi médical approprié, permettant un dépistage systématique des complications

associées connues, voire même parfois l'instauration d'options thérapeutiques ciblées. Enfin, un diagnostic de maladie rare permet une meilleure reconnaissance sociale par les différents organismes (scolaires, maison départementale des personnes handicapées, etc.) en termes d'aides diverses.

Cependant, l'implication d'un très grand nombre de gènes dans la DI et leur faible fréquence relative rendent le diagnostic étiologique génétique difficile. Ceci d'autant que la majorité d'entre eux sont responsables de DI-NS, non différenciables cliniquement, et que de nouveaux gènes de DI restent encore à découvrir.

Ces dernières années, l'émergence des techniques de séquençage haut débit dites de nouvelle génération (NGS), en permettant d'analyser simultanément un nombre considérable de gènes pour un coût parfois inférieur à celui du séquençage d'un seul gène en technique classique (séquençage par la méthode de Sanger), a révolutionné le diagnostic moléculaire des maladies rares à forte hétérogénéité génétique.

Dans ce contexte, un algorithme diagnostique a été proposé par l'ANPGM (Association Nationale des Praticiens de Génétique Moléculaire) pour guider les investigations du bilan génétique dans le cadre d'un diagnostic clinique de déficience intellectuelle (**Figure 1**).

En préambule de ce bilan génétique, il convient toujours d'éliminer une éventuelle étiologie environnementale et de bien définir le phénotype clinique. Pour cela, un interrogatoire étayé (histoire médicale et développementale détaillée, arbre généalogique mentionnant les antécédents familiaux sur 3 générations, etc.) et un examen clinique complet minutieux, notamment neurologique et morphologique, sont essentiels. L'examen dysmorphologique détaillé est incontournable et crucial pour l'interprétation des résultats génétiques. Des examens complémentaires biochimiques et d'imagerie peuvent également être prescrits dans ce but.

A l'issue de cette première étape, il existe, d'un point de vue génétique, différentes approches.

A ce jour, lorsqu'un syndrome génétique connu est fortement suspecté ou un groupe phénotypique identifié (par exemple, DI avec microcéphalie <-3 DS, DI avec excès de croissance, etc.), un examen génétique ciblé, soit sur un locus chromosomique ou sur un gène particulier soit sur un petit panel de gènes, est proposé en première intention. Au

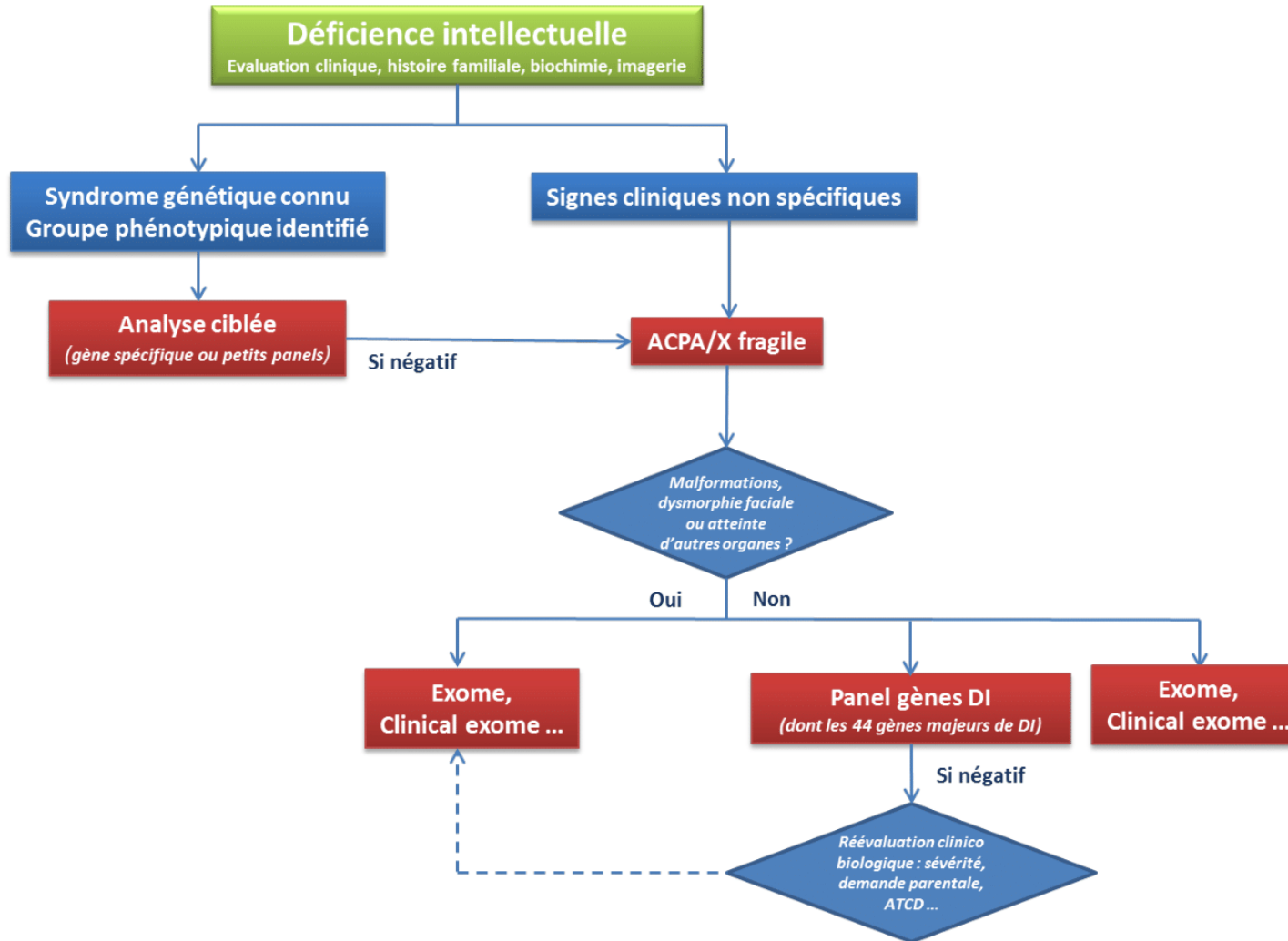
niveau chromosomique, la recherche ciblée de CNV spécifiques peut faire appel aux techniques de FISH ou de PCR* quantitative (comme la qPCR, *quantitative polymerase chain reaction*). Les petits panels de gènes ciblés sont quant à eux élaborés par des laboratoires référents ayant une expertise dans le domaine concerné. Leur taille peut varier de quelques gènes à plusieurs dizaines selon l'hétérogénéité génétique de la pathologie étudiée.

Une ACPA (analyse comparative sur puce à ADN) et une étude du gène de l'X Fragile, sont bien souvent réalisés en première intention et de façon quasi systématique. Ces deux examens sont notamment incontournables pour les cas de DI sans signe clinique d'orientation. L'ACPA, qui permet la recherche de remaniements génomiques déséquilibrés, a en effet un rendement diagnostique* non négligeable. Elle permet de réaliser 5 à 20% de diagnostics étiologiques chez des patients atteints de DI (44-48). Quant au syndrome de l'X Fragile, il représente 2 à 3% des DI chez les garçons et 1 à 2% des DI chez les filles (49).

Dans un second temps, en cas de DI associée à un contexte malformatif avec ACPA normale, un exome* ou un exome clinique* peut être réalisé. En cas de DI isolée (avec ou sans épilepsie ou autisme) avec ACPA normale, il peut être actuellement proposé soit le panel DI44*, soit un panel plus large, soit l'exome ou l'exome clinique. Le panel DI44 est ciblé sur les 44 gènes les plus fréquemment impliqués dans la DI en France ; son rendement diagnostique est estimé à environ 10 à 12% dans le cadre de DI sans orientation clinique évidente. C'est un panel de première intention développé pour permettre d'améliorer à court terme le taux diagnostique de la DI, tout en tenant compte des contraintes actuelles des laboratoires de génétique moléculaire français, telles que l'absence de financement des réactifs de NGS, l'accès restreint à un séquenceur de moyen débit, et l'absence de structure bio-informatique adaptée dans les hôpitaux. Certains laboratoires français ont pu mettre en place de plus grand panels de gènes, avec pour conséquence une sensibilité* plus élevée, par exemple un panel de 450 gènes permettrait le diagnostic d'au moins 25 % des DI isolées. D'autres ont réussi à développer l'exome ou l'exome clinique dont la sensibilité diagnostique attendue est d'au moins 35%. Ainsi, le choix de la stratégie diagnostique à adopter est complexe, et très dépendante des possibilités locales et de l'évolution rapide des connaissances.

L'évaluation coût/efficacité de ces différentes approches diagnostiques notamment celles reposant sur du séquençage non ciblé à très haut débit (exome clinique, exome complet voire génome complet) est actuellement en cours. En effet, l'errance diagnostique fréquente dans la DI comporte un coût sociétal lié au possible cumul des tests génétiques à réaliser, et retarde la prise en charge optimale des patients. La puissance considérable du NGS d'un point de vue diagnostique et son coût technique, qui le rendent de plus en plus accessible en première intention, tendent à remettre rapidement en question les algorithmes décisionnels. Cependant, l'importance des données générées et la difficulté d'interprétation qui en découle sont également à prendre en considération pour le choix d'un test diagnostique de routine. De plus, d'un point de vue éthique, le risque élevé de découverte de variants* non recherchés (*incidental findings* ou variants de découverte fortuite*) et de variants dont l'implication pathogène reste incertaine n'est pas à négliger. Il est donc crucial que le médecin s'interroge sur l'algorithme décisionnel à mettre en œuvre pour le diagnostic de la DI en prenant en compte tous les aspects sous-jacents aux possibilités considérables offertes par le NGS.

Figure 1 : Algorithme décisionnel proposé par l'ANPGM (Association Nationale des Praticiens de Génétique Moléculaire) pour guider les investigations du bilan génétique dans le cadre d'un diagnostic clinique de déficience intellectuelle



II. Le séquençage de nouvelle génération (NGS)

Les techniques de séquençage de l'ADN ont considérablement évolué au cours de l'histoire récente. L'arrivée des méthodes de séquençage de nouvelle génération est en passe de révolutionner le diagnostic des maladies génétiques, notamment celles dont l'étiologie est très hétérogène, comme c'est le cas de la DI.

1. Evolution et principes du séquençage de l'ADN

a. Le séquençage de l'ADN de 1^{ère} génération

Le séquençage de l'ADN dit de 1^{ère} génération a vu le jour au milieu des années 1970 grâce à la mise au point de méthodes reposant sur deux techniques distinctes : d'une part celle de Maxam et Gilbert basée sur le clivage chimique de l'ADN radiomarqué et d'autre part celle développée principalement par Frederick Sanger fondée sur la synthèse d'ADN radiomarqué permise par l'utilisation d'une ADN polymérase. La première fut la plus utilisée jusqu'à ce que Sanger mette au point en 1977 la technique de terminaison de chaîne (*chain-termination technique*) grâce à l'incorporation en cours de synthèse d'analogues de nucléotide appelés didésoxyribonucléotides (ddNTP). Cette méthode dite des didésoxy, qui a valu à F. Sanger le prix Nobel de Chimie (1980), a rapidement supplanté la méthode de Maxam et Gilbert pour devenir l'une des techniques les plus universellement utilisées pour le séquençage de l'ADN. Elle a notamment permis le décryptage de génomes* complets depuis celui du bactériophage en 1977 jusqu'à celui de l'homme en 2003.

Brièvement, la méthode des didésoxy (ddNTP), appelée communément aujourd'hui méthode de Sanger (**Figure 3**), repose sur la synthèse par l'ADN polymérase I d'un brin d'ADN complémentaire radiomarqué ou fluorescent. Un mélange de désoxyribonucléotides (dNTP) et de ddNTP radiomarqués ou fluorescents est utilisé de façon à ce que l'incorporation aléatoire des ddNTP entraîne la terminaison de la

synthèse du brin complémentaire (*chain-termination*). Ainsi, les 4 réactions avec les 4 ddNTP (ddATP, ddCTP, ddGTP ou ddTTP) permettent d'obtenir statistiquement autant de fragments interrompus de tailles différentes que de nucléotides de la séquence analysée. A l'origine tous les fragments obtenus devaient être séparés par électrophorèse dans un gel de polyacrylamide ou dans un gel d'agarose. Les produits radiomarqués des 4 réactions étaient déposés dans le gel sur 4 pistes parallèles puis après migration le gel était exposé à un film radiosensible. La lecture des 4 pistes permettait d'identifier par leur taille tous les fragments se terminant par T, G, C, ou A ce qui en partant des plus petits jusqu'aux plus grands permettait de décrypter des fragments de séquence ADN jusqu'à 1000 nucléotides maximum (1 Kb). C'est ainsi que fut obtenu en 1977 la séquence complète du génome du phage phi 174 (50).

Durant la décennie suivante plusieurs innovations biotechnologiques ont contribué au succès de la méthode Sanger. Ce fut notamment la mise au point de ddNTP marqués par des fluorochromes ayant des longueurs d'onde d'émission différentes permettant ainsi de réunir les 4 réactions (A,C,G,T) en un seul produit réactionnel (**Figure 3**). L'analyse électrophorétique des fragments fluorescents pouvant être réalisée au cours de l'électrophorèse grâce à un détecteur (cellule CCD) (**Figure 2**), cette innovation a rapidement conduit au développement des premiers automates de séquençage (51). Commercialisé au début des années 1990, le séquenceur ABI PRISM 310 utilisant le principe de l'électrophorèse capillaire s'est rapidement imposé dans les laboratoires du monde entier. Son fabricant *Applied Biosystems* a par la suite mis sur le marché des séquenceurs multicapillaires de plus en plus performants capables d'analyser jusqu'à 384 échantillons en un seul *run**. Grâce à ces procédés, la méthode de Sanger a permis le décryptage de génomes de plus en plus complexes, jusqu'à celui de l'homme en 2003.

Figure 3 : Principe du séquençage par la méthode de Sanger

(<https://www.biomnigene.fr/nos-solutions/sanger.html>)

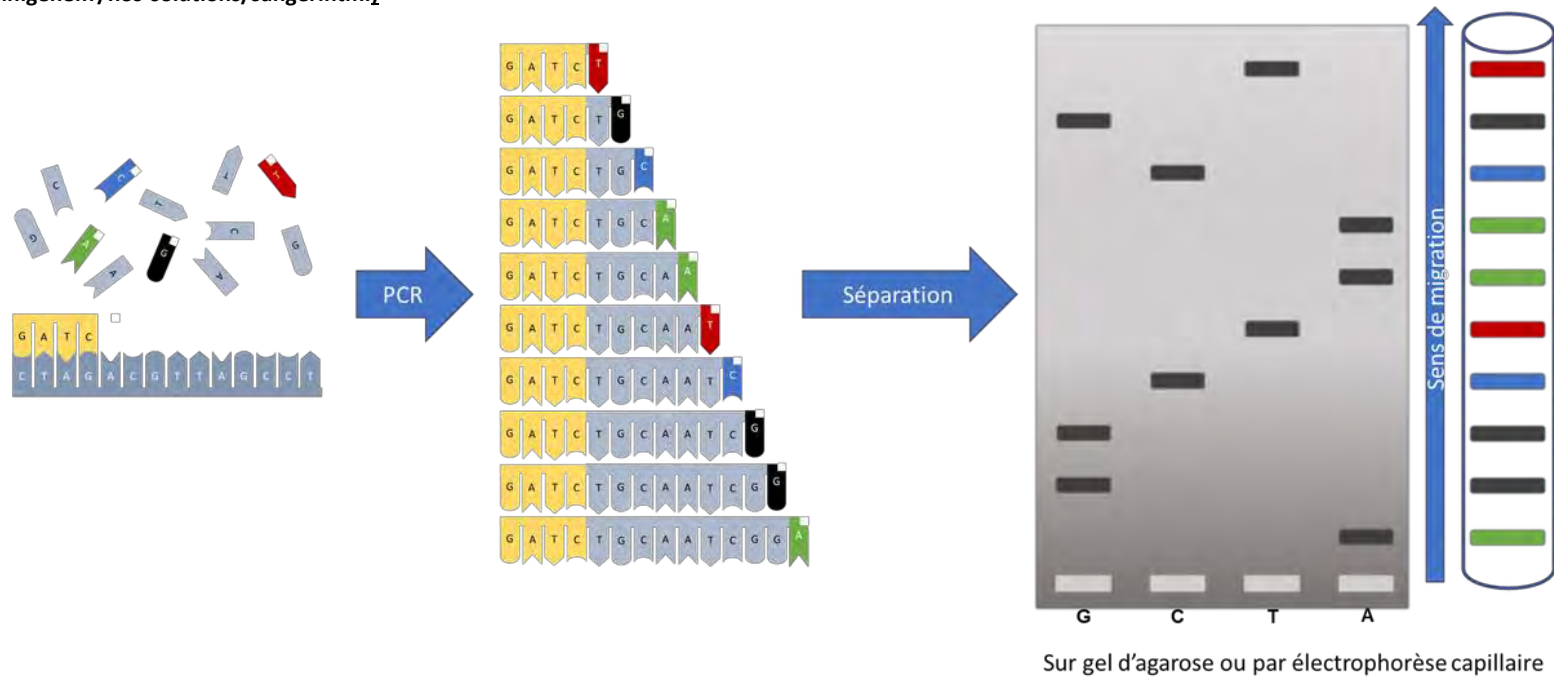
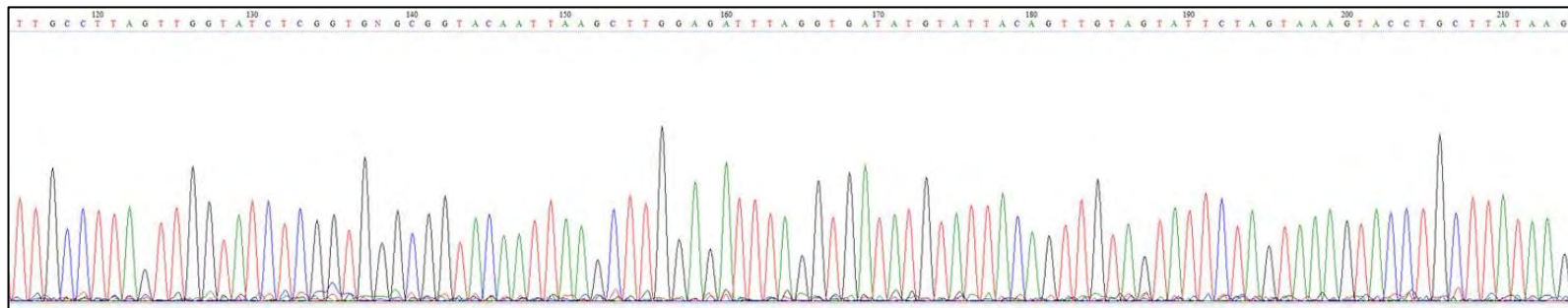


Figure 2 : Profil de migration après électrophorèse capillaire

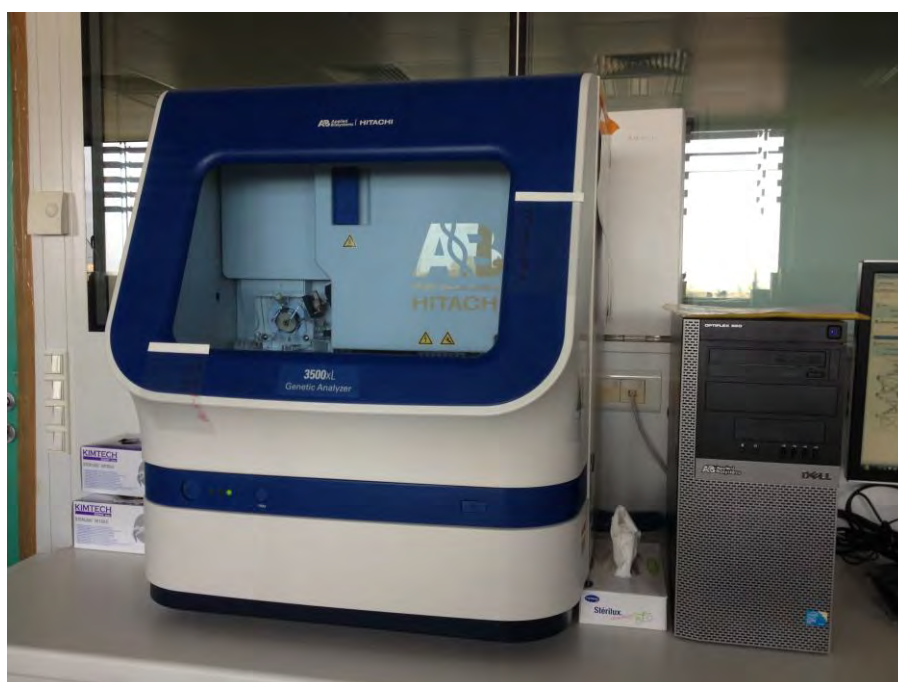
(<https://www.biomnigene.fr/nos-solutions/sanger.html>)



La recherche en génomique n'a pas été la seule à bénéficier de ces avancées biotechnologiques et très tôt le séquençage de 1^{ère} génération a eu des applications médicales incontournables pour le diagnostic des maladies constitutionnelles. Ainsi, le laboratoire de génétique médicale du CHU de Toulouse, pour ses activités de diagnostic moléculaire, utilise depuis près de 20 ans des séquenceurs par électrophorèse capillaire. Il est actuellement équipé d'un ABI 3500XL doté de 24 capillaires et d'un passeur de plaques de 96 puits permettant le séquençage simultané d'environ 600 pb/puits en environ deux heures (soit 57,6 Kb pour la plaque entière en environ 8 heures) (**Figure 4**).

Malgré ses limites en termes de puissance de séquençage (on peut parler de « bas débit ») la méthode de Sanger du fait de sa robustesse* reste encore très utilisée dans les laboratoires. Elle garde encore tout son intérêt pour l'analyse ciblée de petits gènes (pour des enquêtes familiales ou pour la recherche de mutations récurrentes) et pour confirmer certains résultats obtenus par un séquençage de 2^{ème} génération*.

Figure 4 : Photographie du séquenceur ABI 3500XL présent au laboratoire



b. Le séquençage de nouvelle génération (NGS) : une révolution biotechnologique

Les nouvelles méthodes de séquençage sont baptisées « Séquençage massivement parallèle » (MPS), ou plus communément « Séquençage de Nouvelle Génération » (NGS, *Next Generation Sequencing*). Ces méthodes ont pu se développer au début des années 2000, d'une part grâce aux progrès cognitifs accomplis en génomique (décryptage de génomes complets permettant de disposer de séquences de référence), et d'autre part grâce aux progrès accomplis dans le domaine de l'informatique et des biotechnologies (puces à ADN). Alors que les performances de lecture des séquenceurs ADN de 1^{ère} génération se mesurent en kilobases (Kb) celles des séquenceurs de nouvelle génération s'expriment en mégabases (Mb) ou gigabases (Gb) séquencées. C'est pourquoi, on parle de séquençage de bas débit avec la méthode de Sanger et de moyen (dizaines de Mb), haut (centaines de Mb) et très haut débit (plusieurs Gb) pour les procédés NGS. Deux caractéristiques révolutionnaires distinguent radicalement le NGS du séquençage de première génération: l'approche biotechnologique permettant de générer une grande quantité de données en parallèle, et l'indispensable traitement bio-informatique des données générées.

Depuis son apparition au début des années 2000, le séquençage de nouvelle génération a rapidement progressé. Deux approches biotechnologiques très différentes ont émergé successivement, amenant à distinguer aujourd'hui chronologiquement des méthodes de séquençage de 2^{ème} puis de 3^{ème} génération.

i. Le séquençage de 2^{ème} génération

Le séquençage de 2^{ème} génération se caractérise par le séquençage simultané de millions de petits fragments d'ADN à l'aide de techniques, dites de "wash-and-scan"* (ou cycles). Ces techniques sont basées sur l'ajout de polymérase et de réactifs, qui doivent être enlevés à chaque cycle après l'incorporation des bases à lire. Les fragments d'ADN séquencés sont de tailles relativement petites (moins de 500 pb). Ils peuvent être obtenus soit directement par fragmentation d'un génome complet (WGS, *Whole Genome Sequencing*) soit indirectement lorsque le séquençage est limité à une partie spécifique du génome (séquençage ciblé). Dans ce dernier cas, les fragments d'ADN spécifiques sont, préalablement au séquençage, soit amplifiés par PCR soit sélectionnés par des procédés d'hybridation réunis sous le terme de « capture ». L'ensemble des fragments ainsi sélectionnés constitue une librairie*. Préalablement à l'étape du séquençage, une amplification clonale de la librairie d'ADN par PCR est requise afin de multiplier les brins à séquencer et de rendre détectable le signal émis lors du séquençage de chaque brin. Les séquences obtenues à partir des fragments d'ADN sont traitées informatiquement pour devenir, selon la terminologie du NGS, des *reads** pouvant être positionnés sur le génome de référence*. On dit alors que les *reads* sont alignés. Ce processus informatique appelé alignement* des *reads* débouche sur l'assemblage final en une seule séquence continue pouvant avoir la taille d'un chromosome. Les outils bio-informatiques permettent aussi d'exploiter les bases de données génomiques pour identifier et caractériser au sein des *reads* les variations de séquence par rapport à une séquence de référence. Des fichiers informatiques contenant la liste de ces variations sont générés, appelés VCF* (*Variant Calling Files*). Ce sont ces fichiers qui sont utilisés pour annoter les différentes variations identifiées : on parle alors d'annotation*.

Il existe quatre principales méthodes de séquençage de 2^{ème} génération dont les principes techniques sont très différents, expliqués ci-dessous.

1. Le pyroséquençage (technologie Roche 454)

Historiquement, il s'agit d'une technologie utilisée par le premier séquenceur NGS commercialisé avec succès dès 2005, le GS20 de *454 Life Sciences*, inaugurant l'ère du séquençage massivement parallèle (52). C'est grâce à cette technologie que fut obtenu en 2008 le premier séquençage d'un génome humain complet avec une technique de nouvelle génération (53).

La technique de pyroséquençage débute avec une amplification de la librairie d'ADN par une PCR en émulsion*.

Dans une PCR en émulsion, l'ADN simple brin est mis en contact avec des microbilles sur lesquelles sont fixées par l'extrémité 5' des amorces* A (*microbilles A*) ou B (*microbilles B*), complémentaires de la séquence de l'un ou l'autre des adaptateurs situés aux extrémités de cet ADN simple brin. En respectant un ratio de nombre de molécules d'ADN par rapport au nombre de microbilles, un seul fragment d'ADN est théoriquement fixé sur chaque microbille.

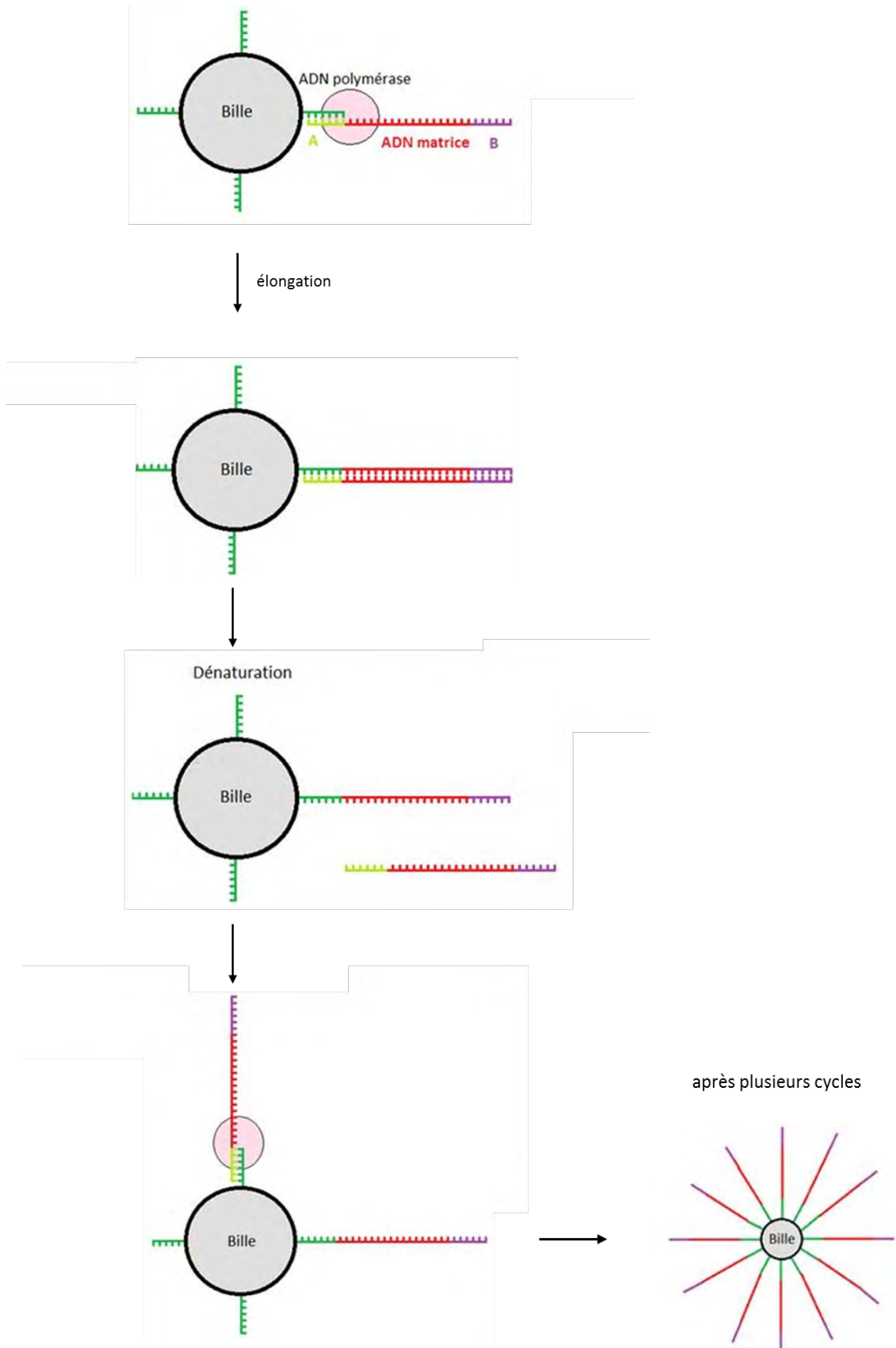
Les microbilles sont ensuite placées dans une émulsion contenant les composés nécessaires à la réaction de PCR et permettant d'individualiser chaque microbille dans une goutte.

Chaque goutte d'émulsion devient un microréacteur dans lequel l'unique fragment d'ADN fixé sur la microbille va pouvoir être amplifié par PCR. En prenant l'exemple des microbilles A, au cours du premier cycle de PCR, l'amorce A associée au fragment d'ADN initial va assurer la synthèse de son brin complémentaire. Au cycle suivant, après dénaturation, la molécule d'ADN simple néo-synthétisée reste fixée sur la microbille et va être utilisée pour la synthèse d'une autre matrice à partir de l'amorce B présente dans le milieu réactionnel. La molécule d'ADN simple brin de départ est quant à elle réutilisée par une autre amorce A fixée sur la microbille.

Après plusieurs cycles de PCR, la microbille est ainsi recouverte de plusieurs millions de molécules d'ADN simple brin identiques (amplicons* clonaux) fixées par l'amorce A (**Figure 5**).

Figure 5 : Principe de la PCR en émulsion

(<http://slideplayer.fr/1148693/3/images/6/Technologie+454.jpg>)

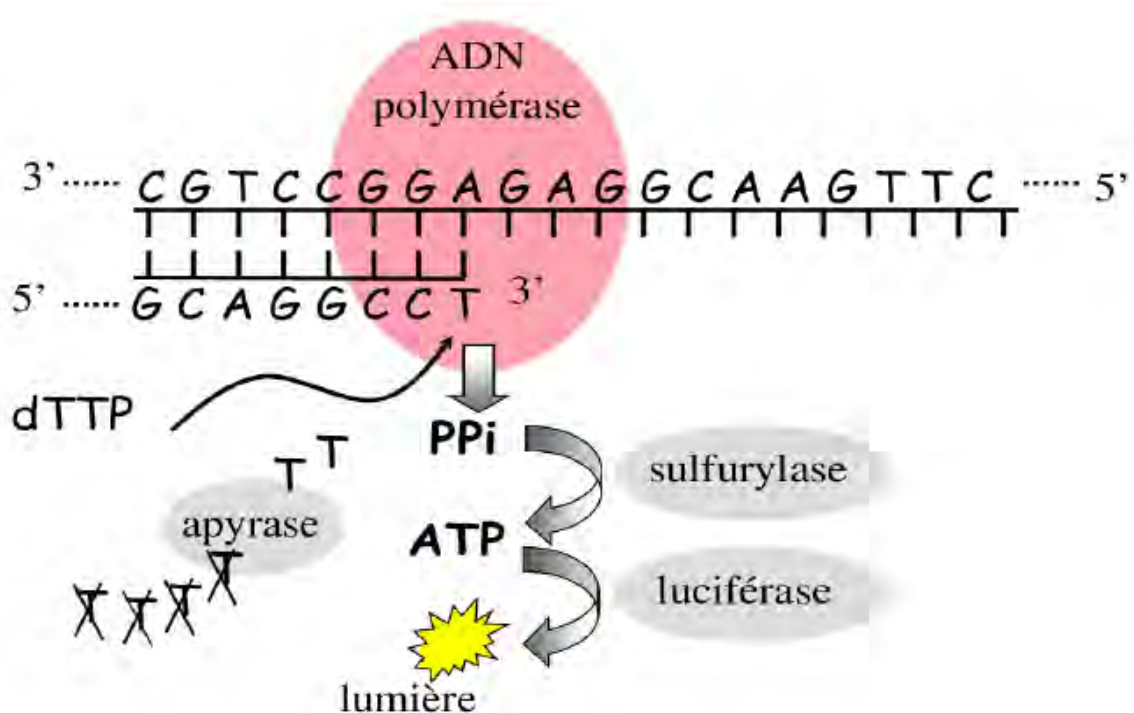


Après amplification, on dépose les microbilles recouvertes d'amplicons sur une plaque PTP (*Pico Titer Plate*) contenant plusieurs millions de puits dont les diamètres vont permettre de ne récupérer qu'une seule bille. C'est au sein de chacun de ces puits que va se dérouler la réaction de pyroséquençage.

Les types de nucléotides (A, T, C, G) sont rajoutés de manière séquentielle. Si le nucléotide ajouté correspond à celui devant être intégré dans l'ADN en cours de séquençage, son incorporation par une ADN polymérase provoque la libération d'un pyrophosphate inorganique. Par une série de réaction, il en résulte une émission de lumière. Une enzyme (apyrase) dégrade les nucléotides en surplus (non intégrés) avant chaque nouveau cycle (**Figure 6**). Le signal lumineux est capté par un capteur CDD (*Charge-Coupled Device*) et les résultats sont présentés sous forme d'un pyrogramme.

Figure 6 : Principe du pyroséquençage

D'après Falque M, 2011 (www.researchgate.net/figure/Principe-du-pyrosequencage_fig4_230690734)



2. Le séquençage par ligation et détection (technologie SOLiD : Sequencing by Oligonucleotide Ligation and Detection, Thermo Fisher Scientific)

La technologie *SOLiD* (54) a été commercialisée en 2006, soit un an après le GS20 de 454 *Life Sciences*.

Comme pour le pyroséquençage, la technologie *SOLiD* met en œuvre une PCR en émulsion pour amplifier la librairie.

Mais son principe de séquençage est très différent de celui du pyroséquençage. Premièrement, les microbilles contenant les fragments d'ADN à séquencer sont fixées à une plaque de verre spécialement traitée alors qu'elles sont réparties dans des micro-cuves lors du pyroséquençage. Deuxièmement, le séquençage s'effectue par ligation avec une ligase et non par synthèse avec une ADN polymérase. Troisièmement, le signal détecté est une fluorescence dont la longueur d'onde diffère selon le nucléotide, et non une bioluminescence.

La technique consiste à ajouter successivement des octamères (blocs de huit bases) constitués d'un dimère sur lequel la fluorescence donne une indication, suivi de 3 bases universelles, puis de dénaturer l'ADN pour éliminer le brin nouvellement synthétisé, et de recommencer la réaction de séquençage quatre fois avec un décalage d'une base à chaque fois (**Figure 8**).

En détail, chaque octamère est constitué d'un enchaînement de deux bases (dimères) en position 1 et 2 que l'on souhaite séquencer, de trois bases dégénérées à séquence universelle en position 3, 4 et 5, et de trois bases à séquence universelle possédant un fluorochrome en position 6, 7 et 8. Il existe 4 couleurs de fluorescence différentes pour 16 dimères possibles, ainsi, une même couleur de fluorochrome code pour 4 dimères différents (**Figure 7**). A chaque cycle, le fragment d'ADN est mis en contact avec 16 octamères de séquences différentes. Au cours du cycle, un octamère va se fixer par complémentarité au fragment d'ADN, et l'excitation du fluorochrome va permettre l'émission d'un signal caractéristique d'un dimère. La détection de la couleur de fluorescence donne une information partielle sur la composition du dimère. Puis les trois dernières bases sont clivées, éliminant avec elles le fluorochrome. Un nouvel octamère peut alors venir se fixer, 5 bases plus loin que l'octamère précédent. La réaction de

séquençage est relancée 4 fois, en la décalant d'une base à chaque fois, grâce à des amorces de taille différente (de 1 à 4 bases de moins que la première amorce). Ainsi, on obtient un séquençage complet permettant de séquencer chaque base 2 fois en superposant les données avec un logiciel informatique (**Figure 8**). Grâce à ce double séquençage, le taux d'erreur de séquençage de la technologie SOLiD est estimé à 1 pour 1000, soit dix fois moins que le pyroséquençage.

Figure 7 : Encodage des bases pour le séquençage SOLiD

D'après Voelkerding et al., 2009 (1)

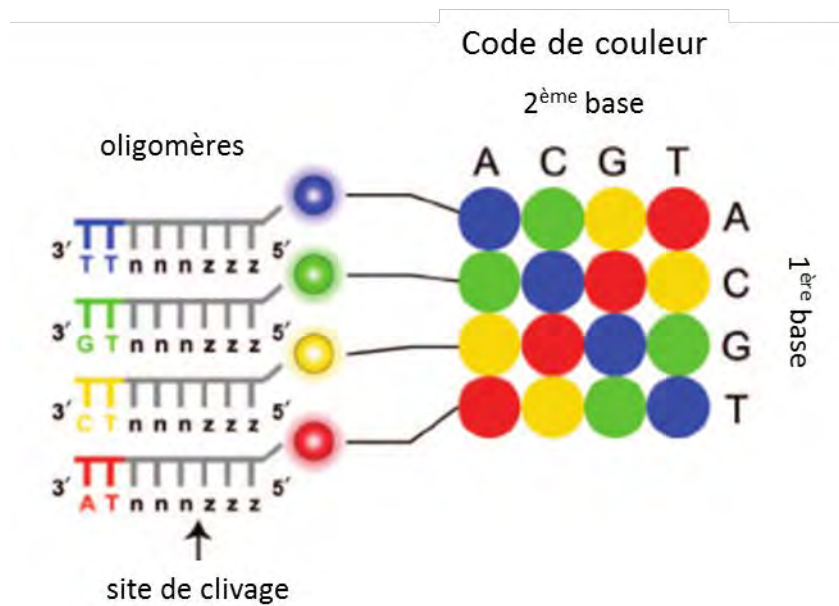
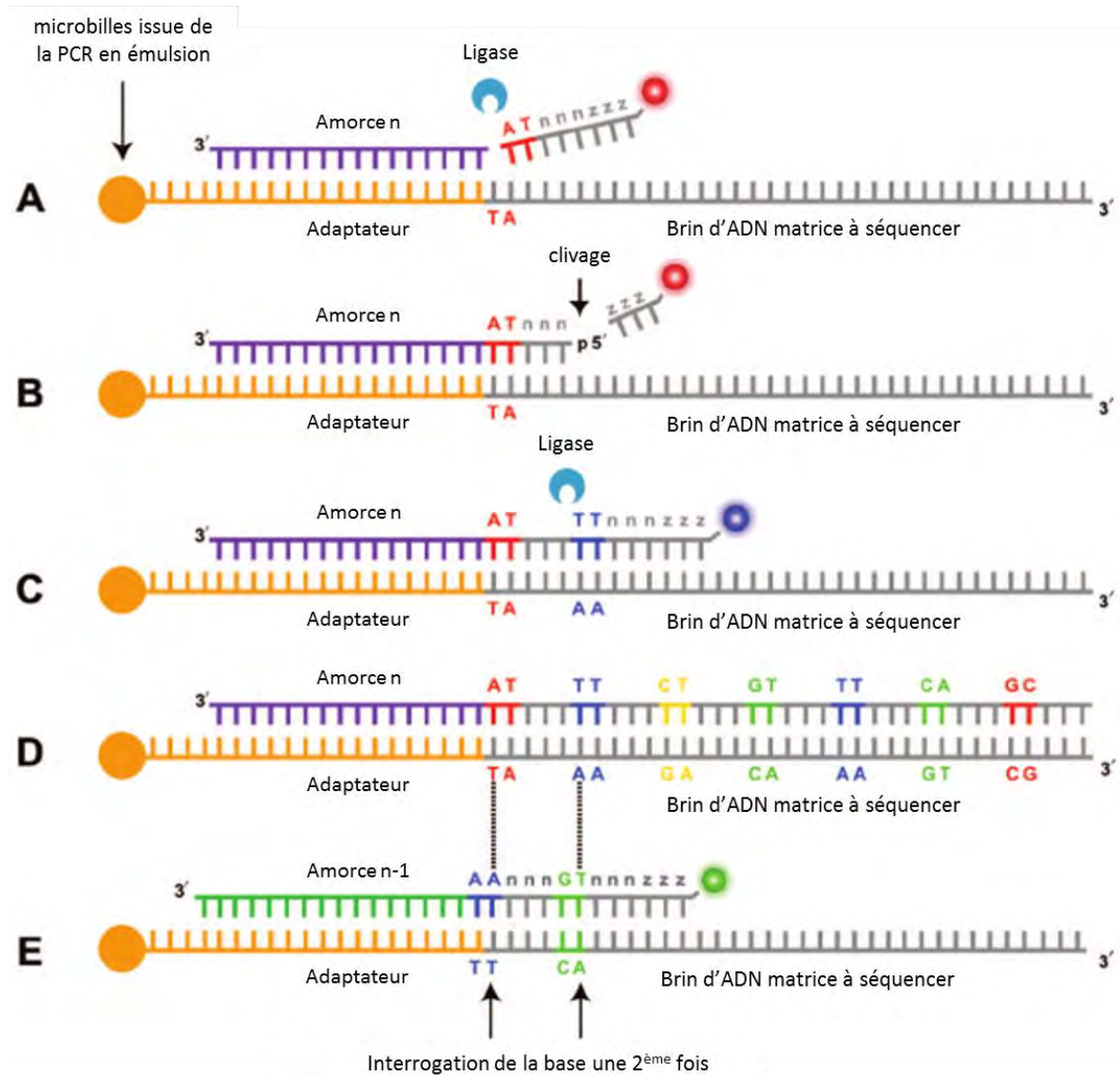


Figure 8 : Les cycles de séquençage de la technologie SOLiD

D'après Voelkerding et al., 2009 (1)

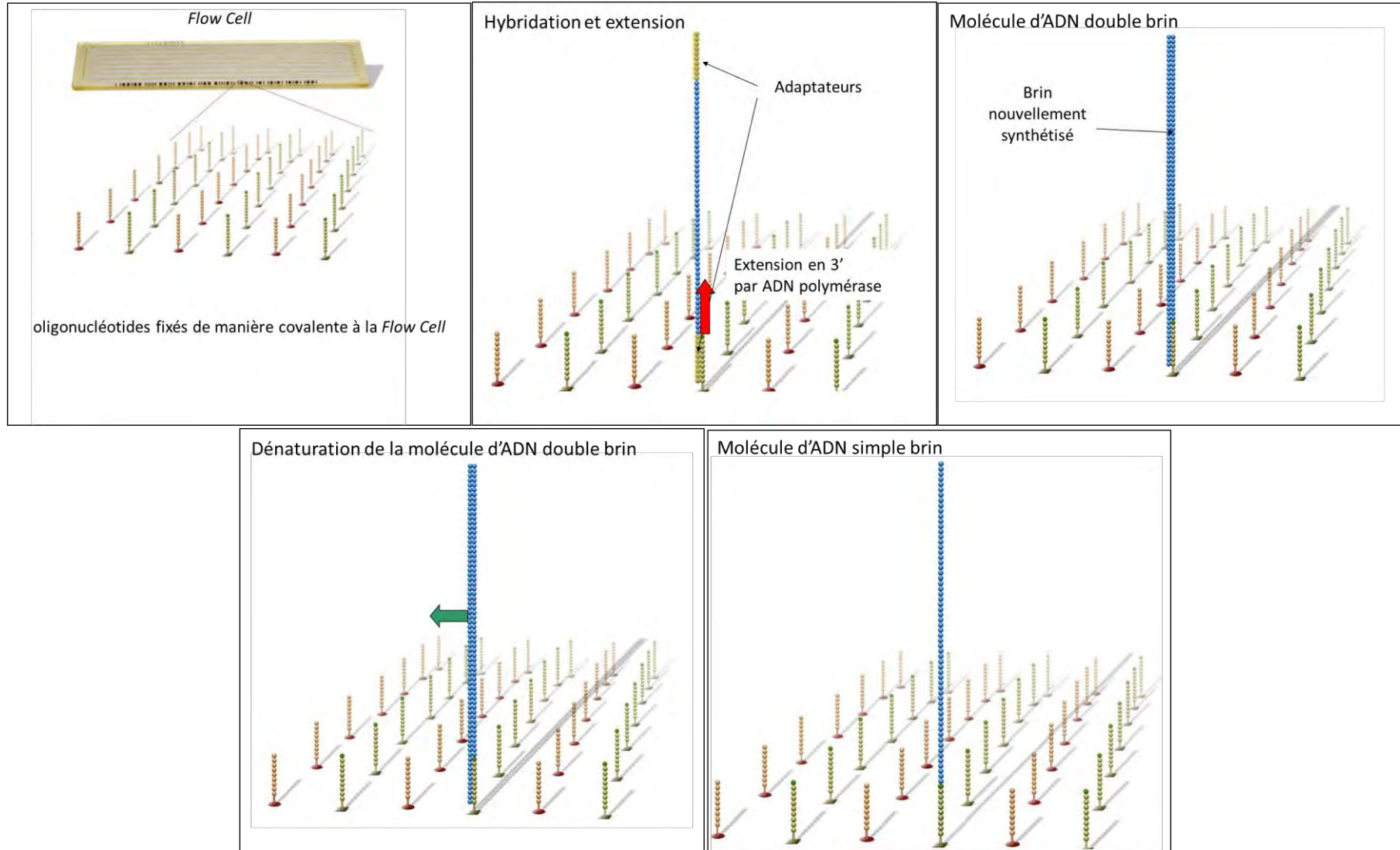


3. Le séquençage par synthèse avec terminaison réversible (technologie Solexa/Illumina)

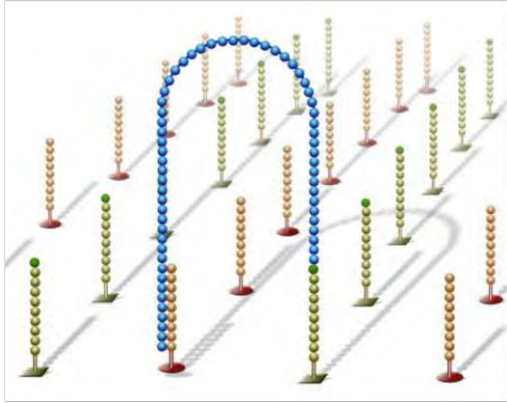
Le premier séquenceur Solexa/Illumina, le *Genome Analyzer*, a été distribué commercialement pour la première fois en 2006. Cette technologie permet de séquencer jusqu'à 1 Gb de données en un *run*.

Contrairement au pyroséquençage et au séquençage *SOLiD*, l'amplification des fragments d'ADN avant leur séquençage ne repose pas sur une PCR en émulsion mais sur une PCR sur phase solide appelée *Flow Cell* (FC) * (55). Il s'agit à l'origine d'une plaque de verre à la surface de laquelle sont fixés de manière covalente des oligonucléotides* complémentaires de la séquence des adaptateurs présents aux extrémités de chaque fragment d'ADN de la librairie. Ces oligonucléotides serviront d'amorces lors de la PCR. Déposé sur la FC, chaque fragment d'ADN dénaturé s'hybride par l'une de ses extrémités avec une amorce de la FC, puis un nouveau brin est synthétisé par une polymérase à partir de cette amorce. Le brin d'origine est éliminé par dénaturation. L'extrémité flottante du brin synthétisé se fixe ensuite sur une amorce adjacente et forme un pont d'ADN simple brin. La polymérase synthétise alors le brin complémentaire pour former un pont d'ADN double brin. Ce dernier est dénaturé pour former des fragments d'ADN simple brin qui servent de substrat pour un nouveau cycle d'amplification par pont (*bridge PCR*) * et ainsi de suite jusqu'à formation d'un groupe d'ADN clonal à la surface de la FC, appelé *cluster** (**Figure 9**). Cette méthode permet ainsi d'amplifier en parallèle un très grand nombre de molécules d'ADN différentes, les clusters monoclonaux permettant une amplification du signal lors de l'étape de séquençage.

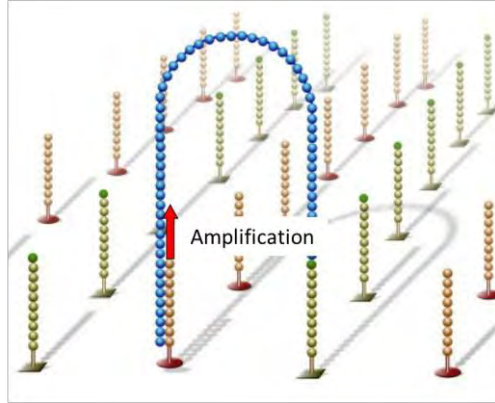
Figure 9 : Amplification par pont (*bridge PCR*) de la technologie *Illumina* (images extraites des documents *Illumina*)



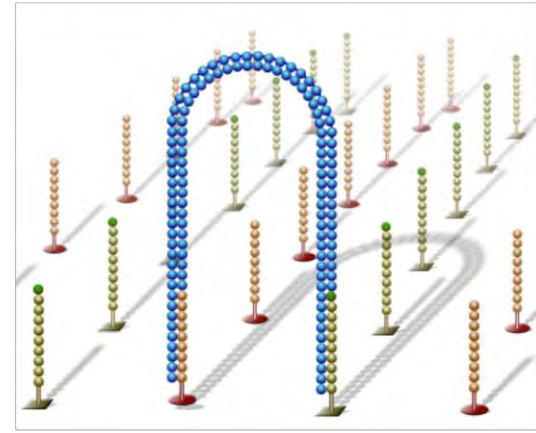
Hybridation à une amorce adjacente et formation d'un pont d'ADN simple brin



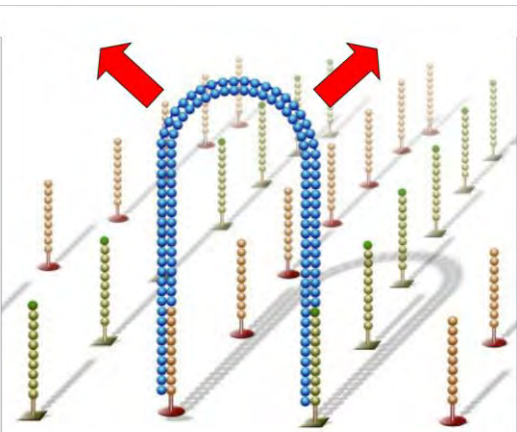
Synthèse du brin complémentaire



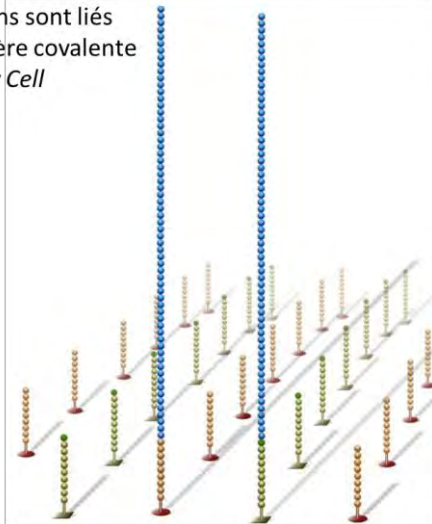
Formation d'un pont d'ADN double brin



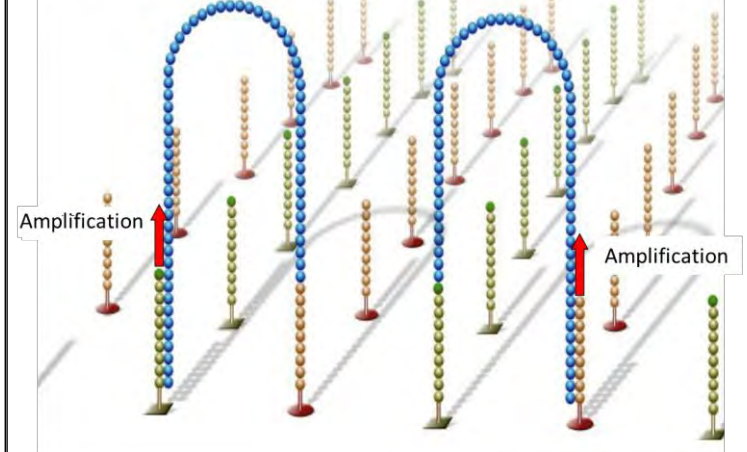
Dénaturation de la molécule d'ADN double brin



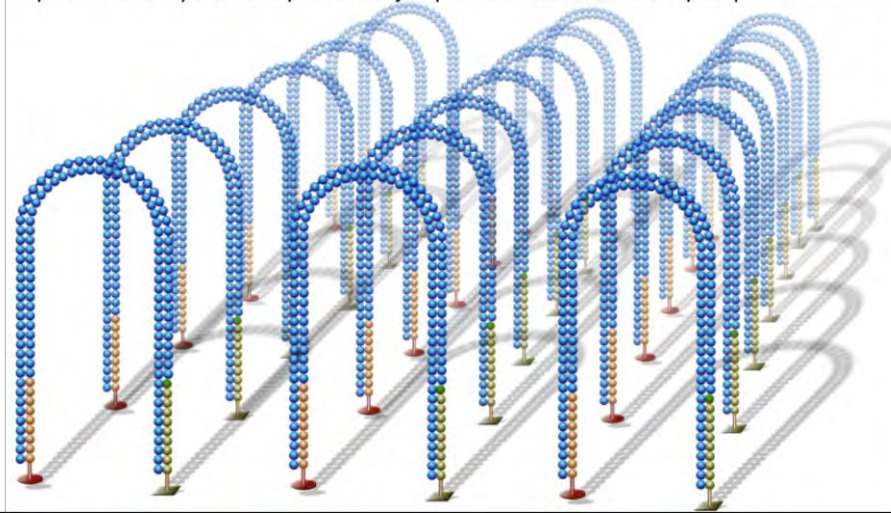
Les 2 brins sont liés de manière covalente à la Flow Cell



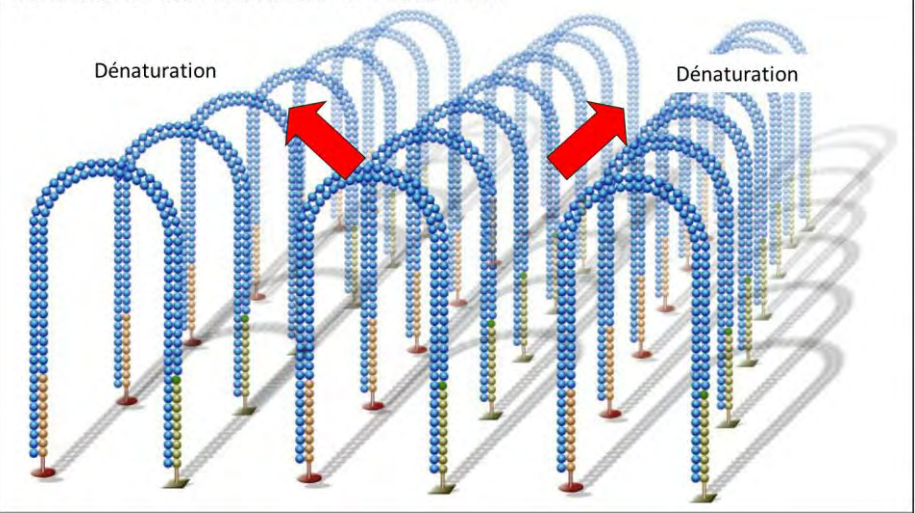
Nouveau cycle d'amplification par pont



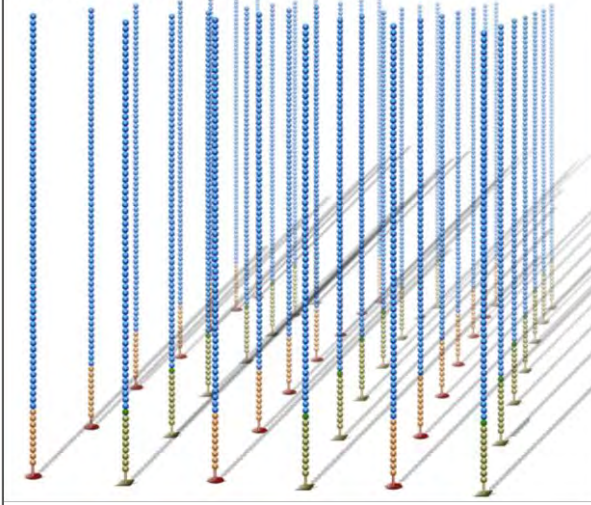
Répétition des cycles d'amplification jusqu'à l'obtention de multiples ponts



Dénaturation des molécules d'ADN double brin



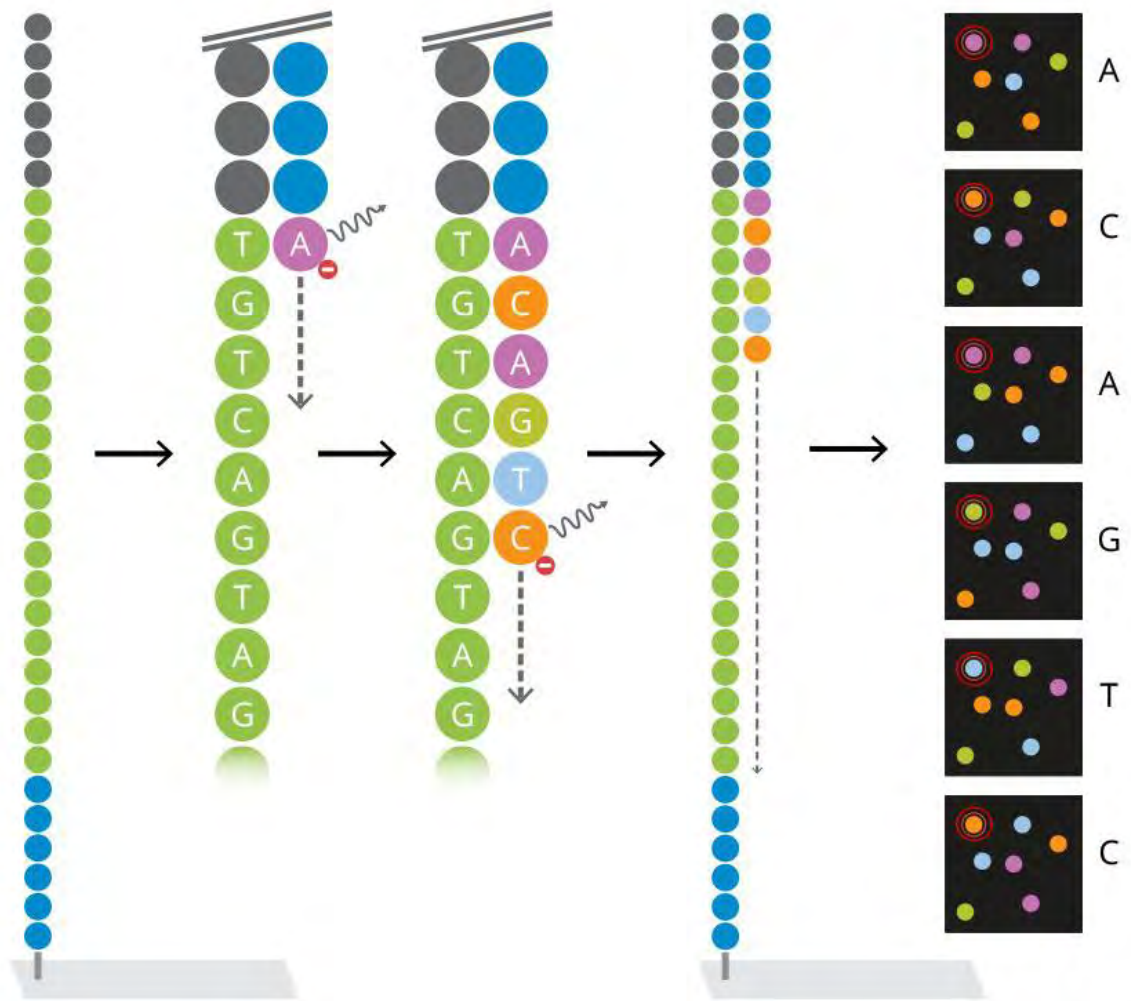
Génération de Clusters



Cette étape de *bridge PCR* est suivie par la réaction de séquençage proprement dite. Elle fait appel à une technique de séquençage par synthèse utilisant des dNTP fluorescents particuliers appelés « terminateurs réversibles » (**Figure 10**). Ces nucléotides sont modifiés par la liaison d'un fluorophore (différent pour chacun des 4 nucléotides) en position 3'OH. Le séquençage s'effectue sur l'ensemble des clusters générés précédemment par l'incorporation d'un premier dNTP modifié dont le fluorophore émet, après excitation par un laser, une fluorescence spécifique à chaque nucléotide A, T, G et C. Le fluorophore bloquant l'extrémité 3'OH, l'incorporation du nucléotide suivant ne peut avoir lieu tant que le fluorophore n'est pas clivé enzymatiquement. Cela permet d'avoir une incorporation par la polymérase qui est synchrone pour les millions de clusters de la *Flow Cell*. Une photographie par un lecteur CCD de l'ensemble des clusters est alors prise avant le clivage des fluorophores. Après un lavage de la FC, un nouveau cycle peut débuter avec l'ajout d'ADN polymérase et de dNTP terminateurs. A chaque cycle, un nouveau nucléotide est incorporé avec émission pour chaque cluster du signal fluorescent propre à ce type de nucléotide. Au final, un *read* est obtenu pour chaque cluster associé à la séquence d'un code-barres* permettant de l'assigner informatiquement à l'échantillon d'un patient.

Figure 10 : Principe du séquençage de l'ADN avec la technologie *Solexa/Illumina*

© UNIL - l'Eprouvette



4. Le séquençage par la technologie Ion Torrent™

La technologie Ion Torrent™ est apparue sur le marché des séquenceurs au début de l'année 2010. Sa grande différence avec les autres NGS est que cette technologie n'utilise pas de luminescence ou de fluorescence (56). Son procédé repose sur la mesure de variations de pH engendrées par l'incorporation de nucléotides.

En revanche, la méthode utilisée pour l'amplification clonale des bibliothèques est la même que celle du pyroséquençage ou du séquençage *SOLiD* ; il s'agit de la PCR en émulsion (cf. paragraphe II.1.b.i.1, page 55).

En amont de l'étape de séquençage, une initialisation du séquenceur est requise, ce qui permet notamment une homogénéisation des valeurs de pH ($\approx 7,8$) au sein des différents réactifs de l'appareil.

Pour mesurer les variations de pH, cette technologie est basée sur des puces semi-conductrices remplies de puits. Les microbilles recouvertes d'amplicons sont chargées sur la puce de manière à avoir une microbille par puits. Des amorces de séquençage et de l'ADN polymérase sont également ajoutées sur la puce.

Le séquençage a lieu par synthèse avec des nucléotides non modifiés. Les types de dNTP (T, A, C, G) sont rajoutés de manière séquentielle, avec une étape de lavage entre chaque. Si le dNTP ajouté correspond à celui devant être intégré, son incorporation provoque la libération d'ions H^+ . Cela entraîne une variation de pH, elle-même détectée au niveau de la couche mince (technologie des semi-conducteurs) située au fond de chaque puits (**Figure 11**).

L'ensemble des données brutes générées est transcrit sous forme d'ionogrammes (**Figure 12**).

Figure 11 : Principe du séquence par mesure du pH de la technologie Ion Torrent™

(www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torre)

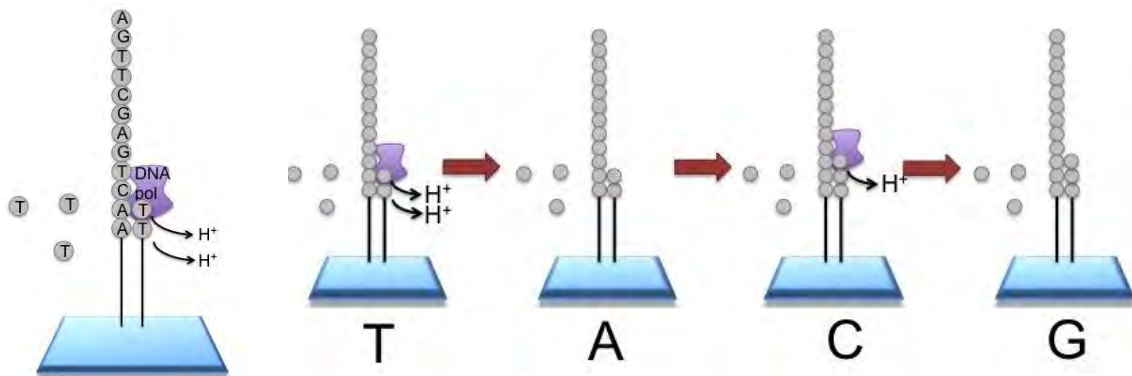
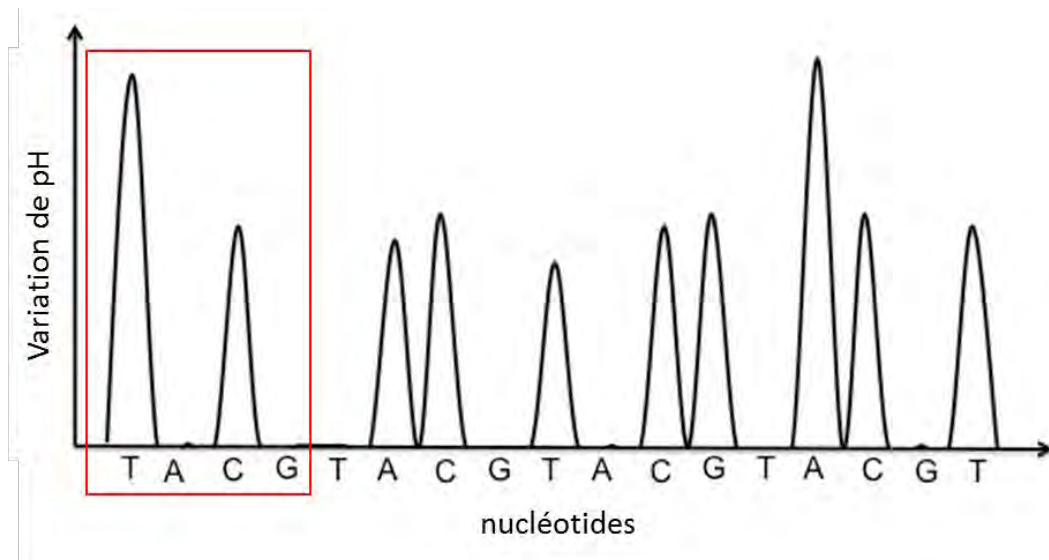


Figure 12 : Représentation des variations de pH mesurées lors du séquençage par la technologie Ion Torrent™

(www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torre)



Les principales différences entre toutes ces méthodes sont :

- la taille des oligonucléotides synthétisés et donc séquencés (longueur de lecture* en pb),
- le nombre de lectures par *run*,
- le débit (nombre de pb totales séquencées au cours du *run*),
- le temps d'exécution,
- le taux d'erreur,
- et le coût du séquençage par pb.

Elles sont résumées dans la figure ci-après (**Figure 13**).

Figure 13 : Principales caractéristiques des séquenceurs NGS de 2^{ème} génération

D'après Blervaque R., 2013 (<http://www.biorigami.com>)

Séquenceurs 2 ^{ème} génération (2013)													
Société	Roche		Illumina				Life technologies						
Plateforme													
Technologie	Titanium	GS FLX+					Chip 314 v2	Chip 316 v2	Chip 318 v2	Chip PI	Chip PII	SOLiD	SOLiD
	Acides nucléiques (matrice)												
	Ligation des adaptateurs												
Méthode d'amplification	PCR en émulsion		« Bridge PCR »				PCR en émulsion						
Méthode de séquençage	Synthèse		Synthèse				Synthèse				Ligation		
Capacité de séquençage/run	35Mb	700Mb	8Gb	95Gb	300Gb	600Gb	100Mb	1Gb	2Gb	10Gb	32Gb	95Gb	48Gb
Taille moyenne des reads	400b	700b	2x300b	2x150b	2x100/150b	2x100/150b	400b	400b	400b	200b	100b	2x60b	2x60b
Exactitude de séquençage	Q20	Q20	Q30	Q30	Q30	Q30	Q20	Q20	Q20	Q20	Q20	Q40	Q40
Coût machine + annexes	125K\$	550K\$	125K\$	300K\$	590K\$	690K\$	50K\$ + 20K\$			149K\$		600K\$	350K\$
Coût/run	1K\$	6K\$	1K\$	17K\$	11K\$	23K\$	350\$	550\$	750\$	1K\$	1K\$	10K\$	5K\$
Durée de run de séquençage	10h	23h	27h	14j	8,5j	11j	4h	5h	7h	4h	4h	6j	6j

Q20 = taux d'erreur de 1%
 Q30 = taux d'erreur de 0,1%
 Q40 = taux d'erreur de 0,01%

ii. Le séquençage NGS de 3^{ème} génération

Les techniques de séquençage NGS de 2^{ème} génération, bien qu'offrant des débits de lecture largement plus élevés que le séquençage Sanger, ont toutes en commun d'avoir des *reads* relativement courts (moins d'1 Kb). Or le génome comprend beaucoup d'éléments complexes dont la taille est supérieure à la longueur de ces *reads*, comme les régions fortement répétées et les variations de structure (SV). Les *reads* d'une région répétée sont difficiles à aligner correctement sur le génome de référence si leur taille ne permet pas un empiètement sur une région adjacente aidant à les localiser. La détection de variations de structure (SV) de grande taille est délicate avec des *reads* courts qui ne chevauchent pas forcément ce réarrangement génétique. Ceci est d'autant plus gênant que les SV sont fréquemment pathogènes, et leur implication dans la DI est reconnue (cf. partie Introduction, paragraphe I.3.b, page 38). De plus, la faible longueur des *reads* du séquençage de 2^{ème} génération ne permet pas de connaître si des variants sont localisés sur le même brin d'ADN et donc de reconstruire des haplotypes* complets. Deux variants pathogènes peuvent se trouver ou non sur le même brin d'ADN, et affecter ainsi un même allèle* ou deux allèles différents d'un même gène. Le défaut de longueur de *read* inhérent à ces techniques pose de réels problèmes pour réaliser un séquençage exhaustif du génome humain.

De plus, les régions riches en bases G et C sont difficiles à amplifier par PCR du fait de la formation possible de structures secondaires. Avec le séquençage de 2^{ème} génération, une amplification clonale par PCR est nécessaire avant l'étape de séquençage afin d'augmenter le signal détecté ; le séquençage de ces régions n'est donc pas optimum. La phase d'amplification clonale par PCR dans le séquençage de 2^{ème} génération est aussi source de biais si une erreur est amplifiée lors de la PCR.

Pour finir, le séquençage de 2^{ème} génération ne permet pas de détecter les modifications épigénétiques*, telle que la méthylation des îlots CpG des régions promotrices des gènes. Le séquençage est réalisé sur une matrice d'ADN qui correspond à une copie de la librairie initiale, l'information épigénétique (comme la méthylation) portée par les bases initiales est donc perdue.

Bien que le séquençage de 2^{ème} génération soit une importante révolution technologique, il est rapidement apparu qu'il était nécessaire de développer d'autres méthodes de séquençage de l'ADN afin de résoudre ses limites. Ainsi, le séquençage de 3^{ème} génération* a émergé peu de temps après. Il s'agit d'un procédé de séquençage de molécules uniques en temps réel ne nécessitant pas de cycles de "wash-and-scan", ni d'amplification par PCR préalable. Cela permet de réduire significativement le temps de chaque séquençage. Les *reads* ont une longueur beaucoup plus importante dans les séquenceurs de 3^{ème} génération. Ils atteignent plusieurs milliers de pb, contre des centaines de pb au maximum pour le séquençage de 2^{ème} génération. Les modifications épigénétiques de l'ADN peuvent être détectées par ces méthodes.

Deux technologies principales se sont imposées :

- celle de Pacific Biosciences (*PacBio*), commercialisée en 2011, qui a pour principe le séquençage de molécules uniques, avec observation en temps réel de l'incorporation de nucléotides fluorescents (*Single Molecule Real Time™ (SMRT) sequencing*) (57) (**Figure 14**). L'ADN polymérase est fixée au fond de structures nanophotoniques (*zero-mode waveguide, ZMW*) où se déroule l'incorporation enzymatique des nucléotides fluorophores. Ces ZMW permettent une détection confinée de l'incorporation du fluorophore à l'aide une caméra haute résolution. Le diamètre de chaque ZMW est inférieur à la longueur d'onde de la fluorescence émise, ainsi la fluorescence est restreinte au ZMW, et ne parasite pas les autres ZMW adjacentes. Le séquençage se fait à l'aide d'une puce contenant jusqu'à un million de ZMW, chaque ZMW pouvant séquencer un brin d'ADN. Ses limites principales sont son encombrement, le coût de la machine et le coût par séquençage. Le taux d'erreur est important ($\approx 15\%$) avec un séquençage unique, mais il peut être extrêmement amélioré avec un séquençage en boucle de la molécule d'ADN (moins 1 d'une erreur pour 100000 avec 20 passages) (**Tableau 1**). Cette technologie connaît un développement rapide et ses progrès sont spectaculaires année après année (58).

- celle d'Oxford Nanopore Technologies (MinION™), mise sur le marché en 2014, qui a pour principe le séquençage par nanopore (**Figure 15**). Le séquençage par nanopore s'effectue en mesurant les différences de potentiels à la surface d'un nanopore lors du passage de la molécule d'ADN simple brin à séquencer au travers de ce nanopore. Le déplacement de la molécule d'ADN à travers le nanopore est provoqué par

électrophorèse. Une molécule de cyclodextrine est positionnée sur la surface intérieure du nanopore pour agir en tant que site de liaison à des nucléotides individuels du brin d'ADN, et ainsi mesurer la différence de densité de courant électrique à leur passage (59). La densité de courant électrique est mesurée plusieurs milliers de fois par seconde. L'avantage principal de cette technologie est sa portabilité et son coût réduit, car elle ne se base pas sur des principes de fluorescence ou luminescence. La taille des *reads* n'est théoriquement pas limitée, des *reads* d'1 Mb ont pu être produits par cette technologie. Toutefois, son taux d'erreur est aujourd'hui très élevé (de 3 à 15%), et des erreurs systématiques apparaissent avec les homopolymères* (**Tableau 1**). Elle est encore considérée en développement.

Les technologies de séquençage de 3^{ème} génération sont capables de produire de très longues lectures (plusieurs Kb au minimum). Cela permet le chevauchement des régions d'ADN constituées de longues répétitions et des SV de grande taille, et revêt un intérêt pour l'assemblage de novo des génomes*.

Cependant, ces nouvelles techniques produisent encore beaucoup d'erreurs de séquençage, du fait du décryptage en temps réel (et non rythmé par des cycles), et de la non amplification du signal détecté par une PCR préalable.

Elles ne sont donc que très rarement utilisées en diagnostic pour le moment.

Par conséquent, les technologies de séquençage de 2^{ème} génération restent celles utilisées pour le diagnostic biomédical par les laboratoires de biologie moléculaire.

Figure 14 : Principe du séquençage de la technologie de *Pacific Biosciences (PacBio)*

D'après Metzker, 2010 (3)

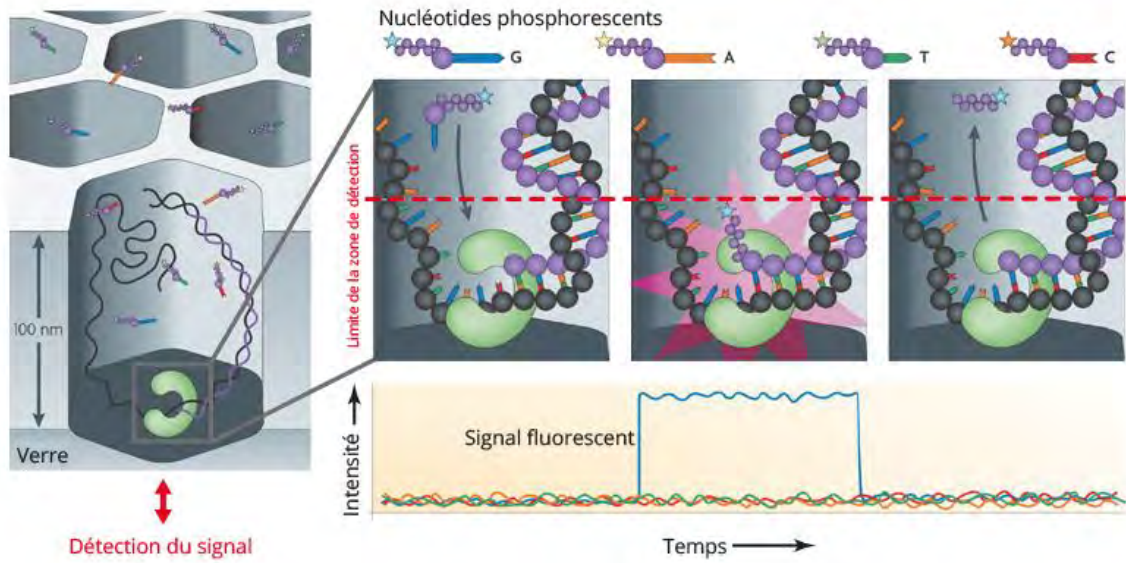


Figure 15 : Principe du séquençage par nanopore d'*Oxford Nanopore Technologies*

D'après Schaffer et MacNeill, 2012

(<http://www2.technologyreview.com/news/427677/nanopore-sequencing/>)

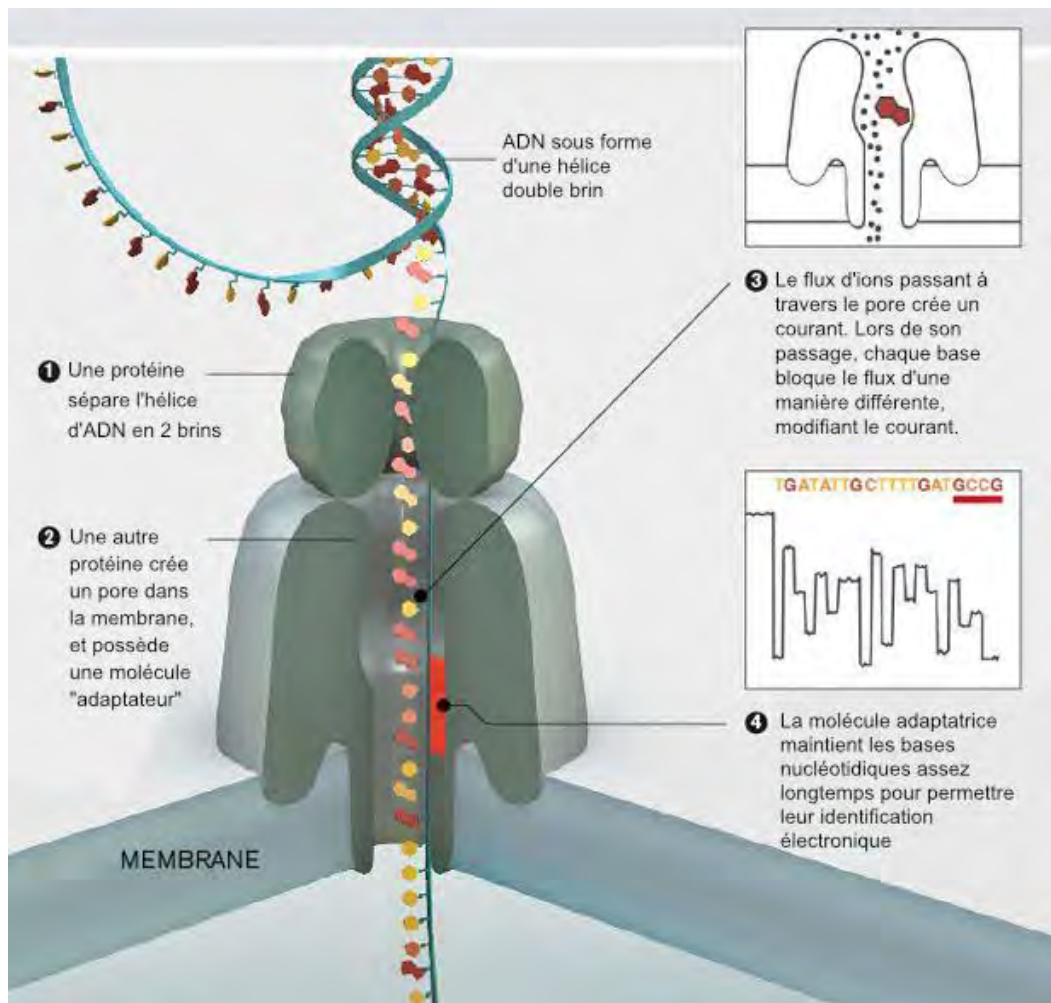


Tableau 1 : Principales caractéristiques des séquenceurs NGS de 3^{ème} génération

Plateforme	Pacific Biosciences		Oxford Nanopore Technologies		
Technologie	PacBio RSII	Sequel	MinION	GridION	PromethION
Capacité de séquençage / run	0,5 à 1 Gb	5 à 10 Gb	jusqu'à 20 Gb	jusqu'à 100 Gb (5 flow cells)	jusqu'à 125 Gb (1 flow cell) - jusqu'à 6 Tb (48 flow cells)
Taille moyenne des reads	10 à 15 Kb	10 à 15 Kb	variable	variable	variable
Taille maximum des reads	> 80 Kb	> 80 Kb	variable	variable	variable
Taux d'erreur	15% si lecture unique; <1% si lecture circulaire		15% si séquençage d'un seul brin d'ADN, 3% si séquençage des 2 brins		
Coût machine	700 000 \$	350 000 \$	1 000 \$	49 955 à 125 000 \$ selon volume de flow cells commandées	135 000 \$
Coût / run	400 \$	850 \$	475 à 900 \$	475 à 900 \$ par flow cell	625 à 2000 \$ par flow cell
Coût / Gb	400 \$	85 \$	24 \$	24 \$	5 \$
Durée de run de séquençage	jusqu'à 4 h	jusqu'à 4 h	jusqu'à 48 h	jusqu'à 48 h	jusqu'à 64 h

* la longueur des reads n'est limitée que par la taille de la molécule d'ADN de l'échantillon

D'après Van Dijk et al. 2018 (2)

2. Principes méthodologiques du séquençage NGS de 2^{ème} génération : les différentes phases techniques, bases bio-informatiques de l'analyse des données et terminologie

Les méthodes de séquençage de 2^{ème} génération comportent généralement 4 grandes phases : la préparation des librairies (banques de fragments d'ADN cible que l'on veut séquencer), l'amplification par PCR de ces librairies, le séquençage proprement dit avec la génération de données par le séquenceur, et enfin le traitement bio-informatique des données générées.

a. La préparation des librairies

Cette étape consiste à obtenir, pour chaque patient, un ensemble de fragments d'ADN d'une ou de plusieurs régions du génome désignées comme cibles du séquençage. Idéalement, ces fragments sont chevauchants et couvrent en totalité les régions cibles. La première préoccupation du biologiste sera donc d'appliquer une méthode qui lui permette d'obtenir une librairie ayant une couverture théorique* de 100 %. Deux possibilités méthodologiques s'offrent aujourd'hui à lui : soit une approche de type « amplicon », soit une approche de type « capture » des régions d'intérêt.

L'approche par amplicon repose sur la mise au point de PCR multiplexes* conduisant à une amplification sélective des régions cibles du séquençage. Un exemple de cette approche est la technologie AmpliseqTM (*Thermo Fisher Scientific*) utilisée dans ce travail de thèse (cf. partie Matériel et méthodes, paragraphe IV.2.a, page 86).

L'approche par capture s'appuie, quant à elle, sur la mise au point d'un ensemble de sondes nucléotidiques d'environ une centaine de paires de bases, complémentaires des régions d'intérêt. Ces sondes sont ainsi conçues pour s'hybrider spécifiquement à l'ADN génomique préalablement fragmenté. Divers procédés permettent ensuite de récupérer les fragments d'ADN capturés cibles du séquençage (par exemple par fixation des sondes sur des microbilles magnétiques via un système biotine/streptavidine*). En exemple, on

peut citer la technologie SureSelectQXT (*Agilent*) utilisée dans ce travail de thèse (cf. partie Matériel et méthodes, paragraphe VI.1, page 96).

Quelle que soit l'option choisie pour obtenir une librairie, le laboratoire doit concevoir l'ensemble des amorces ou de sondes permettant d'obtenir la meilleure couverture possible, en obtenant un nombre de fragments d'ADN à séquencer chevauchants suffisamment élevés. En effet, le chevauchement de différents fragments au niveau d'une séquence cible accroît la possibilité pour une base donnée d'être séquencée ou « lue » plusieurs fois. Le nombre de lectures (*reads*) des bases séquencées est un critère capital pour la précision du séquençage de 2^{ème} génération, car le taux d'erreur du séquençage est directement dépendant du nombre de lecture. On utilise le terme de profondeur de lecture* pour désigner ce paramètre. Ainsi, une profondeur de lecture dite de 30X pour une base donnée signifie que cette base a été séquencée 30 fois.

Le travail de conception de l'ensemble des amorces ou des sondes d'une librairie est par conséquent crucial pour le succès d'un séquençage ciblé de 2^{ème} génération. Les généticiens utilisent le terme de « dessin » ou « design »* de la librairie pour cette opération. Divers outils bio-informatiques aident le laboratoire à obtenir un dessin optimal. Les fabricants d'amorces de PCR ou de sondes permettent de bénéficier de librairies sur-mesure ou adaptées pour chaque test.

Quel que soit le dessin des amorces de PCR ou des sondes, tous les procédés de fabrication des librairies incluent des systèmes conditionnant les traitements ultérieurs des fragments d'ADN à séquencer. Ces fragments d'ADN de la librairie sont systématiquement liés à de courtes séquences appelées « adaptateurs » qui sont nécessaires aux étapes techniques suivantes (amplification par PCR et séquençage). Cela permet aussi d'indexer* chaque échantillon provenant de patients différents et ainsi de les traiter simultanément dans un seul et même *run* (l'adaptateur contient une séquence oligonucléotidique faisant office de code-barres).

b. L'amplification des librairies par PCR

Afin d'accroître la puissance du séquençage, le contenu en fragments d'ADN des librairies peut être amplifié par PCR. La méthode utilisée pour cette amplification clonale des librairies va différer selon le séquenceur utilisé.

L'amplification va être réalisée :

- soit par une PCR en émulsion au cours de laquelle les amplicons sont rattachés à des microbilles (cf. paragraphe II.1.b.i.1, page 55). Ce type d'amplification est utilisé par les technologies de pyroséquençage, SOLiD et Ion Torrent™ ;
- soit par une PCR avec formation de ponts (*bridge amplification*) au cours de laquelle les amplicons sont fixés à une surface solide plane, comme la *flow cell* du procédé Illumina (cf. paragraphe II.1.b.i.3, page 61).

L'objectif de cette étape est de multiplier les fragments d'ADN ciblés lors de l'étape de séquençage haut débit.

c. Le séquençage proprement dit ou comment sont générées les données du séquenceur

Le séquençage massivement parallèle d'un échantillon signifie que les millions de fragments d'ADN contenus dans cet échantillon sont autant de matrices pour des réactions de séquençage simultanées. Chaque incorporation de base génère l'émission d'un signal qui est capté et traité individuellement pour chaque fragment. Ainsi, le séquençage s'effectue en parallèle pour tous les fragments de l'échantillon. La nature des signaux résultant de la réaction de séquençage dépend de la technique utilisée.

Ces signaux sont traités informatiquement pour être convertis en une séquence de bases (*read*) dont la qualité est ensuite analysée. Ainsi, grâce à des algorithmes informatiques complexes, les logiciels du séquenceur attribuent à chaque base de la séquence obtenue un score de qualité, qui traduit la probabilité que la base appelée soit correcte. Ce premier processus analytique est appelé « *base-calling* »*.

Si le *run* comprend différents échantillons, le traitement informatique des *reads* permet aussi grâce à leur code-barres oligonucléotidique (cf. paragraphe II.2.a, page 76) d'être identifiés et analysés par échantillon spécifique. C'est le démultiplexage*.

Finalement, des fichiers informatiques de type FASTQ*, contenant les *reads* associées à leurs scores de qualité, sont générés par le séquenceur pour chaque échantillon du *run*.

d. Le traitement bio-informatique des données de séquençage

L'analyse bio-informatique des données générées par le NGS nécessite le recours à un ensemble d'outils d'analyse complémentaires. Cela comprend notamment des outils pour l'alignement des séquences et pour la détection et l'annotation des variations de séquence. L'ensemble de ces outils informatiques constitue ce qui est communément appelé le pipeline bio-informatique*.

Les données de séquençage sont tout d'abord « nettoyées », c'est l'étape de *trimming**. Les adaptateurs et les portions de *reads* de mauvaise qualité sont éliminés par un logiciel informatique.

Les séquences retenues (*reads*) doivent ensuite être positionnées dans le bon ordre afin de reconstituer le génome du patient. C'est l'étape d'alignement. Elle utilise comme modèle un génome de référence (base de données de séquences nucléiques faisant consensus pour une espèce, représentant l'assemblage du génome de différents donneurs). Le logiciel « aligne », de la manière la plus pertinente possible, sur ce génome de référence tous les *reads* de quelques centaines de bases, en comparant un à un tous les *reads* avec la séquence d'ADN du génome de référence. On obtient ainsi un fichier texte SAM* (*Sequence Alignment/Map Format*, d'extension .sam) qui contient pour chaque *read* les coordonnées génomiques sur lesquelles il a été aligné, sa séquence en nucléotides, et les scores de qualité de chacune de ses bases. Ce fichier peut être compressé dans un format binaire, le fichier BAM* (*Binary Alignment/Map Format*, d'extension .bam).

L'étape suivante, appelée *Variant-calling**, vise à détecter des variations de séquences parmi les fragments séquencés. Des algorithmes de calcul comparent, base par base, les séquences de l'ADN du patient par rapport au génome de référence pour rechercher des anomalies (modification d'une base, délétion ou insertion d'une base, de plusieurs bases, etc.). On obtient alors des fichiers VCF (*Variant Call Format*, d'extension .vcf).

Les fichiers VCF sont utilisés pour la dernière étape du processus analytique : l'annotation des séquences. Plusieurs paramètres sont pris en compte lors de l'annotation d'une variation de séquence : la fréquence dans les populations, l'impact au niveau protéidique, les conséquences fonctionnelles (conséquences sur la transcription, l'épissage*, etc.), le niveau de conservation phylogénique, etc. Des logiciels exploitant diverses bases de données et contenant des algorithmes de calculs de prédiction sont nécessaires pour réaliser cette annotation, qui doit permettre *in fine* de classer chaque variation selon un score de pathogénicité allant de 1 à 5 (1 = bénin, 2 = probablement bénin, 3 = variant de signification inconnue, 4 = probablement délétère, 5 = délétère).

Matériel et méthodes

I. Sélection d'une cohorte de patients témoins

Tous les patients témoins inclus dans cette étude ont été vus en consultation au sein du service de génétique clinique du CHU de Toulouse.

Nous avons présélectionné 16 patients DI dont l'ADN était disponible au laboratoire. Nous avons retenu 15 patients qui avaient une mutation dans les gènes inclus dans le test DI117 (**Annexe 4**), et 1 patient qui n'avait pas encore de diagnostic moléculaire confirmé (**Tableau 2**).

Afin de comparer les performances du test DI117 par rapport à celles du test DI44, nous avons sélectionné 7 patients dans cette cohorte (**Tableau 2**, Patient 1 à 7). 6 patients ont une mutation dans des gènes différents communs aux 2 tests ; 1 patient n'a pas encore de diagnostic génétique posé. Toutefois, un échantillon (**Tableau 2**, Patient 7) n'a pas atteint les critères de qualité nécessaires au séquençage sur le PGM Ion Torrent, et n'a donc pas pu être analysé via le test DI44. Seuls 6 d'entre eux ont ainsi été re-séquencés avec le test DI117 pour comparaison des résultats. Il n'a pas été jugé nécessaire d'inclure d'autres patients dans cette première analyse, le nombre important de données déjà générées étant suffisant pour une évaluation comparative des performances de la méthode.

Pour éprouver notre test DI117, nous l'avons également testé sur les 9 autres patients présélectionnés (**Tableau 2**, Patient 8 à 16).

Tableau 2 : Présentation des 16 patients témoins de notre cohorte

Présentation des 16 patients témoins de notre cohorte

	Caractéristique du variant pathogène				Sévérité de la DI	Principaux signes cliniques associés au retard psychomoteur	Age au diagnostic
	Gène concerné	Nomenclature HGVS	Hétéro ou homozygote	Hérité ou <i>de novo</i>			
Patient 1	ANKRD11	NM_001256182.1:c.7535G>C p.(Arg2512Pro)	hétérozygote	<i>de novo</i>	légère à modérée	petite taille postnatale (-2 DS) syndrome dysmorphique crano-facial anomalies dentaires anomalies des extrémités (brachydactylie, clinodactylie des V ^{ème} bilatérale) fossette sacro-coccygienne troubles du sommeil déficit auditif trouble du langage troubles visuels (strabisme, hypermétropie, astigmatisme)	10 ans
Patient 2	CDKL5	NM_003159.2:c.719G>C p.(Ser240Thr)	hétérozygote	Hérité (mais inactivation de l'X muté chez la mère)	non évaluée	TSA microcéphalie progressive postnatale (-3DS à 1 an) troubles visuels (strabisme, hypermétropie)	2 ans
Patient 3	/	/	/	/	non évaluée	∅	/
Patient 4	RAI1	NM_030665.3:c.1958_1959delTG p.(Val653Glyfs35*)	hétérozygote	<i>de novo</i>	modérée	troubles du comportement (hyperactivité, troubles obsessionnels compulsifs) syndrome dysmorphique crano-facial troubles du sommeil (réveils précoces)	14 ans
Patient 5	SCN8A	NM_014191.3:c.3943G>C p.(Val1315Leu)	hétérozygote	<i>de novo</i>	sévère	TSA épilepsie	11 ans
Patient 6	MECP2	NM_004992.3:c.806delG p.(Gly269Alafs*19)	hétérozygote	<i>de novo</i>	non évaluée	régression psychomotrice depuis l'âge de 10 mois microcéphalie progressive postnatale (-2DS à 18 mois) syndrome pyramidal stéréotypies difficultés alimentaires	1 ans et 7 mois
Patient 7	ARID1B	NM_020732.3:c.1871delC p.(Pro624Hisfs*44)	hétérozygote	<i>de novo</i>	sévère	syndrome dysmorphique crano-facial troubles du comportement (instabilité motrice, difficultés de sociabilisation) anomalie de la pilosité cryptorchidie bilatérale infections récidivantes ORL (otites chroniques) Anomalies à l'IRM (hypoplasie du corps calleux et kyste arachnoïdien volumineux)	7 ans
Patient 8	ANKRD11	NM_001256182.1:c.3066del p.(Thr1023Glnfs*295)	hétérozygote	<i>de novo</i>	légère	petite taille pré et postnatale syndrome dysmorphique crano-facial pectus recurvatum déficit auditif bilatéral troubles de la déglutition néonatales hypotonie néonatale	20 ans

	Caractéristique du variant pathogène				Sévérité de la DI	Principaux signes cliniques associés au retard psychomoteur	Age au diagnostic
	Gène concerné	Nomenclature HGVS	Hétéro ou homozygote	Hérité ou <i>de novo</i>			
Patient 9	ARID1B	NM_020732.3:c.4275T>A p.(Tyr1425*)	hétérozygote	<i>de novo</i>	non évaluée	retard de croissance staturo-pondéral (taille : -0,75DS, poids: -2DS) syndrome dysmorphique cranio-facial pilosité au niveau du front hyperéchogénéicité rénale bilatérale	4 ans et 4 mois
Patient 10	EEF1A2	NM_001958.3:c.370G>A p.(Glu124Lys)	hétérozygote	<i>de novo</i>	non évaluée	épilepsie	6 ans
Patient 11	MAGEL2	NM_019066.4:c.461_481del p.(Pro154_Ala160del)	en mosaïque	<i>de novo</i>	modérée	syndrome dysmorphique cranio-facial obésité troubles du comportement alimentaire	19 ans
Patient 12	SCN2A	NM_021007.2:c.1199C>G p.(Thr400Arg)	hétérozygote	<i>de novo</i>	profonde	encéphalopathie épileptique microcéphalie (>-3 DS) cassure pondérale à 6 mois hypotonie axiale syndrome tétrapyrimal incapacité à s'alimenter (sonde naso-gastrique) hypotonie néonatale sommolence Anomalies à l'IRM (atrophie cortico sous corticale et dilatation ventriculaire)	11 ans et 7 mois
Patient 13	SYNGAP1	NM_006772.2:c.1746delG p.(Asp583Thrfs*67)	hétérozygote	<i>de novo</i>	non évaluée	microcéphalie progressive postnatale épilepsie	6 ans et 5 mois
Patient 14	IQSEC2	NM_001111125.2:c.2776C>T p.(Arg926*)	hétérozygote	<i>de novo</i>	non évaluée	régression psychomotrice TSA (difficulté de contact, hétéroagressivité) épilepsie stéréotypies	4 ans et 6 mois
Patient 15	MECP2	réarrangement complexe (délétion de la partie 3' de l'exon 4 avec duplication de l'exon 3 et de la partie 5' de l'exon 4)	hétérozygote	<i>de novo</i>	légère	TSA troubles du comportement (difficultés attentionnelles) troubles du sommeil (réveils fréquents à type de terreurs nocturnes) hypotonie globale	8 ans
Patient 16	GRIN1	NM_007327.3:c.2452A>C p.(Met818Leu)	hétérozygote	<i>de novo</i>	non évaluée	encéphalopathie épileptique troubles visuels (hypermétropie) mouvements anormaux (bras, opsoclonies)	10 mois

II. Recueil des données de séquençage (fichier .bam, .bai, et .vcf) de certains patients établies dans d'autres laboratoires

Au sein de notre cohorte, le diagnostic moléculaire de 5 patients (**Tableau 2**, Patient 1, 3, 4, 9, et 10) faisait suite à un séquençage haut débit réalisé sur un panel dédié à la DI dans d'autres laboratoires français de génétique moléculaire, ceux des CHU de Marseille et Strasbourg. Une étroite collaboration nous a permis de récupérer les données de séquençage (fichiers .bam et .bai), ainsi que les données issues de leur pipeline d'analyse bio-informatique (fichiers .vcf) pour comparaison.

Le laboratoire de Strasbourg possède ainsi un test de 456 gènes impliqués dans les troubles cognitifs (TC-456), mis à jour récemment en un test de 520 gènes (TC-520). Les tests TC-456 et TC-520 disposent respectivement de 92 et 110 gènes en commun avec notre test DI117. Le CHU de Marseille propose un test de 142 gènes responsables de DI (DI-142) qui présente 73 gènes en commun avec le test DI117. Les gènes communs à l'ensemble des 4 tests (TC-456, TC-520, DI-142 et DI117) sont au nombre de 67 (**Annexe 1**).

Les principales caractéristiques techniques de ces tests sont résumées dans le **Tableau 3** ci-dessous.

Tableau 3 Tableau récapitulatif des caractéristiques techniques des différents tests

Différents tests NGS de DI	Méthode d'enrichissement des librairies	Technologie de séquençage	Nombre de patient(s)
Test TC-456 (CHU de Strasbourg)	Capture de séquence SureSelect XT2 (<i>Agilent</i>)	HiSeq, Illumina	1 patient
Test TC-520 (CHU de Strasbourg)	Capture de séquence Seq Cap Ez (<i>Roche</i>)	NextSeq550, Illumina	1 patient
Test DI-142 (CHU de Marseille)	Capture de séquence SureSelect XT Inherited Diseases Panel (<i>Agilent</i>)	Ion Proton, Thermo Fisher Scientific	3 patients

III. Extraction des ADN

Les prélèvements sanguins des patients témoins ont été effectués après obtention par écrit de leur consentement éclairé conformément à la réglementation en vigueur.

L'ADN a été extrait et purifié à partir de sang périphérique (prélèvement sur EDTA) à l'aide de l'extracteur MagnaPure Compact (*Roche*) utilisant la technologie des billes magnétiques, avec le kit Magna Pure Compact Nucleic Acid Isolation Kit I - Large Volume (ref. 03730972001, *Roche*).

Chaque échantillon d'ADN a été dosé à l'aide du spectrophotomètre Nanodrop NanoDrop ND-1000 (*Labtech*) afin de déterminer sa concentration et s'assurer de sa pureté (rapport $DO_{260nm}/DO_{280nm} > \text{à } 1,7 \text{ et } < \text{à } 2,15$).

IV. Mise en pratique expérimentale du test DI44

1. Caractéristiques techniques du test DI44

Le test DI44 est conçu pour cibler les régions codantes des 44 gènes les plus fréquemment impliqués dans la DI en France, avec 10 nucléotides de marge de part et d'autre.

Sa couverture théorique est de 99,94 % pour une taille des régions ciblées de 181,69 kb.

La préparation des bibliothèques est réalisée par amplification avec la technologie Ampliseq™ de *Thermo Fisher Scientific*. Le séquençage NGS a lieu sur un séquenceur PGM™ Ion Torrent™.

2. Préparation de la librairie jusqu'au séquençage sur le PGM™ Ion Torrent™

Le protocole utilisé est celui fourni avec le système et mis au point par *Thermo Fisher Scientific* (**Annexe 2**).

a. Préparation de la librairie

Cette étape utilise la technologie Ampliseq™ de *Thermo Fisher Scientific*.

La technique Ampliseq™ repose, tout d'abord, sur l'amplification de l'ADN génomique de chaque patient par PCR multiplexe (**Figure 16**) grâce aux 2 pools d'amorces du test DI44 (incluant respectivement 1108 et 1116 amorces) et au pool d'amorces « Sample ID ». Le test DI44 est constitué de 1112 amplicons au total, dont la taille varie entre 125 et 375pb. Les « Sample ID » sont des amorces qui amplifient des variations nucléotidiques connues et fréquentes, ils permettent ainsi l'identification des patients sous forme d'un code unique.

Les amorces de PCR sont ensuite partiellement éliminées par digestion enzymatique (**Figure 16**).

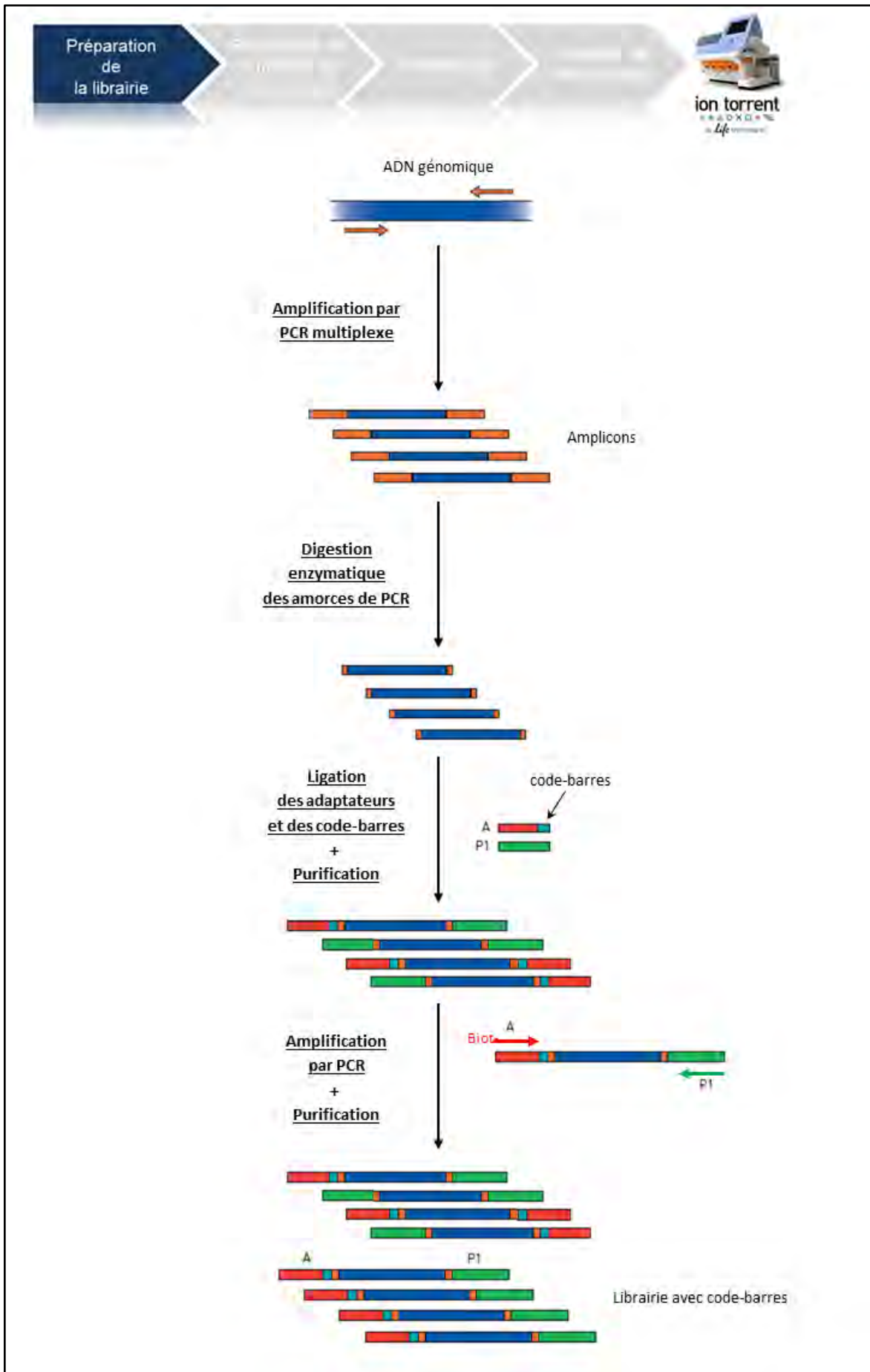
Puis, on réalise une ligation des adaptateurs de séquençage, A et P1, et des codes-barres, pour individualiser chaque patient et permettre leur identification ultérieure par l'analyse bio-informatique (**Figure 16**).

Les amplicons sont ensuite purifiés sur billes magnétiques grâce au système Agencourt AMPure XP (*Beckman Coulter*), pour éliminer les sels et les enzymes.

Une réamplification avec les amorces spécifiques A (biotinylées) et P1 a lieu afin d'enrichir la librairie en produit « ligué » correctement avec les amorces A en 5' et P1 en 3' (**Figure 16**).

Enfin, les amplicons sont purifiés de nouveau sur billes magnétiques grâce au système Agencourt AMPure XP (*Beckman Coulter*), pour éliminer les sels, les enzymes, les amorces et les dNTPs non utilisés.

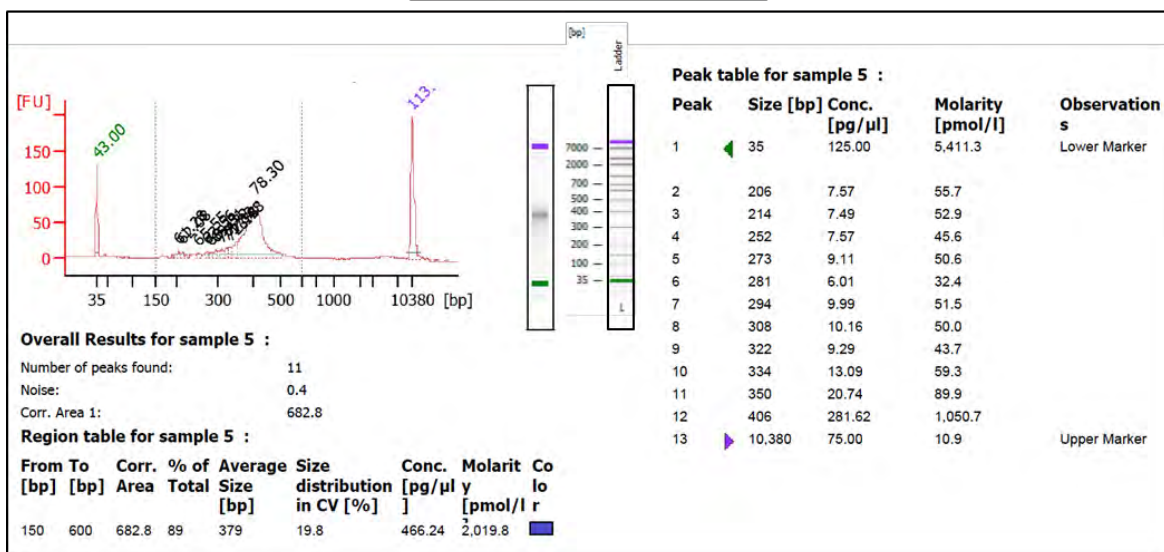
Figure 16 : Préparation de la librairie avec la technologie Ampliseq™ de *Thermo Fisher Scientific*



La librairie est contrôlée sur puce DNA High sensitivity à l'aide du Bioanalyseur 2100 (Agilent) avec comme objectif de calculer sa concentration et d'identifier la taille moyenne des fragments qui la composent (**Figure 17**). Ces valeurs permettront ainsi de déterminer la concentration molaire de la librairie et d'y appliquer le facteur de dilution nécessaire pour favoriser le ratio idéal 1/1 (Fragment ADN de la librairie/*Ion Sphere Particle*) pour l'étape suivante de PCR en émulsion.

Les librairies des patients préalablement diluées sont, pour finir, combinées de façon équimolaire.

Figure 17 : Photographie du Bioanalyseur 2100 (Agilent) et de la puce utilisée, et visualisation d'un exemple de résultats obtenus pour un échantillon donné



b. Préparation de la matrice de séquençage

Cette étape permet l'amplification clonale suivi d'un enrichissement en « *Ion Sphere Particles* » (ISPs) à la surface desquelles un fragment de la librairie est amplifié.

i. Amplification clonale

La PCR en émulsion (emPCR) réalisée sur le système Ion OneTouch™ 2 (OT2) permet l'amplification clonale en microréacteurs « *Ion Sphere Particles* » (ISPs) à la surface desquelles un fragment de librairie est amplifié (**Figure 18**). Cette étape contribue à atteindre un seuil de détection du signal nécessaire et suffisant au moment du séquençage.

Malgré une optimisation du ratio 1/1 (ISP / Fragment ADN de la librairie), plusieurs configurations de microréacteurs sont envisageables : un amplicon par ISP (amplification monoclonale), plusieurs amplicons par ISP (amplification polyclonale), aucun amplicon (ISP vide) (**Figure 19**). Seule la configuration de monoclonalité est souhaitée car elle seule est source de données de séquençage. Les autres configurations génèreront des données qui seront filtrées lors de la primo-analyse par les logiciels bio-informatiques de la « Torrent suite ».

Figure 18 : Préparation de la matrice de séquençage : étape d'amplification clonale par PCR en émulsion (emPCR) réalisée sur le système Ion OneTouch™ 2 (OT2)

(image extraite des documents du fournisseur)

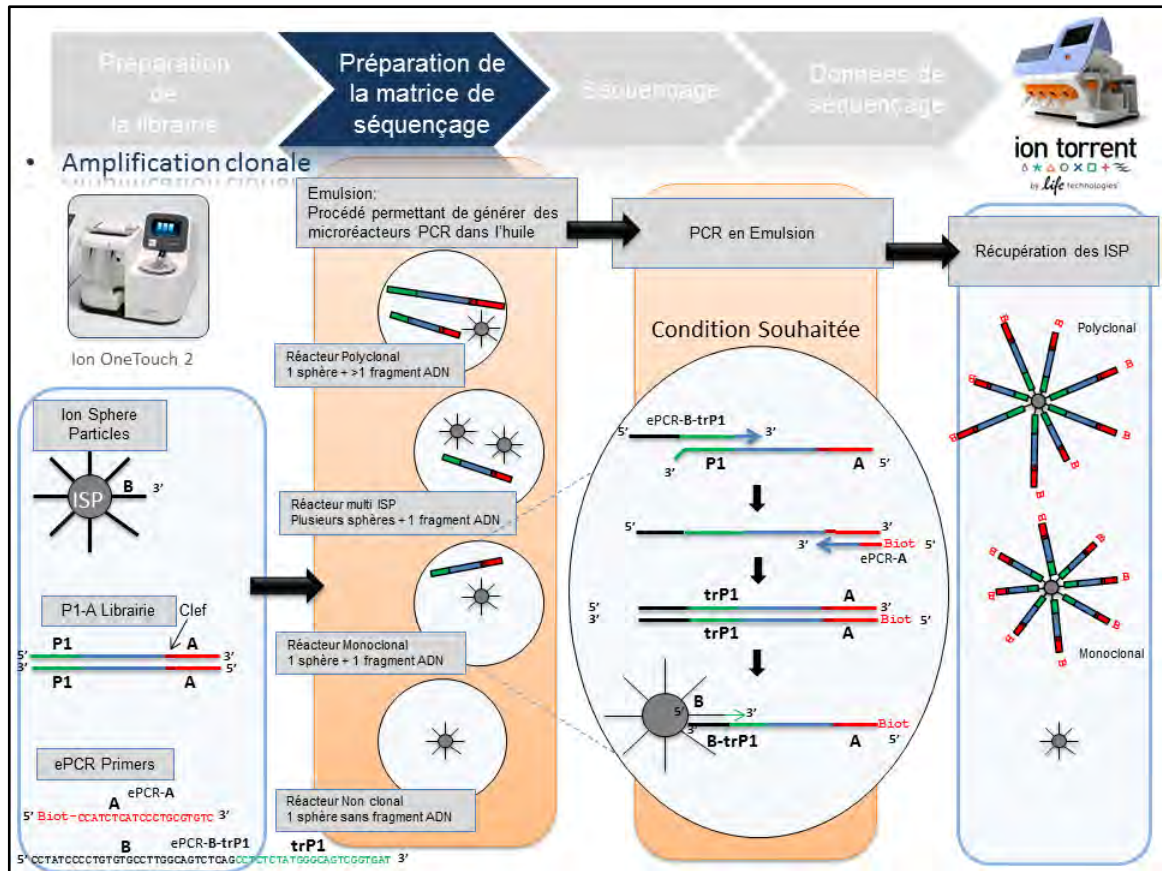
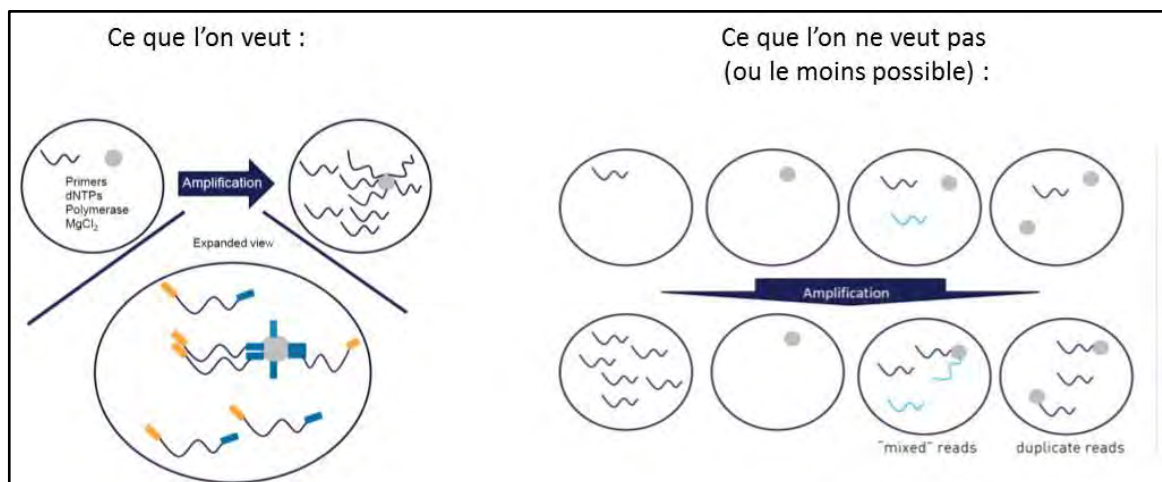


Figure 19 : Configuration de microréacteur souhaitée parmi les différentes possibilités

(image extraite des documents du fournisseur)



ii. Enrichissement en ISP contenant des fragments d'ADN

Cette étape, réalisée sur le module Ion OneTouch™ ES (ES), permet la sélection des ISP sur lesquelles sont fixés les fragments (monoclonaux + polyclonaux) et donc l'élimination des ISP vides. Des billes magnétiques liées à la streptavidine vont capturer et sélectionner les ISP non vides par liaison avec les amorces A biotinylées (**Figure 20**).

Figure 20 : Préparation de la matrice de séquençage : étape d'enrichissement en ISP contenant des fragments d'ADN réalisée sur le module Ion OneTouch™ ES (ES)

(image extraite des documents du fournisseur)

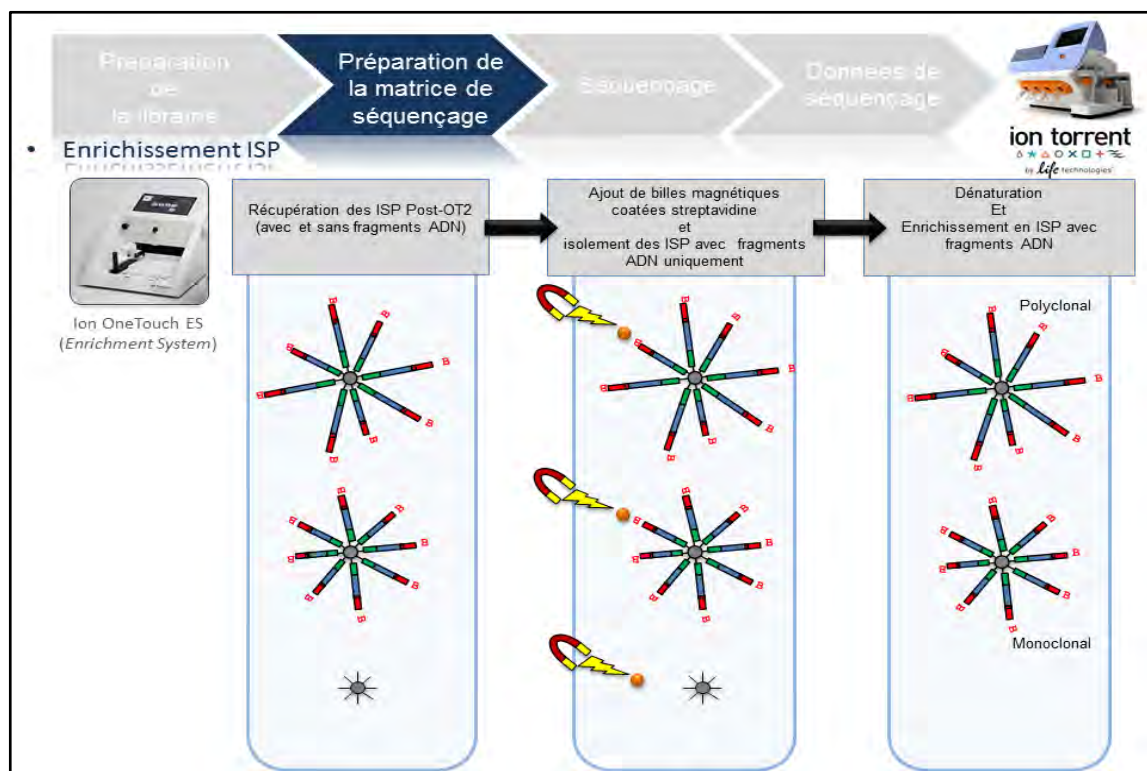


Figure 22 : Photographie du séquenceur PGM™ Ion Torrent™

(www.thermofisher.com)



A l'issue du *run* de séquençage, le fichier DAT* (extension .dat) regroupe l'ensemble des données brutes (ionogrammes). Ces fichiers sont transférés du séquenceur vers le serveur informatique du *Torrent Server*. Les fichiers DAT sont alors convertis en fichiers WELLS* (extension .wells). Les fichiers WELLS sont ensuite traités par le logiciel de « *base-calling* » (conversion des données en séquences de nucléotides – *reads* - associés chacun à un score de qualité) (**Figure 23**).

Un prétraitement bio-informatique des *reads* générés est appliqué. Il est constitué :

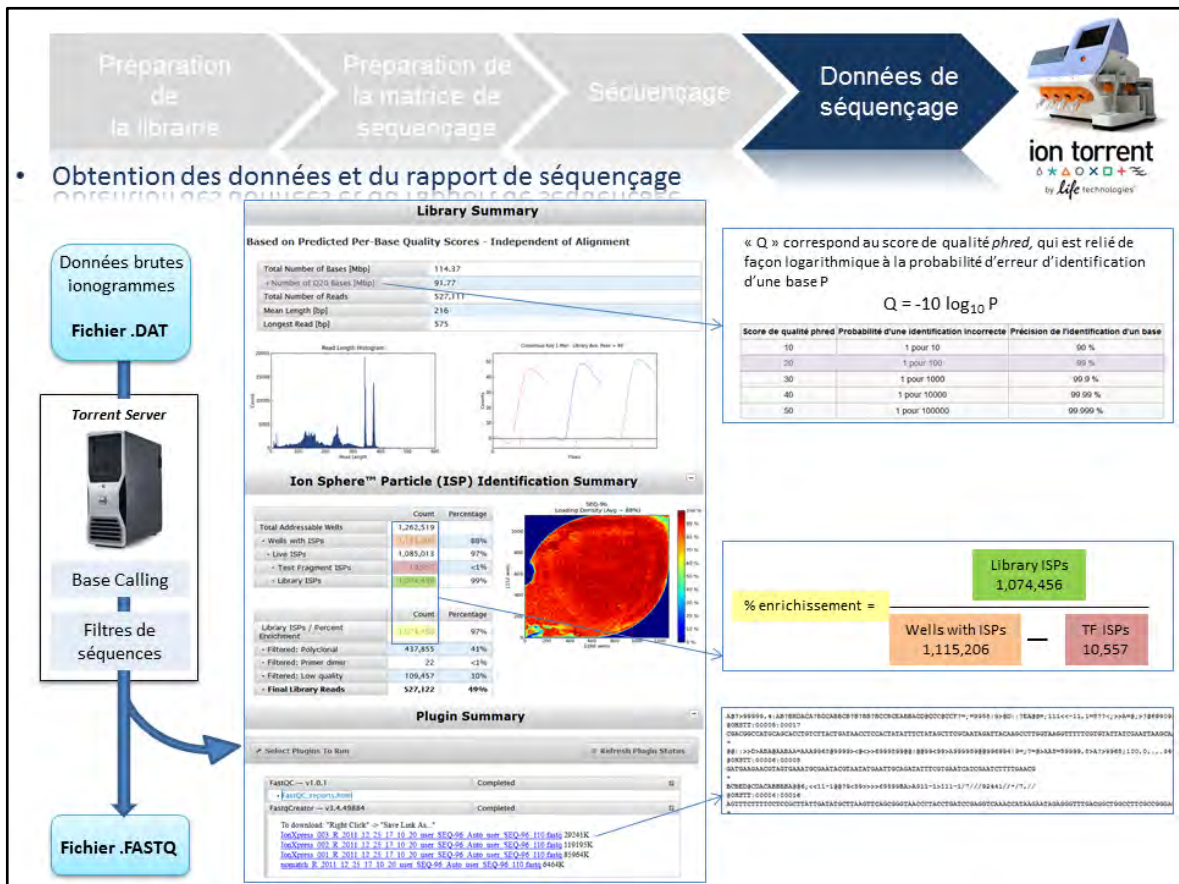
- d'une étape de *trimming* qui élimine les adaptateurs et les portions de *reads* de mauvaise qualité ;
- et de filtres qui éliminent les *reads* de petites tailles, de mauvaise qualité, et les polyclonaux.

Cette étape permet ensuite d'aligner les *reads* et de générer des données dites brutes sur les séquences nucléotidiques. Ces données sont stockées sous la forme de fichiers UBAM (extension .ubam) pouvant ensuite être convertis en fichiers BAM (extension .bam) (après alignement) et en fichiers FASTQ (extension .fastq) (données de séquençage associées à des valeurs de qualité) (**Figure 23**).

L'ensemble de ces informations est repris au travers du « report » généré à l'issue du séquençage et du prétraitement. Y sont également renseignés, le nombre de *reads* générés ainsi que leur taille moyenne (Figure 23).

Figure 23: Génération des données de séquençage par le logiciel du PGM™ Ion Torrent™

(image extraite des documents du fournisseur)



V. Conception des sondes de capture du test DI117 pour un séquençage sur technologie Illumina

Le développement de notre test a été réalisé en étroite collaboration avec le Dr Amélie Piton du laboratoire de génétique moléculaire du CHU de Strasbourg.

Le logiciel SureDesign (<https://earray.chem.agilent.com/suredesign/>) a été utilisé pour concevoir les sondes de capture.

Les paramètres de sélections ont été les suivants :

- Région : Régions codantes \pm 10 pb

Groupe de sondes	Tiling Density	Masking	Boosting	Extension into repeat	Strand
« DI44modera »	3x	Moderately stringent	MaximizePerformance	20	Sens
« 70genes a ajouter »	2x	Moderately stringent	Balanced	20	Sens
« Pws toulouse »	3x	Least Stringent	MaximizePerformance	20	Sens
« DI44 V03.1 least »	3x	Least Stringent	MaximizePerformance	20	Sens
« DI44 diffic regions »	5x	Least Stringent	MaximizePerformance	20	Sens
« 70genes_no masking »	4x	No Masking	Balanced	20	Sens

VI. Préparation de la librairie du test DI117 en capture et séquençage sur le MiSeq Illumina

1. Préparation de la librairie

La préparation des librairies a été réalisée avec la technologie SureSelectQXT (*Agilent*) selon le protocole délivré par le fournisseur (**Annexe 3**).

La première étape consiste à normaliser les échantillons d'ADN génomique à une concentration de 25 ng/μl au moyen de 2 dosages successifs, réalisés avec le kit Qubit dsDNA BR ASSAY (*Thermo Fisher Scientific*) selon les recommandations du fournisseur.

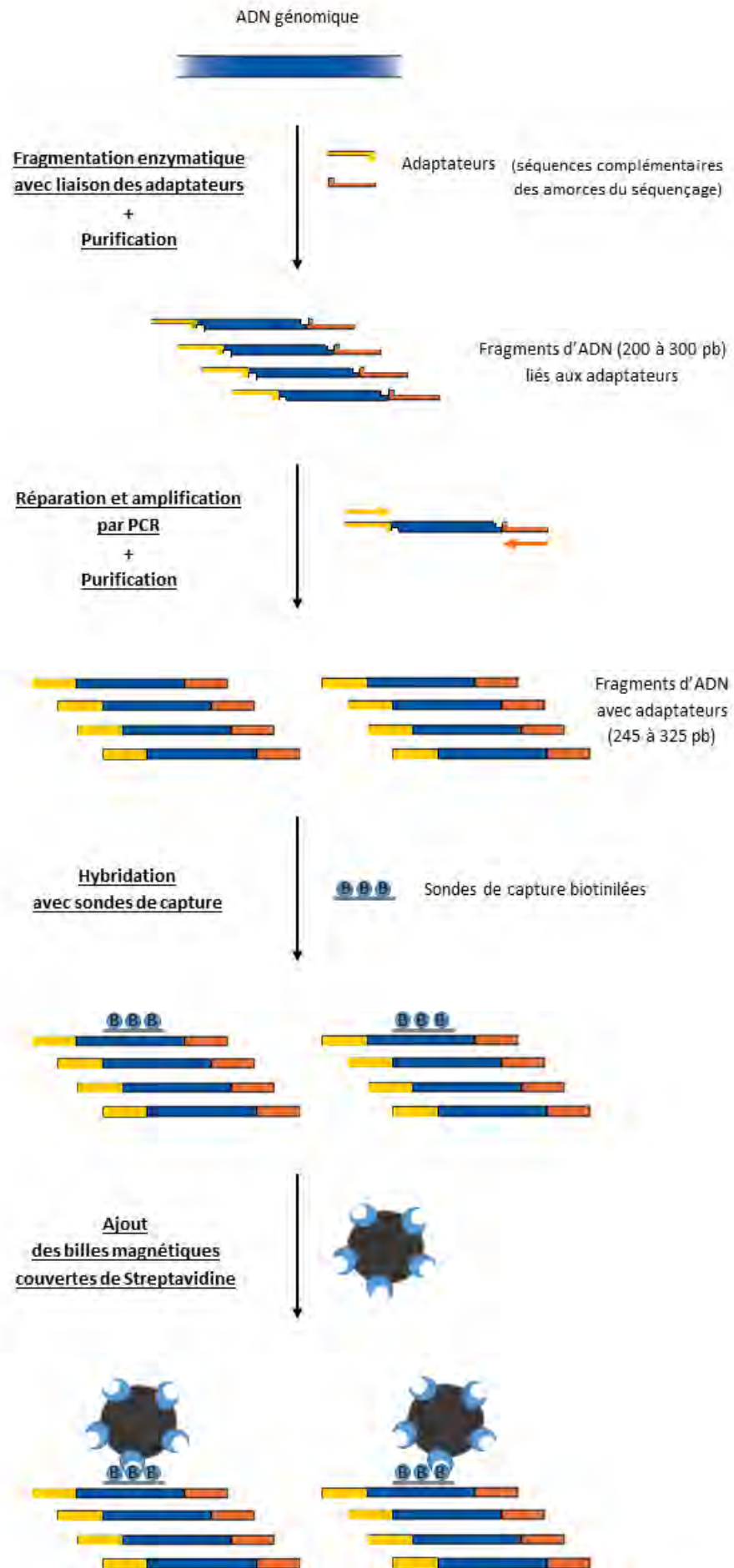
La deuxième étape consiste à « découper » à l'aide d'une enzyme l'ADN génomique en fragments d'environ 200 à 300 pb, en liant des séquences adaptatrices aux extrémités (**Figure 24**). Ultérieurement, au cours de l'étape de séquençage réalisée sur le MiSeq Illumina, ces séquences adaptatrices permettront l'hybridation des amorces utilisées. On purifie ensuite l'ADN lié aux adaptateurs sur billes magnétiques grâce au système Agencourt AMPure XP (*Beckman Coulter*). Puis on réalise une réparation et une amplification par PCR, suivie d'une purification sur billes AMPure XP (**Figure 24**). La qualité et la concentration des échantillons sont contrôlées sur puce DNA High sensitivity à l'aide du Bioanalyseur 2100 (*Agilent*), la taille des fragments d'ADN devant être comprise entre 245 et 325 pb.

La troisième étape consiste à hybrider les sondes d'ARN biotinylées, spécifiques de chaque région à capturer, avec les échantillons d'ADN fragmenté amplifié. Des billes magnétiques couvertes de Streptavidine permettent de récupérer les fragments d'ADN capturés (**Figure 24**). On purifie ensuite ces billes, et on amplifie l'ADN par PCR grâce à des amorces « barcodées ». Ces amorces « barcodées » contiennent des séquences Index* spécifiques à chaque patient et des séquences adaptatrices (**Figure 24**). Lors du

run sur le séquenceur MiSeq Illumina, ces séquences adaptatrices permettront la liaison des fragments d'ADN à séquencer aux amorces de la *flow cell* au cours de l'étape d'amplification clonale des librairies par *Bridge PCR* (PCR avec formation de ponts). On récupère les fragments d'ADN amplifiés et indexés, puis on les purifie sur billes AMPure XP. La qualité et la concentration des échantillons sont de nouveau contrôlées à l'aide du Bioanalyseur 2100 (*Agilent*), la taille des fragments d'ADN devant être comprise, cette fois-ci, entre 325 et 450 pb.

La quatrième étape consiste à quantifier de manière précise les librairies par dosage Qubit DsDNA HS Assay (*Thermo Fisher Scientific*), en suivant le protocole du fournisseur, à les diluer à 4 nM, et à les combiner de manière équimolaire. Puis le pool de librairies est dénaturé avec de la soude, et combiné à 1% de PhiX (*Illumina*), contrôle pour l'étape de séquençage (**Figure 25**).

Figure 24 : Préparation de la librairie avec la technologie SureSelectQXT d'Agilent



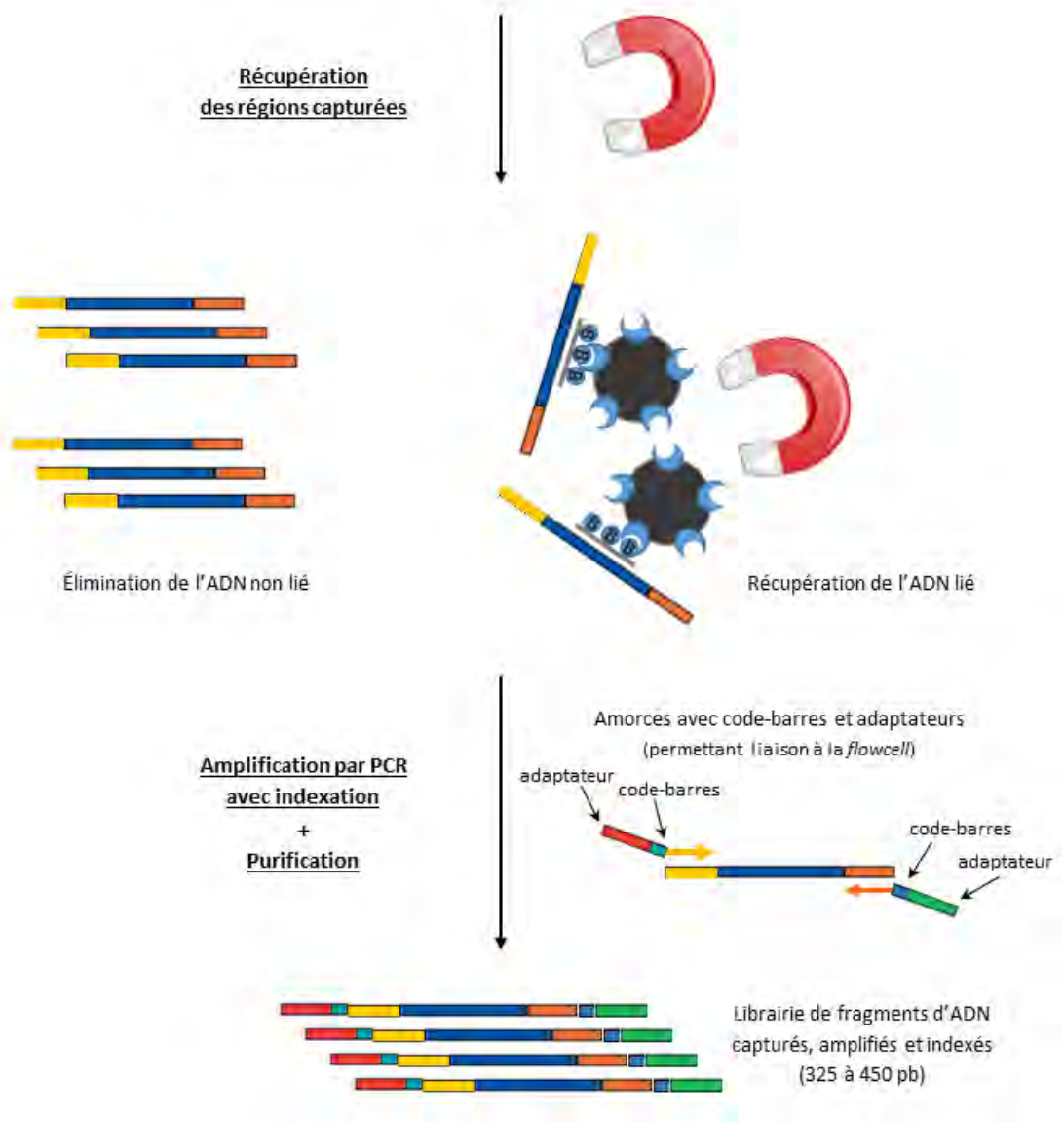
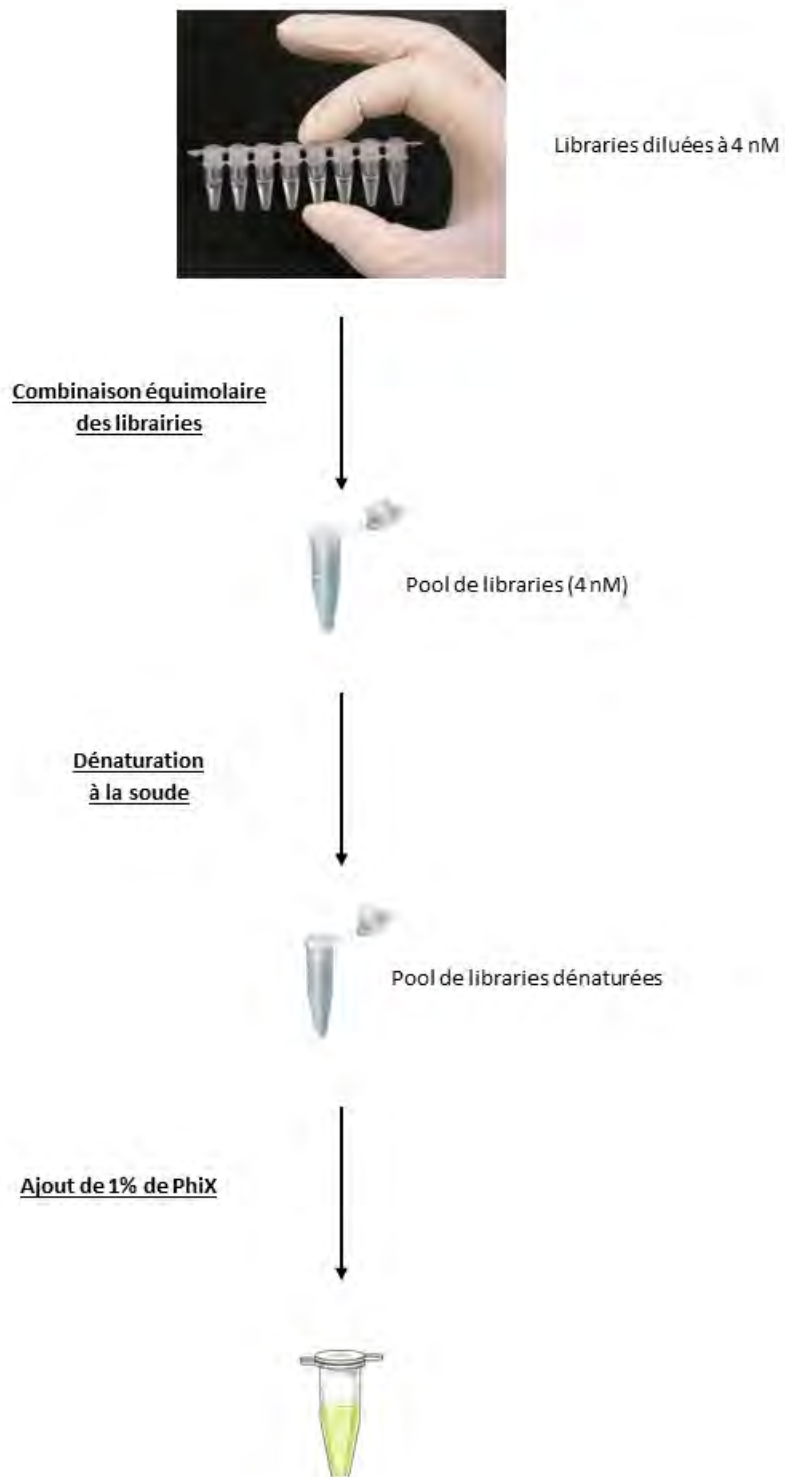


Figure 25 : Combinaison des librairies et préparation pour le séquençage



2. Séquençage par le MiSeq Illumina avec la génération de données par le logiciel associé au séquenceur

Le chargement des réactifs, de la *Flow cell*, et de la cartouche de réactifs (MiSeq Reagent Kit v2 (300-cycles), *Illumina*), ainsi que le lancement du *run* sur le séquenceur MiSeq Illumina (logiciels Real Time Analysis version 1.18.54 et MiSeq Reporter version 2.5.1.3) (**Figure 26**) sont réalisés selon les instructions du fournisseur.

L'analyse des images photographiées et l'algorithme de « *base-calling* » (conversion du signal en séquences de nucléotides - *reads* - associés chacun à un score de qualité) sont réalisés au fur et à mesure du *run* par le logiciel du séquenceur. On obtient ainsi un fichier binaire au format BCL* (Illumina CASAVA Base Call Format, extension .bcl).

Ensuite, un prétraitement bio-informatique par le logiciel *bcl2fastq* permet le démultiplexage associé à un *trimming* des *reads*, c'est-à-dire la réattribution de chaque *read* au bon échantillon en supprimant les séquences des adaptateurs. Un fichier FASTQ (extension .fastq) est ainsi généré pour chaque patient, il contient les séquences d'ADN avec le score de qualité attribué à chaque nucléotide. On obtient également un rapport de qualité FastQC.

Figure 26 : Photographie du séquenceur MiSeq Illumina

(www.illumina.com)



VII. Traitement bio-informatique à partir des fichiers FASTQ

Le pipeline bio-informatique utilisé au CHU de Toulouse est une adaptation locale du pipeline DEVA développé à l'IGR (Institut Gustave Roussy).

A partir des fichiers FASTQ, les *reads* sont alignés sur le génome humain de référence (version hg19/GRCh37) à l'aide de l'algorithme BWA-MEM du programme BWA v.0.7.10 (*Burrows-Wheeler Aligner*).

La suite d'outils bio-informatiques de GATK v.3.3 (*Genome Analysis Toolkit*) est utilisée pour repérer et reprendre les alignements locaux connus dans la littérature : *RealignerTargetCreator*, *IndelRealigner*, *BaseRecalibrator*, *PrintReads*.

Un fichier BAM (*Binary Alignment Map*, extension .bam) est alors généré et permet de visualiser les données sur le logiciel IGV v.2.3.95 (*Integrative Genome Viewer*)*.

Pour la détection des variants, 2 logiciels de recherche de variants (*Variant Caller**) sont utilisés : *HaplotypeCaller* de GATK v.3.3 et *VarScan 2* v.2.3.7. Deux fichiers VCF (*Variant Call Format*, extension .vcf) sont générés.

Chacun des deux fichiers VCF est annoté avec les outils d'annotation *SNPEff* v.4.3, *dbNSFP* v.3.5a, et la base de donnée *COSMIC* v.83 (*Catalogue Of Somatic Mutations In Cancer*).

Les fichiers VCF annotés sont ensuite importés dans le logiciel *BD_NGS* qui propose une synthèse de ces deux résultats.

Résultats

Préambule

La révolution biotechnologique du séquençage de nouvelle génération offre de nouvelles perspectives pour le diagnostic génotypique des affections ayant une forte hétérogénéité génétique. Parmi elles, la DI est probablement celle dont l'approche diagnostique a le plus profité de cette avancée technologique du NGS. Ainsi, de nouveaux algorithmes décisionnels dans l'exploration génétique des patients atteints d'une DI se sont imposés (cf. partie Introduction, paragraphe I.5, page 44).

Dans ce contexte, pour offrir aux cliniciens prescripteurs de notre CHU la possibilité d'améliorer la sensibilité diagnostique du bilan génétique de la DI, notre laboratoire s'est engagé à mettre en place un nouveau test ADN explorant par NGS un large panel de gènes impliqués dans la DI. Plusieurs étapes ont été nécessaires pour développer ce test, depuis le choix de la technologie utilisée jusqu'à la création des sondes de capture. Nous avons voulu comparer ce test avec un test déjà existant qui fait appel à une approche technologique différente, le test DI44. L'objectif souhaité étant de mettre en évidence un rendement diagnostique et une fiabilité des résultats supérieurs. Pour montrer l'intérêt médical de ce test comparatif, nous avons constitué une cohorte de patients témoins atteints de DI. Finalement, nous avons confronté les performances ce nouveau test à celles d'autres tests NGS de DI réalisés par d'autres laboratoires.

I. Développement d'une méthode de séquençage de nouvelle génération de 117 gènes pour le diagnostic de la déficience intellectuelle (DI)

Le développement d'un nouveau test diagnostique commence par le choix de la technologie utilisée. La sélection des gènes ciblés par le test est réalisée sur la base d'arguments médicaux, scientifiques, et technologiques. La dernière étape de développement du test consiste à générer les sondes de capture utilisées correspondant aux régions des gènes ciblées par le test.

1. Choix de la technologie utilisée

Pour la préparation des librairies et pour le séquençage NGS, nous avons opté pour une méthode d'enrichissement par capture* et pour la technologie de séquençage Illumina (cf. partie Introduction, paragraphes II.2.a et II.1.b.i.4, pages 76 et 67). Ce choix repose sur plusieurs arguments.

Concernant la mise au point de la librairie, des études ont montré que la technique d'enrichissement par capture est plus robuste que l'enrichissement par amplification* (60).

La première repose sur un phénomène physique d'hybridation des sondes, alors que la technique par amplification fait appel à un processus enzymatique d'amplification (PCR) qui peut être, en fonction des caractéristiques de certaines séquences cibles, inefficace ou source d'erreurs.

De plus, les sondes de capture sont environ 5 fois plus larges que les amorces utilisées dans le procédé d'enrichissement par amplification. Ainsi, une amorce de PCR peut ne pas s'hybrider en cas de variation nucléotidique conduisant à l'échec d'amplification du

fragment d'ADN concerné. Un variant peut ne pas être détecté en raison de ce biais. En revanche, les sondes de capture, du fait de leur plus grande taille, s'hybrideront même en cas de variation nucléotidique (mésappariement toléré).

De plus, la méthode d'enrichissement par capture est mieux adaptée pour la détection de variations du nombre de copie (CNV). En effet, l'amplification par PCR n'est pas toujours homogène entre les différents fragments d'ADN. Cela peut biaiser la mesure quantitative de leur fréquence allélique respective et donc la mise en évidence des CNV.

Enfin, la technologie d'enrichissement par capture s'avère techniquement plus souple dans son développement et offre davantage de possibilités de modification du panel de gènes. Lorsque l'on souhaite rajouter un gène à un panel développé avec la technologie d'enrichissement par amplification, il est nécessaire de reconcevoir toutes les amorces de la PCR multiplexe. En effet, il existe dans une PCR un phénomène de compétition entre les amorces lors de l'hybridation à l'ADN. Ce phénomène de compétition est basé sur leur niveau de spécificité, et un mélange minutieusement contrôlé des amorces est nécessaire pour le déroulement optimal d'une PCR multiplexe. Les amorces peuvent également s'hybrider entre elles si elles partagent un pourcentage élevé d'homologie. En revanche, avec une librairie par capture, il n'est pas nécessaire de modifier les sondes déjà réalisées lorsque l'on désire ajouter un gène à un panel.

En ce qui concerne la technologie utilisée pour le séquençage proprement dit, nous avons comme alternative soit de passer la librairie sur un séquenceur de type Ion Torrent™ soit sur un séquenceur de type Illumina. Le premier nécessite une étape d'amplification clonale en PCR émulsion puis une incorporation par synthèse de dNTP générant l'émission de protons (ionogramme), tandis que dans le procédé Illumina l'étape préalable d'amplification s'effectue sur un support solide appelé *flow cell*, suivie d'une incorporation par synthèse de dNTP fluorescent de type bloqueur réversible (cf. partie Introduction, paragraphes II.1.b.i.3 et II.1.b.i.4, pages 61 et 67). Bien que le temps de *run* soit plus long avec un séquençage sur un appareil Illumina, plusieurs arguments plaident pour ce choix technologique à des fins de diagnostic. Le premier est que le taux d'erreur de séquençage s'avère moins élevé qu'avec la technologie Ion Torrent (cf. **Figure 13**, page 70). Or, plus le nombre de variants faux-positifs* dus à ces erreurs de séquençage est

important, plus l'analyse des résultats est fastidieuse. Le variant causal doit être distingué des variants faux-positifs, son identification est simplifiée s'il existe moins de faux-positifs. En l'absence de variant causal évident, les variants faux-positifs devront parfois être infirmés avec un séquençage par la méthode de Sanger.

Le choix technologique d'enrichissement par capture et de séquençage sur séquenceur Illumina nous a donc paru le plus pertinent pour l'ensemble de ces raisons.

2. Sélection des gènes cibles

La conception de notre panel a été réalisée en étroite collaboration avec le Dr Amélie Piton du laboratoire de génétique moléculaire du CHU de Strasbourg. Le CHU de Strasbourg et le CHU de Toulouse appartiennent au réseau DéfiScience* (Filière Nationale de Santé pour les maladies rares du développement cérébral et pour la Déficience Intellectuelle). Le Dr Piton possède une expertise indéniable dans la DI avec de nombreuses publications à son actif dans ce domaine (61). Le CHU de Strasbourg a d'ailleurs contribué à l'élaboration du test DI44 que notre nouveau test vise à remplacer dans notre laboratoire. La sélection des gènes de notre panel a ainsi été construite au cours d'échanges fructueux avec le CHU de Strasbourg.

Le choix de la taille du panel et des gènes à inclure repose sur des arguments médicaux, scientifiques, et économiques. Nous souhaitons en effet développer un test génétique de DI de première intention. Par conséquent, sa sensibilité et son rendement diagnostique doivent être optimisés. Cependant, le coût de revient du test par patient doit aussi être acceptable par rapport à la rentabilité diagnostique : notre but est de trouver le meilleur compromis entre un nombre de gènes suffisant et bien sélectionnés pour une bonne sensibilité diagnostique, et une offre de test qui soit accessible au plus grand nombre de patients.

Nous avons ainsi décidé de développer un panel NGS de première intention en sélectionnant environ une centaine de gènes.

Nous avons conservé les 44 gènes ciblés dans le test DI44. Nous avons ajouté à ceux-ci les 70 gènes les plus fréquemment impliqués dans la déficience intellectuelle en France non-ciblés par le test DI44 (données du réseau DéfiScience). Cela donne un total de 114 gènes, auquel sont joints 3 gènes responsables du syndrome de Prader-Willi like, une maladie rare vis-à-vis de laquelle le CHU de Toulouse est le site coordonnateur du centre de référence en France. Comportant un total de 117 gènes ciblés en lien avec la Déficience Intellectuelle, notre test a ainsi été nommé DI117 (**Annexe 4**).

Sur la base des données communiquées par la filière nationale de santé DéfiScience concernant les patients ayant un variant pathogène diagnostiqué dans les gènes inclus dans notre panel, nous pouvons prédire un rendement diagnostique du test NGS DI117 à environ 20% (données recueillies auprès des laboratoires de Strasbourg, Lille, Rennes, La Pitié Salpêtrière, Lyon, et Dijon). Ce rendement est de 9,33% en se basant uniquement sur les données issues des publications scientifiques (Pubmed), et de 10,58% avec l'ensemble des données. De la même manière, le taux diagnostique du panel DI44 peut être évalué à 10,06% selon les données DéfiScience, à 4,30% selon les données Pubmed, et à 5,09% en agrégeant l'ensemble de ces données. Il apparaît ainsi que la mise en œuvre de ce nouveau test DI117 devrait doubler la détection de mutations par rapport au test DI44 pour les patients déficients intellectuels (**Annexe 5**).

3. Conception des sondes de capture

Pour concevoir les sondes de capture, nous avons utilisé le logiciel SureDesign (*Agilent*). Les variants pathogènes d'un gène sont presque exclusivement localisés dans les exons codants* ou dans les sites d'épissage* (62). Les sites d'épissage se situent dans les 6 premières bases des jonctions intron-exon. Nous avons donc choisi de cibler les séquences codantes (CDS) *, en ajoutant un minimum de 10 nucléotides de marge de part et d'autre. Ce choix permet l'inclusion des exons codants et des sites d'épissage en minimisant le risque d'un défaut de couverture.

Ainsi, nous avons conçu 54 078 sondes d'ARN différentes de 120 pb capables de capturer les fragments d'ADN destinés à être séquencés. La taille totale des régions ciblées s'élève à 762 464 pb.

II. Comparaison des performances du test DI117 (Illumina) avec le test DI44 (Ion Torrent)

Pour justifier la mise en place d'un nouveau test diagnostique dans un laboratoire à la place d'un autre existant pour la même indication, il convient de démontrer qu'il a des performances supérieures ou au moins équivalentes. Après avoir comparé les caractéristiques théoriques du test DI117 à celles du DI44, nous avons confrontés expérimentalement les deux tests, en évaluant notamment leur sensibilité et leur spécificité* sur leurs 44 gènes communs.

Comme détaillé dans le chapitre 1, la librairie du test DI117 est obtenue par capture, et son séquençage effectué sur un séquenceur Illumina de type MiSeq. La librairie du test DI44 est, quant à elle, obtenue par amplification et son séquençage réalisé sur un appareil Ion Torrent™ de type PGM™.

1. Performances théoriques des tests NGS DI44 versus DI117 pour les 44 gènes communs

a. Couverture attendue par le design du test

La couverture théorique est calculée à partir des amorces ou des sondes conçues dans le cadre du développement de chaque test. Elle est transmise par le fabricant des amorces ou des sondes.

i. Sur les régions codantes avec une marge de ± 25 pb

Les performances théoriques ont d'abord été évaluées concernant la couverture attendue sur les régions codantes des 44 gènes avec une marge de 25 pb (CDS \pm 25) (**Tableau 4**). La région codante ou CDS (*Coding DNA Sequence*) est la partie de l'ADN d'un

gène traduite en protéine. La CDS commence au niveau d'un codon (séquence de trois nucléotides) initiateur de la traduction et se termine au niveau d'un codon stop.

La CDS ± 25 permet de séquencer les exons codants, les régions d'épissages et d'ajouter potentiellement le séquençage du point de branchement d'épissage*, séquence qui peut aussi être la cible de mutations. La CDS ± 25 est le seuil de couverture adopté pour le rendu de résultat dans notre laboratoire. Toutefois, il a été constaté que le point de branchement peut se situer parfois hors de cette région, la couverture CDS ± 25 n'est alors pas suffisante pour le séquencer.

Tableau 4 : Couverture théorique des tests DI44 et DI117 sur les régions CDS ± 6 pb et CDS ± 25 pb

Gènes		Test DI44						Test DI117 (sur les 44 gènes)						Taux de couverture par gène (Régions codantes ± 25 pb vs ± 6 pb)				
		Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		DI44		DI117				
			n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène		n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6			
ANKRD11	NM_001256182.1	exon10 exon10 exon10	exon10 (c.7067_7075)	9		9		exon10 (c.1346_1379) exon10 (c.4379_4421)	34 43		34 43							
ARID1B	NM_020732.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
ATRX	NM_000489.4	exon9 intron12 exon15 intron17 intron20 intron24	intron17 (c.4809+12_4809+25) intron20 (c.5272+21_5272+25)	14 5				exon9 (c.2465_2482) intron12 (c.4121-25_4121-23) exon15 (c.4322_4386) intron24 (c.5787-25_5787-21)	18 3 65 5		18 65							
CASK	NM_003688.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
CDKL5	NM_003159.2	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
CTNNB1	NM_001904.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
CUL4B	NM_003588.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
DLG3	NM_021120.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
DYRK1A	NM_001396.3	exon12			0		0	exon12 (c.1836_1847)	12		12	12	100,00%	100,00%	99,58%	99,58%		
EP300	NM_001429.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
FOXG1	NM_005249.4	UTR 5' exon1 UTR 3'	exon1 (c.279_311) UTR 3' (c.*23_*25)	33 3		33		UTR 5' (c.-25_-15)	11									
FOXP1	NM_032682.5	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
GATAD2B	NM_020699.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
GRIA3	NM_000828.4	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
GRIN1	NM_007327.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
GRIN2A	NM_000833.4	intron10 intron12	intron12 (c.2357-25_2357-21)	5				intron10 (c.2008-25_2008-22)	4									
					5		0			4		0	99,87%	100,00%	99,92%	100,00%		

n. pb. n. c. = nombre de bases non couvertes

Gènes		Test DI44						Test DI117 (sur les 44 gènes)						Taux de couverture par gène (Régions codantes ± 25 pb vs ± 6 pb)			
		Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		DI44		DI117			
			n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène		n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6		
GRIN2B	NM_000834.3	intron9 intron11		0		0	intron9 (c.2011-25_2011-22) intron11 (c.2360-25_2360-10)	4 16				100,00%	100,00%	99,60%	100,00%		
IL1RAPL1	NM_014271.3	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
IQSEC2	NM_001111125.2	intron4 exon5 exon15		0		0	intron4 (c.1402-25_1402-12) exon5 (c.1950_1981) exon15 (c.3716_3731)	14 32 16		32 16	48	100,00%	100,00%	98,81%	99,08%		
KDM5C	NM_004187.3	intron23	intron23 (c.4039-25_4039-18)	8	8	0			0		0	99,85%	100,00%	100,00%	100,00%		
KMT2A	NM_001197104.1	exon1 intron7 intron31	exon1 (c.6_195) intron7 (c.4012+20_4012+25) intron31 (c.11146+20_11146+25)	190 6 6		190					0	98,53%	98,62%	100,00%	100,00%		
MECP2	NM_004992.3	/	/	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
MED13L	NM_015335.4	intron2 intron10				0	intron2 (c.311-25_311-22) intron10 (c.2013-25_2013-15)	4 11			0	100,00%	100,00%	99,82%	100,00%		
NAA10	NM_003491.3	/	/	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
PQBP1	NM_005710.2	/	/	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
PTCHD1	NM_173495.2	/	/	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
RAI1	NM_030665.3	exon3 exon3				0	exon3 (c.4896_4917) exon3 (c.843_856)	22 14		22 14	36	100,00%	100,00%	99,39%	99,39%		
SATB2	NM_001172509.1	UTR 5'	UTR 5' (c.-25_-11)	15	15	0			0		0	99,75%	100,00%	100,00%	100,00%		
SCN2A	NM_021007.2	exon13 intron18	intron18 (c.2920-25_2920-21)	5	5	0	exon13 (c.1523_1547)	25		25		99,81%	100,00%	99,66%	99,66%		
SCN8A	NM_014191.3	intron20	intron20 (c.3820-25_3820-20)	6	6	0	intron20 (c.3820-25_3820-22)	4	4		0	99,92%	100,00%	99,94%	100,00%		
SETBP1	NM_015559.2	/	/	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%		
SHANK3	NM_033517.1	UTR 5' UTR 5' exon 1 intron1 intron1 intron2 intron10	UTR 5' (c.-25_-1) exon 1 (c.1_63) intron1 (c.63+1_63+25) intron1 (c.64-25_64-19) intron2 (c.267+19_267+25) intron10 (c.1303-25_1303-10)	25 63 25 7 7 16		6 63 6	UTR 5' (c.*15_*25)	11				97,73%	98,81%	99,83%	100,00%		

n. pb. n. c. = nombre de bases non couvertes

Gènes	Test DI44							Test DI117 (sur les 44 gènes)							Taux de couverture par gène (Régions codantes ± 25 pb vs ± 6 pb)			
	Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		Régions non couvertes en théorie	CDS ± 25 pb		CDS ± 6 pb		DI44		DI117					
		n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène		n. pb n. c.	n. pb n. c. total/gène	n. pb n. c.	n. pb n. c. total/gène	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6				
SLC16A2 NM_006517.4	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%			
SLC2A1 NM_006516.2	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%			
SLC6A8 NM_005629.3	exon1 intron1	exon1 (c.23_32)	10		10		intron1 (c.262+10_262+25)	16			0	99,49%	99,49%	99,37%	100,00%			
SLC9A6 NM_001042537.1	intron14	intron14 (c.1728-25_1728-16)	10			0		0		0		99,61%	100,00%	100,00%	100,00%			
SMARCA2 NM_003070.4	exon4 intron4	intron4 (c.791-25_791-17)	9				exon4 (c.647_741)	95			95							
	exon32 intron32	intron32 (c.4594+13_4594+25)	13			0	exon32 (c.4541_4561)	21			21							
			22			0		116			116	99,66%	100,00%	98,19%	98,19%			
SMC1A NM_006306.3	intron5	intron5 (c.855-25_855-24)	2			0		0			0	99,97%	100,00%	100,00%	100,00%			
STXBP1 NM_003165.3	intron3	intron3 (c.169+23_169+25)	2			0		0			0	99,96%	100,00%	100,00%	100,00%			
SYNGAP1 NM_006772.2	intron10 intron11						intron10 (c.1677-25_1677-22) intron11 (c.1913+22_1913+25)	4 4										
	exon15		0			0	exon15 (c.2869_2886)	18			18	100,00%	100,00%	99,48%	99,64%			
								26			18							
TBR1 NM_006593.3	UTR 3'					0	UTR 3' (c.*18_*25)	8			0	100,00%	100,00%	99,66%	100,00%			
TCF4 NM_001083962.1	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%			
UPF3B NM_080632.2	exon7			0		0	exon7 (c.652_729)	78			78	100,00%	100,00%	96,10%	96,10%			
WDR45 NM_007075.3	/	/	0	0	0	0	/	0	0	0	0	100,00%	100,00%	100,00%	100,00%			
Au total pour les 44 gènes			511		317			612		493		99,74%	99,84%	99,68%	99,75%			

n. pb. n. c. = nombre de bases non couvertes

On remarque que les 2 tests ont une couverture théorique moyenne des CDS ± 25 proche de 100% sur pour les 44 gènes étudiés (99,68% pour le test DI117, et 99,74% pour le test DI44, **Tableau 4**).

La majorité des 44 gènes sont en moyenne théoriquement couverts à plus de 99% par l'un ou l'autre des tests, mais il existe des différences selon les gènes. Ainsi, 27 gènes sont couverts à 100% par le test DI117, 14 entre 99 et 100%, et 3 en deçà de 99% (*IQSEC2* à 98,81%, *SMARCA2* à 98,19%, et *UPF3B* à 96,10%). Le test DI44 en compte respectivement 29, 12, et 3 (*FOXG1* à 97,63%, *KMT2A* à 98,53%, et *SHANK3* à 97,73%) (**Figure 27, Tableau 4**).

Ces données démontrent que la couverture moyenne théorique sur CDS ± 25 est sensiblement identique entre les 2 tests.

On note que certains gènes sont mieux couverts sur CDS ± 25 par l'un ou l'autre des tests. Ainsi, 10 des 44 gènes (*FOXG1*, *GRIN2A*, *KDM5C*, *KMT2A*, *SATB2*, *SCN8A*, *SHANK3*, *SLC9A6*, *SMC1A*, et *STXBP1*) ont une meilleure couverture théorique avec le test DI117, et 13 (*ANKRD11*, *ATRX*, *DYRK1A*, *GRIN2B*, *IQSEC2*, *MED13L*, *RAI1*, *SCN2A*, *SLC6A8*, *SMARCA2*, *SYNGAP1*, *TBR1*, et *UPF3B*) avec le test DI44 (**Figure 28**). Les 21 autres gènes ont une couverture théorique de 100% avec les 2 tests (**Figure 27, Tableau 4**).

Il est donc possible d'affirmer que les deux tests ont une couverture théorique sensiblement équivalente et a priori satisfaisante pour les 44 gènes étudiés.

Figure 27 : Graphique représentant la couverture théorique des tests DI44 et DI117 pour les 44 gènes (CDS ± 25 pb)

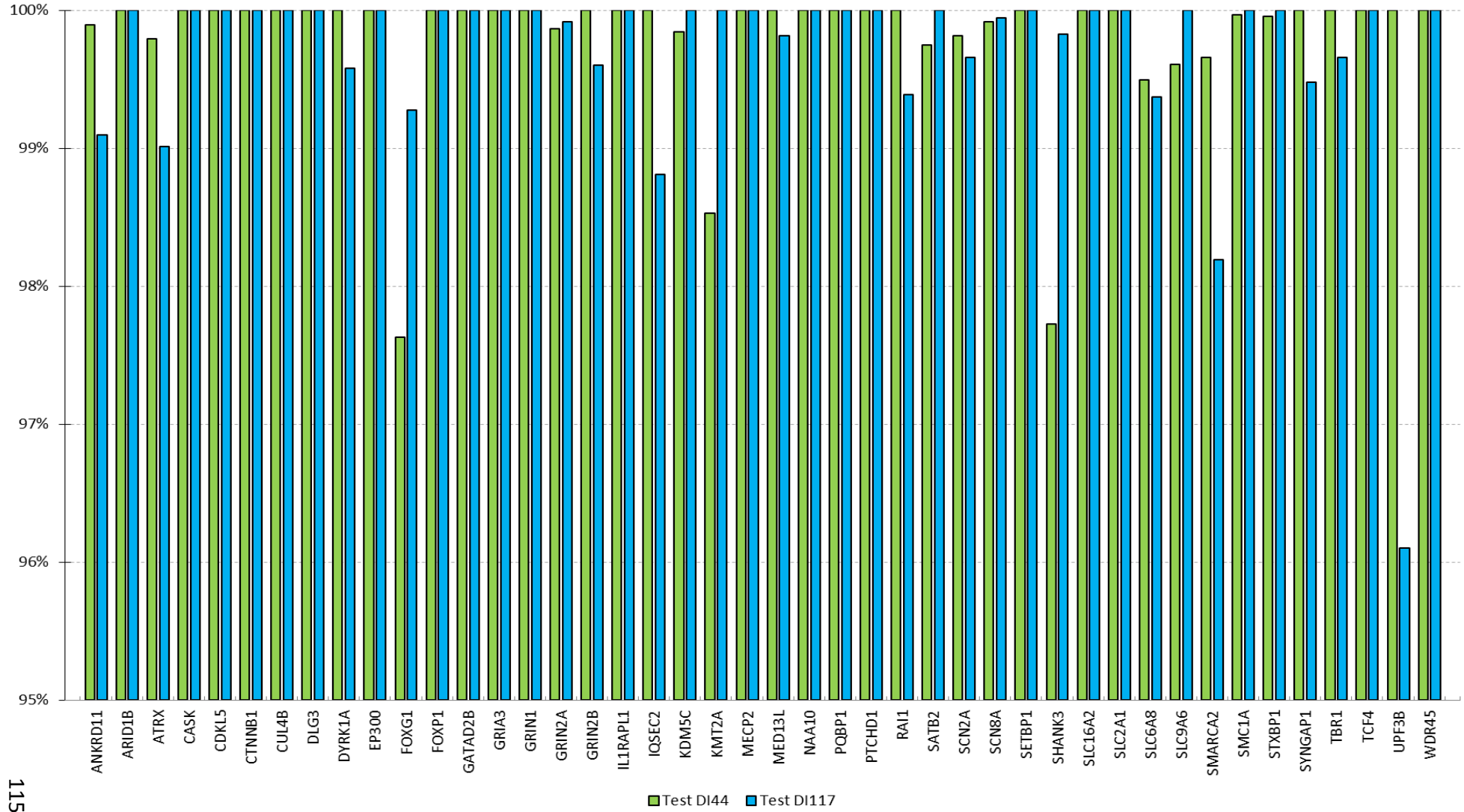
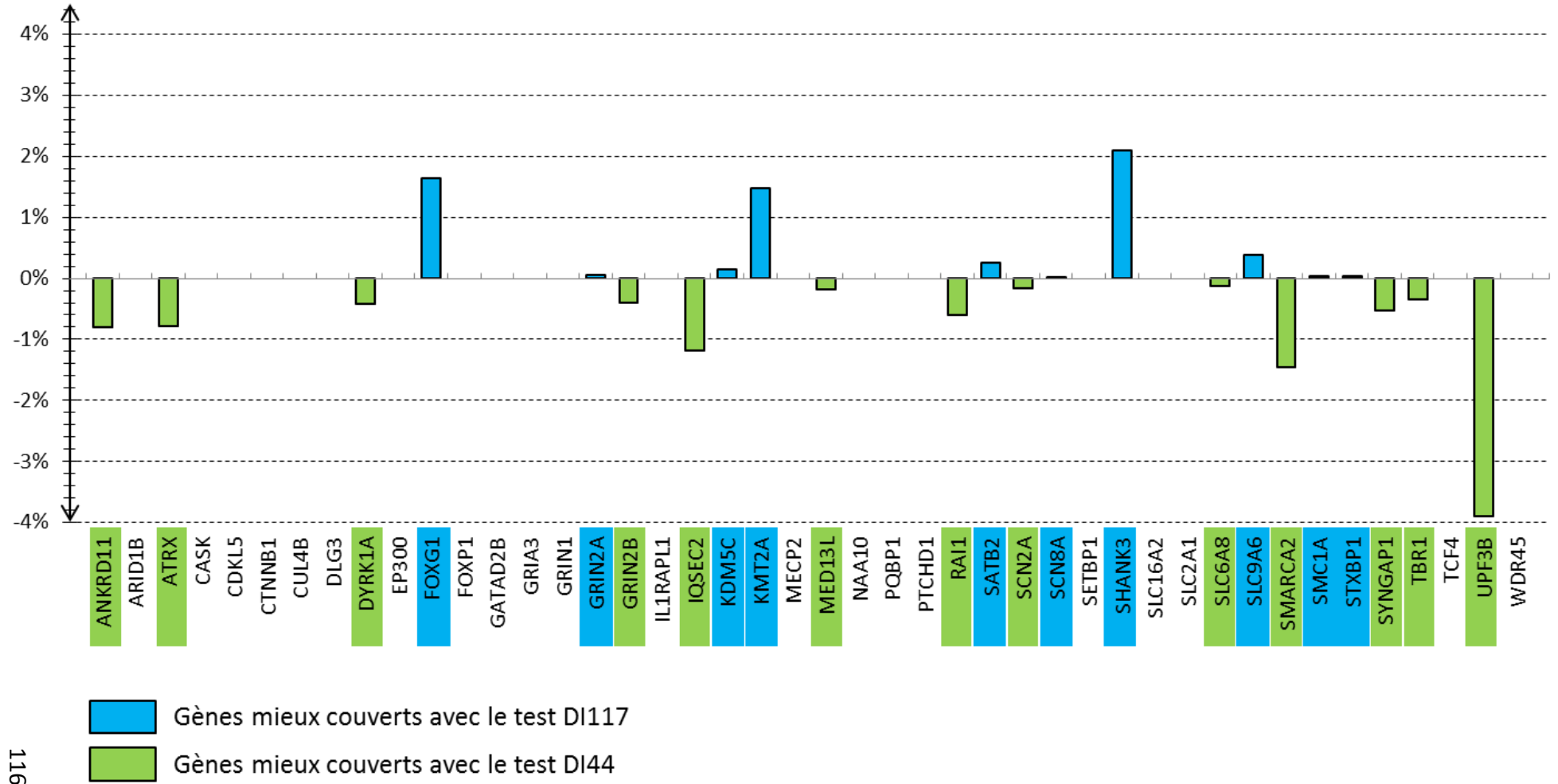


Figure 28 : Graphique représentant la différence (en %) entre les couvertures théoriques pour les 44 gènes (CDS ± 25 pb) : Test DI117 comparé au Test DI44



ii. Sur les régions codantes avec une marge de 6 pb

Il nous a semblé ensuite utile d'observer la couverture théorique pour les régions codantes des 44 gènes avec une marge de 6 pb (CDS ± 6) (**Tableau 4**). La CDS ± 6 est en effet la couverture minimale critique, incluant uniquement les exons codants et les régions d'épissage. Plus encore que la CDS ± 25 , Il est pertinent d'assurer une couverture complète pour ces régions, dans lesquelles se concentre la quasi-totalité des variants pathogènes.

L'objectif de couverture recherché sur ces régions CDS ± 6 est une couverture à 100%. En cas de couverture inférieure à 100%, un complément par séquençage Sanger serait nécessaire.

De la même manière que sur CDS ± 25 , les 2 tests ont une excellente couverture théorique sur CDS ± 6 , avoisinant les 100% pour les 44 gènes étudiés. Elle atteint 99,75% pour le test DI117, et 99,84% pour le test DI44 (**Tableau 4**).

La quasi-totalité des 44 gènes sont couverts à 100% par l'un ou l'autre des tests. Dans le test DI117, on dénombre 35 gènes à 100%, 7 entre 99 et 100% (*SCN2A* à 99,66%, *SYNGAP1* à 99,64%, *DYRK1A* à 99,58%, *RAI1* à 99,39%, *ANKRD11* et *ATRX* à 99,10%, et *IQSEC2* à 99,08%), et 2 en deçà de 99% (*SMARCA2* à 98,19%, et *UPF3B* à 96,10%) ; dans le test DI44, on en recense respectivement 39, 2 (*ANKRD11* à 99,89%, et *SLC6A8* à 99,49%), et 3 (*SHANK3* à 98,81%, *KMT2A* à 98,62%, et *FOXG1* à 97,83%)(**Figure 29, Tableau 4**).

Comme pour CDS ± 25 , la couverture théorique sur CDS ± 6 de certains gènes diffère entre les 2 tests. 4 gènes (*FOXG1*, *KMT2A*, *SHANK3*, *SLC6A8*) sont ainsi mieux couverts avec le test DI117, et 9 gènes (*ANKRD11*, *ATRX*, *DYRK1A*, *IQSEC2*, *RAI1*, *SCN2A*, *SMARCA2*, *SYNGAP1*, et *UPF3B*) avec le test DI44 (**Figure 30**). Les 31 autres gènes ont une couverture théorique de 100% sur CDS ± 6 avec les 2 tests (**Tableau 4**).

A l'aune de ces éléments, les deux tests DI44 et DI117 sont encore sensiblement équivalents dans leur couverture théorique pour leurs 44 gènes en commun, avec un très léger avantage pour le test DI44.

Figure 29 : Graphique représentant la couverture théorique des tests DI44 et DI117 pour les 44 gènes (CDS ± 6 pb)

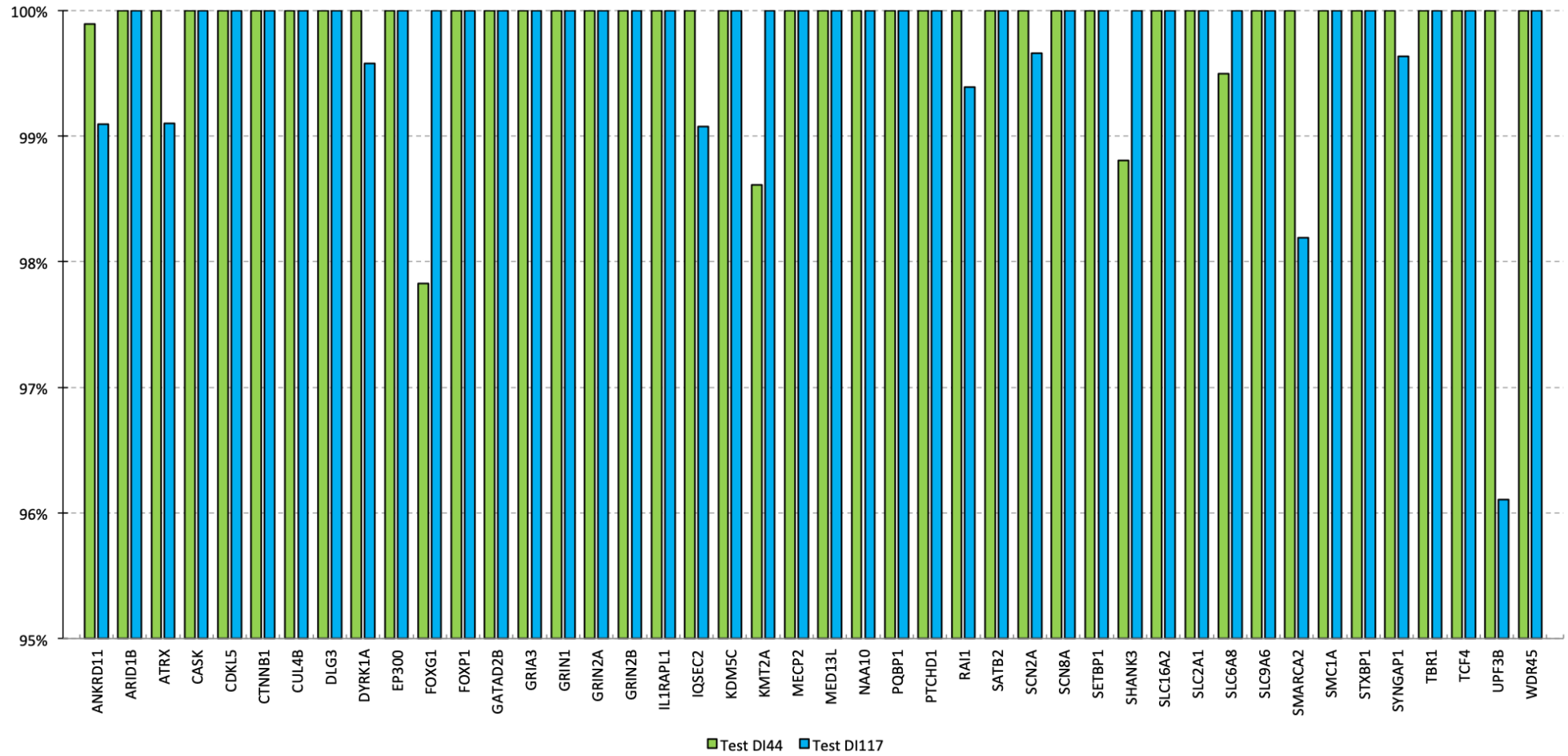
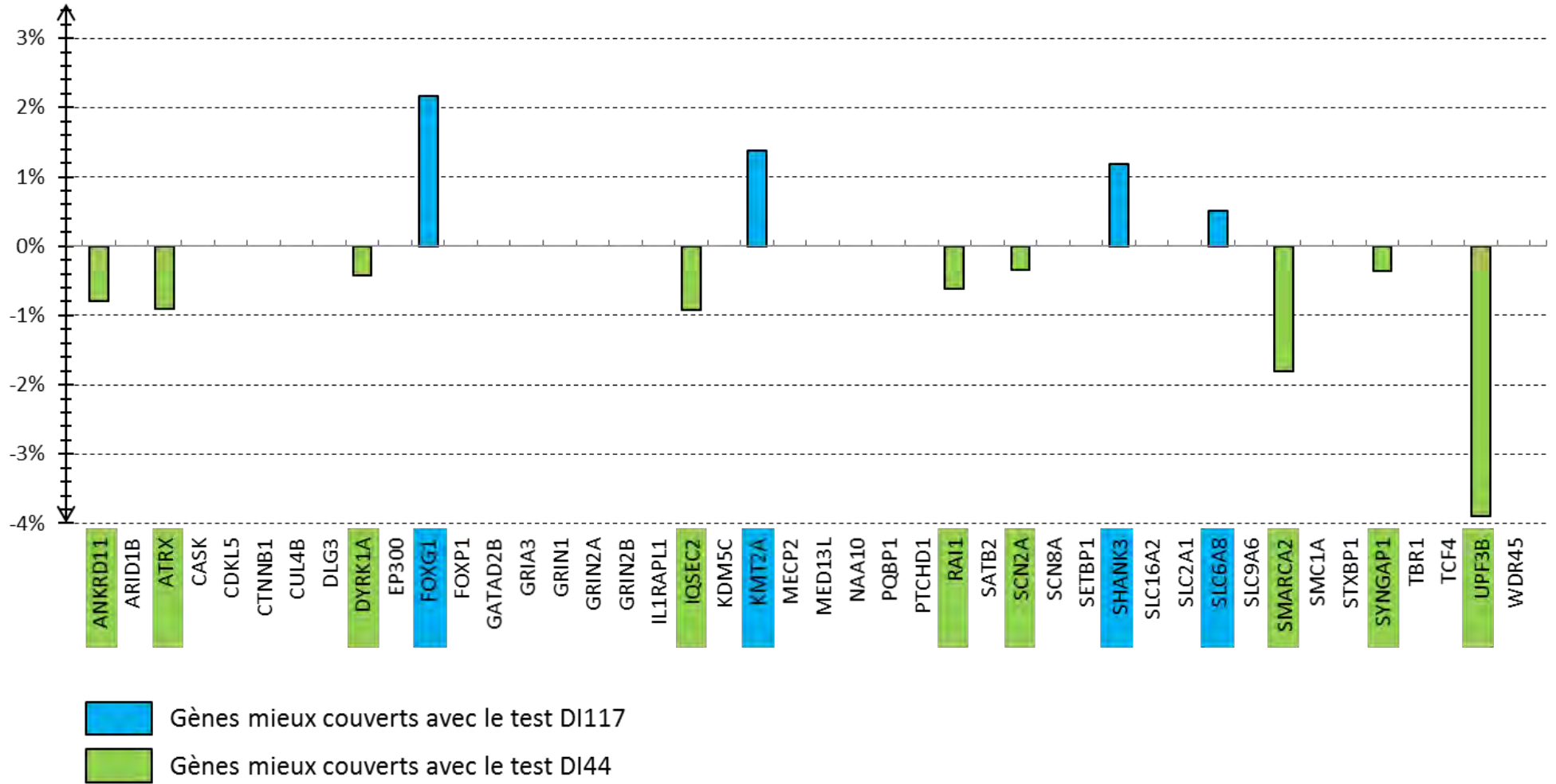


Figure 30 : Graphique représentant la différence (en %) entre les couvertures théoriques pour les 44 gènes (CDS ± 6 pb) : Test DI117 comparé au Test DI44



b. Nombre attendu de « Sanger » (amplicons séquencés par la technique de Sanger) pour compléter la couverture sur CDS ± 6

Nous avons pu constater dans le paragraphe précédent que la couverture sur les régions CDS ± 6 n'était pas systématiquement de 100% pour les deux tests DI117 et DI44. Or, pour détecter la totalité des mutations des régions codantes et des sites canoniques d'épissage des gènes cibles, l'objectif de couverture à atteindre est de 100%.

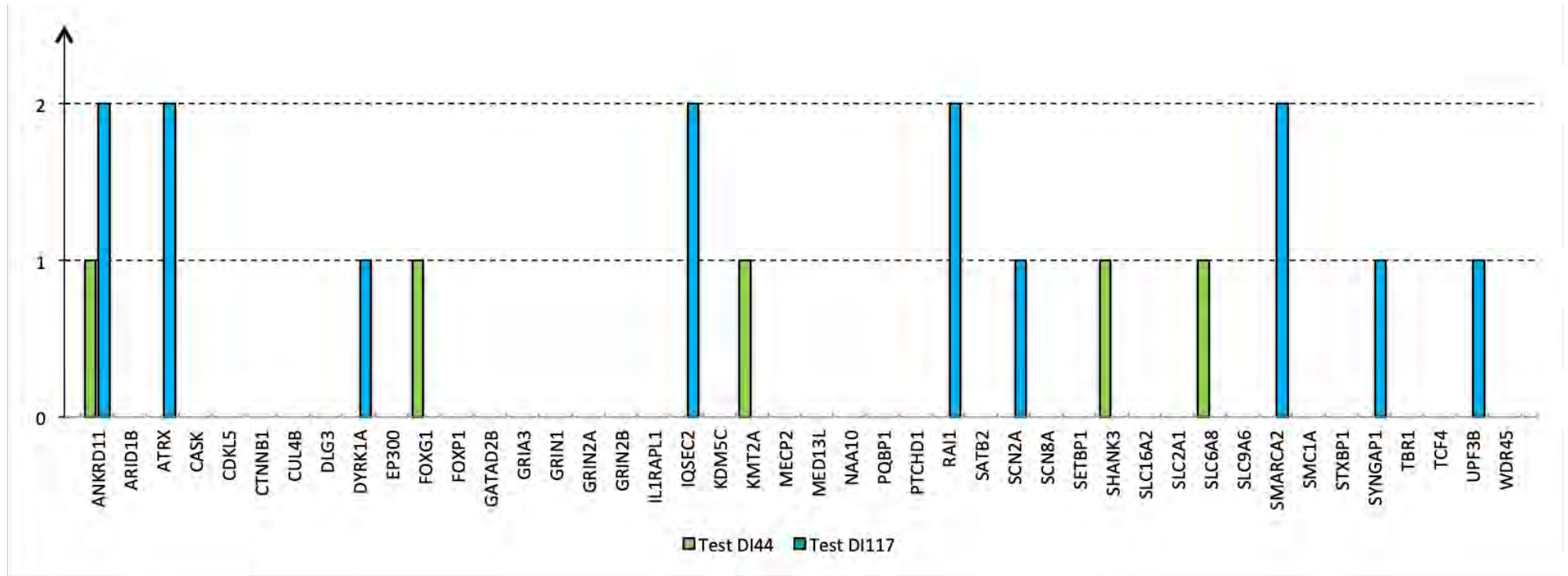
En cas de couverture inférieure par un test, un séquençage complémentaire par la technique de Sanger est à envisager pour étudier la région non couverte. Le séquençage par la méthode de Sanger est chronophage, il retarde le rendu du résultat au patient et il ajoute un coût supplémentaire au test initial. Ces éléments sont donc à prendre en considération dans l'évaluation globale d'un test à visée diagnostique.

Nous avons calculé pour les 44 gènes le nombre de « Sanger » (amplicons séquencés par la technique de Sanger) complémentaires nécessaires pour atteindre une couverture théorique de 100% sur CDS ± 6 pour chacun des deux tests (**Annexe 6**).

En théorie, 14 « Sanger » complémentaires sont nécessaires pour le test DI117, contre 5 pour le test DI44 (**Figure 31, Annexe 6**).

Il n'y a pas de différence significative concernant les performances théoriques entre les 2 tests. Néanmoins, la couverture théorique sur les régions CDS ± 6 est légèrement inférieure pour le test DI117 par rapport au test DI44 sur leurs 44 gènes communs ; concrètement cela signifie que 9 « Sanger » devraient être effectués en plus par rapport au test DI44. Ce résultat pourrait apparaître comme dommageable pour la mise en production de ce nouveau test DI117. Toutefois, avant de conclure sur d'éventuels défauts de couverture du test DI117 par rapport au test DI44 pour ces 44 gènes, il est nécessaire d'évaluer leurs performances réelles. Celles-ci peuvent différer des valeurs théoriques et elles sont à considérer avec encore plus d'égards, car elles sont issues de la réalité expérimentale.

Figure 31 : Graphique représentant le nombre de "Sanger" nécessaire(s) pour compléter la couverture théorique des 44 gènes (sur CDS ± 6) : Test DI44 versus Test DI117



2. Performances réelles déduites de la pratique expérimentale

Nous avons évalué trois caractéristiques qui nous semblaient pertinentes en pratique expérimentale :

- la couverture sur CDS ± 25 et sur CDS ± 6 ,
- la profondeur de lecture,
- et la performance diagnostique (capacité à détecter les vrais variants candidats).

Les données ont été obtenues avec 6 patients communs aux deux tests, auxquels sont ajoutés 9 patients supplémentaires pour le test DI117. Un patient du test DI44 avait un ADN de mauvaise qualité et a donc été exclu des analyses. Ce patient n'a pas été séquencé avec le test DI117 et n'apparaît pas dans les résultats.

a. Couverture observée expérimentalement

La couverture a été évaluée expérimentalement pour les deux tests et pour les 44 gènes communs en utilisant les mêmes critères que pour la couverture théorique : sur CDS ± 25 , sur CDS ± 6 et le nombre de « Sanger » complémentaires attendus.

i. Sur les régions codantes avec une marge de 25 pb

Dans un premier temps, nous avons observé de façon globale le pourcentage moyen de couverture des régions CDS ± 25 des 44 gènes à une profondeur soit de 30X soit de 100X.

Le pourcentage moyen de couverture* à une profondeur de 30X ou 100X correspond au nombre de bases séquencées au moins 30 ou 100 fois par rapport au nombre total de bases ciblées.

Ces deux profondeurs de lecture ont été étudiées en se basant sur les recommandations de l'ANPGM (Association Nationale des Praticiens de Génétique Moléculaire), qui préconise une profondeur minimale de 30X pour l'analyse des variants SNV, et une profondeur minimale de 100X pour l'analyse des variants CNV.

Comme on peut le constater sur les **Figure 32**, **Figure 33** et **Figure 34**, la différence est significative entre les résultats du *run* de séquençage du test DI44 réalisé sur le séquenceur PGM™ Ion Torrent™ à partir d'une librairie obtenue par amplification, et ceux des 2 *runs* de séquençage du test DI117 réalisés sur le séquenceur MiSeq Illumina à partir de bibliothèques obtenues par capture. Ainsi, pour ce dernier les résultats montrent une couverture moyenne des 44 gènes cibles de plus de 98% à 30X et supérieure à 94% à 100X (**Figure 33**, **Figure 34**), contre, respectivement, environ 90% à 30X et 84-88% à 100X pour le test DI44 (**Figure 32**).

Figure 32 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le *run* de séquençage du Test DI44 réalisé sur le séquenceur PGM™ Ion Torrent™

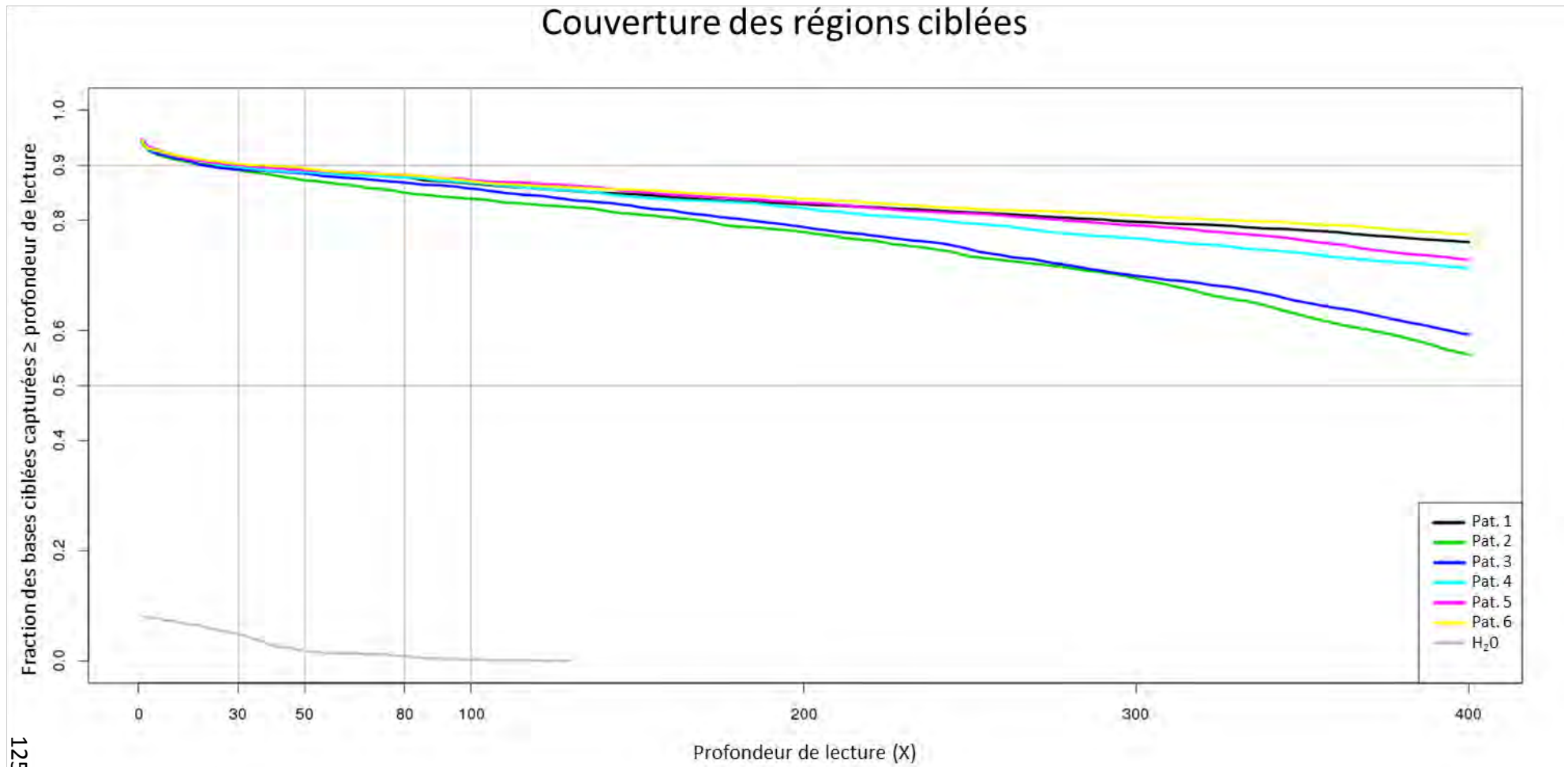


Figure 33 : Graphique représentant la fraction couverte des bases ciblées (avec la région cible limitée aux 44 gènes communs avec le Test DI44) en fonction de la profondeur de lecture pour le 1er *run* de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina

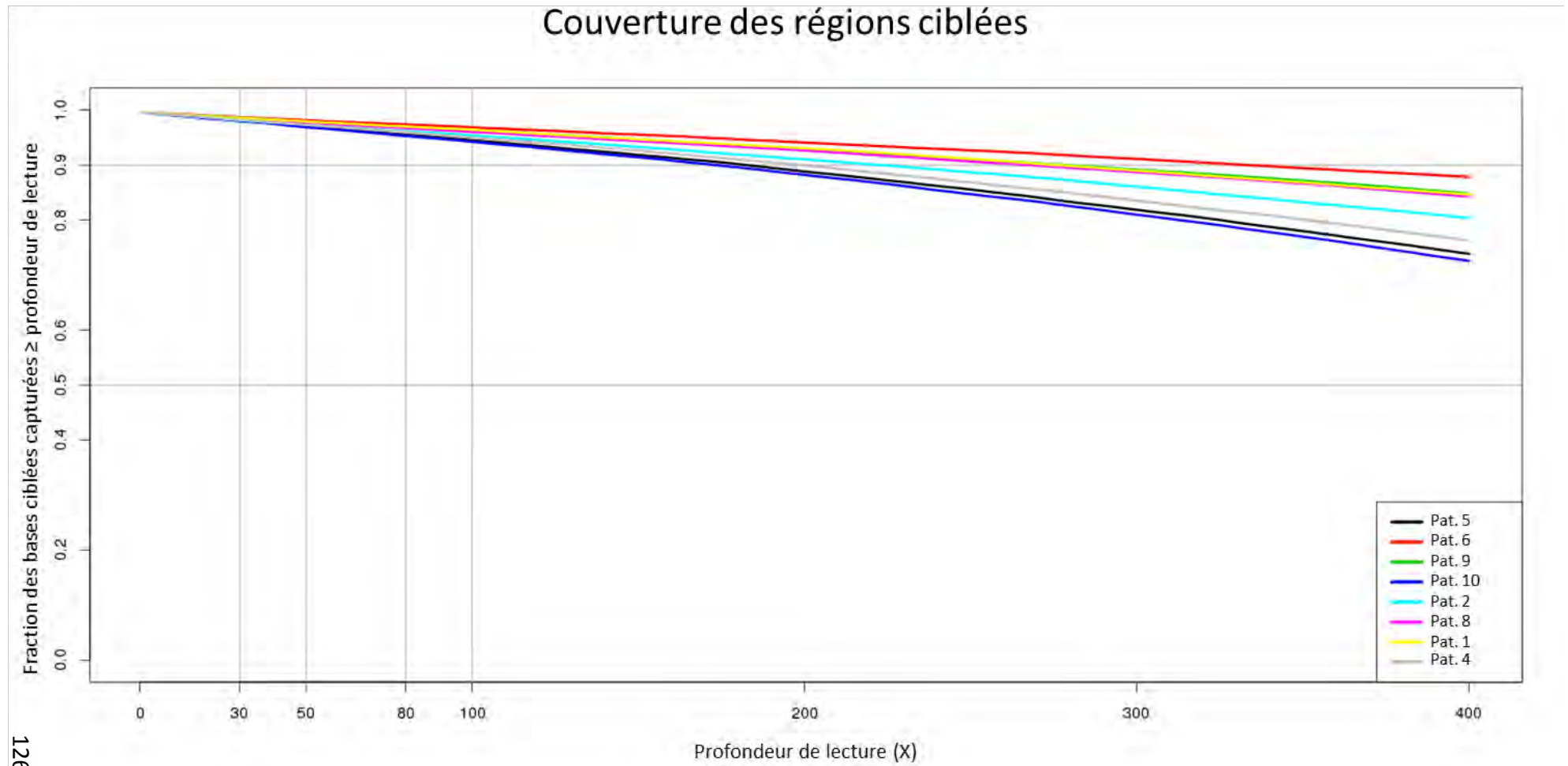
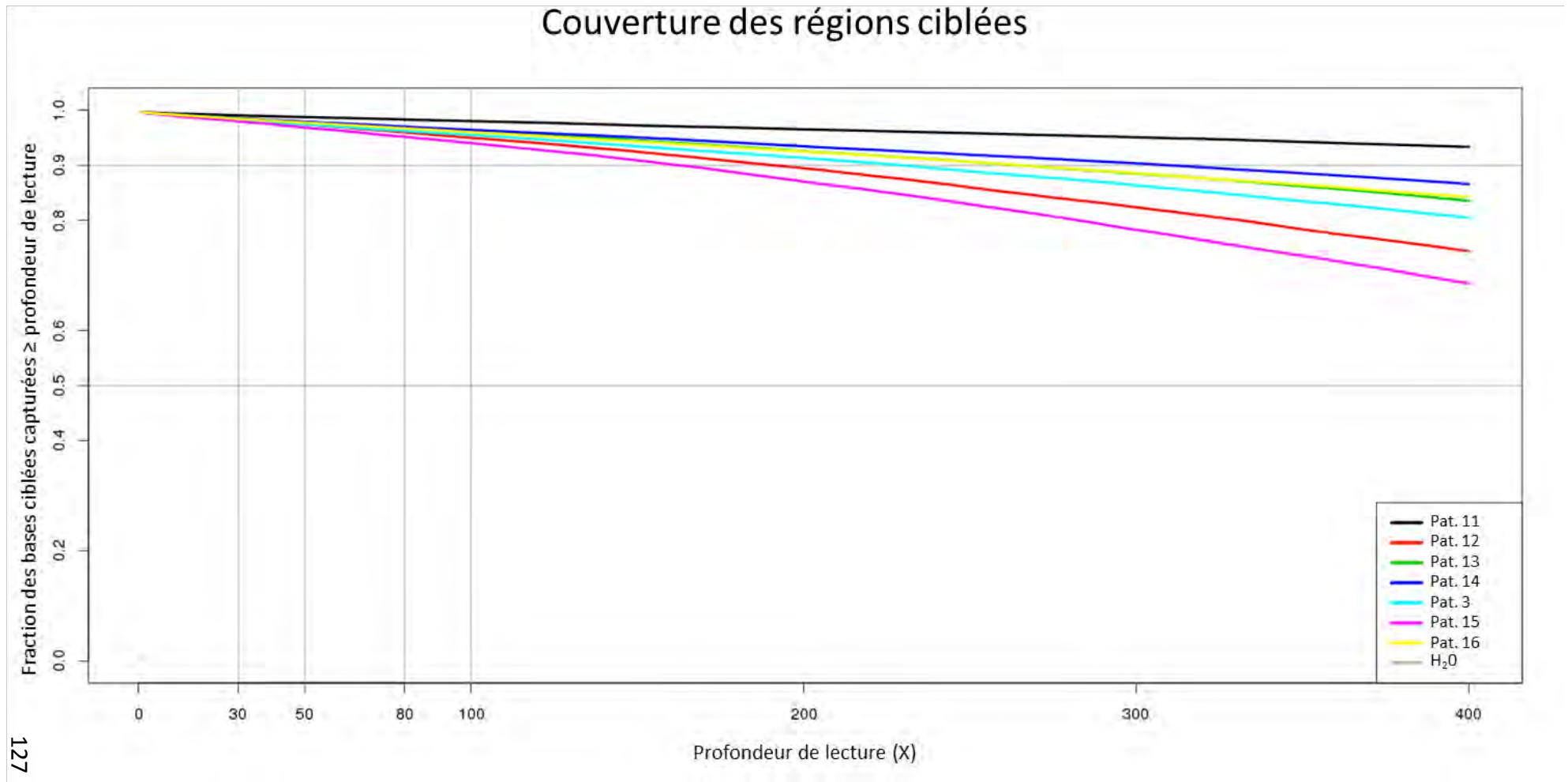


Figure 34 : Graphique représentant la fraction couverte des bases ciblées (avec la région cible limitée aux 44 gènes communs avec le Test DI44) en fonction de la profondeur de lecture pour le 2^{ème} run de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina



Nous nous sommes ensuite intéressés aux régions ayant une profondeur de lecture inférieure à 30X (*gap 30X*) et de fait non considérées comme couvertes en diagnostic. Nous les avons répertoriées pour chacun des 44 gènes et pour chacun des patients analysés (6 patients pour le test DI44 et 15 pour le test DI117). Ces régions non couvertes qui sont considérées comme des *gaps** incluent par définition les régions non couvertes par le dessin de la librairie (cf. paragraphe II.1.a.i, page 109).

Les *gaps 30X* liés aux conditions expérimentales peuvent varier d'un patient à l'autre, notamment en fonction de la qualité de l'ADN ou de la charge (quantité d'ADN) du patient au sein du *run*. Lorsque de telles variations ont été constatées, nous avons considéré le plus petit *gap 30X* (« gap min ») et le plus grand *gap 30X* (« gap max ») au sein de la série de patients. Ceci nous a conduit à définir en pratique une « fourchette » de couverture expérimentale comprise entre une couverture minimale (ensemble des « gaps max » à 30X) et une couverture maximale (ensemble des « gaps min » à 30X) (**Annexe 7, Annexe 8**).

Comme le montre la **Figure 35**, la différence de couverture en pratique sur CDS ± 25 entre les 2 tests est clairement à l'avantage du test DI117.

Ainsi, le test DI117 a 43 gènes couverts à plus de 90% (seul *FOXP1* a une couverture inférieure, égale à 85,46%) dont 38 à plus de 99% ; alors que pour le test DI44 c'est seulement 37 gènes qui sont couverts à plus de 90%, dont 16 à plus de 99% (**Figure 35, Annexe 8**).

Comparativement, il apparaît que sur les 44 gènes, 38 sont mieux couverts par le test DI117 contre 2 par le test DI44 (**Figure 36**). Les 4 autres ont une couverture équivalente entre les 2 tests, égale à 100 % (**Annexe 7**).

Globalement, la couverture minimale calculée pour l'ensemble des 44 gènes est de 99,24% pour le test DI117 contre 92,84% pour le test DI44 (**Annexe 8**).

Une différence aussi significative est observée lorsque l'on considère la couverture expérimentale maximale sur CDS ± 25 , c'est-à-dire prenant en compte les « *gaps min* » (Annexe 7, Annexe 8).

Figure 35 : Graphique représentant la couverture minimale en pratique des tests DI44 et DI117 pour les 44 gènes (CDS ± 25 pb)

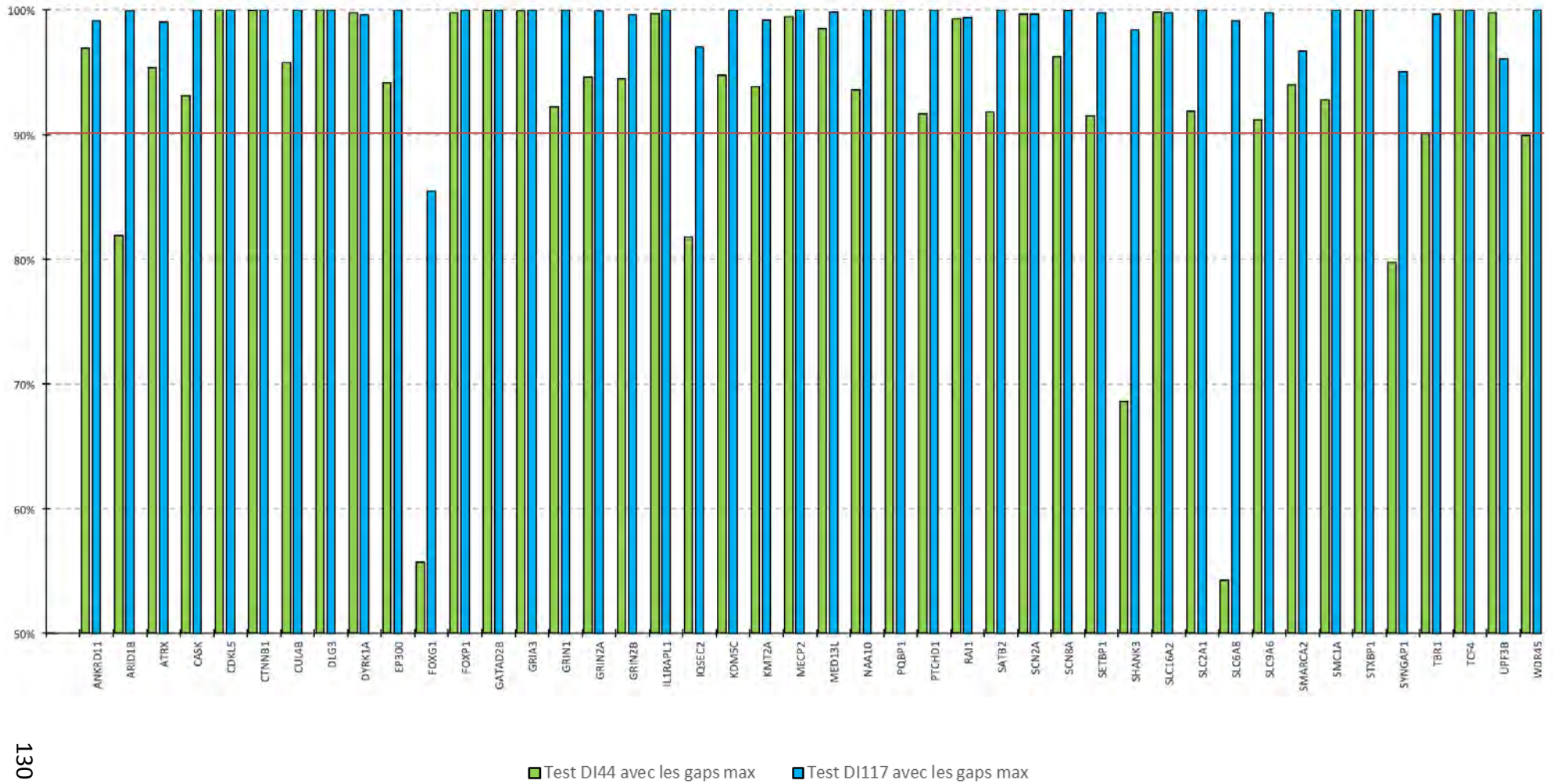
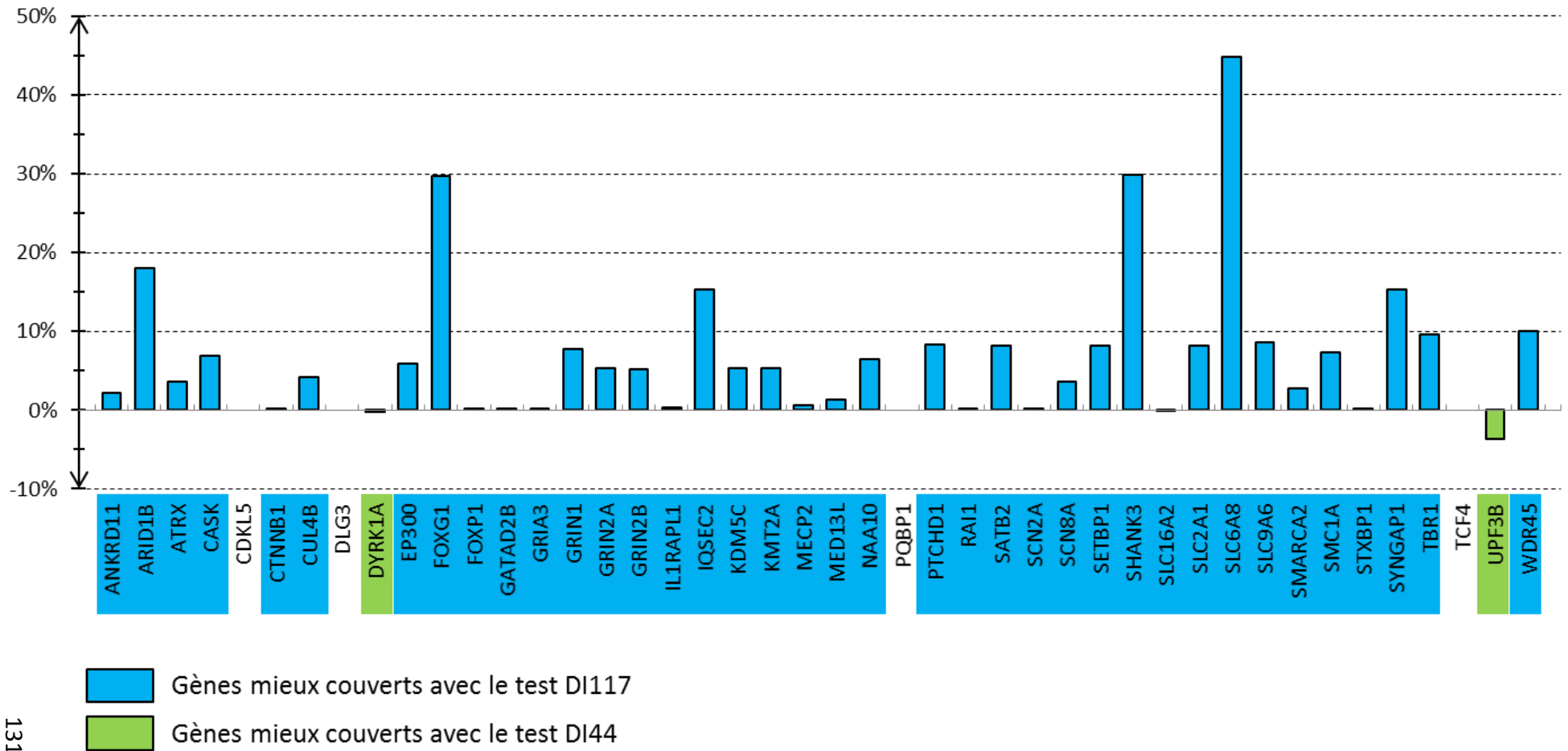


Figure 36 : Graphique représentant la différence (en %) entre les couvertures minimales en pratique pour les 44 gènes (CDS ± 25 pb) : Test DI117 comparé au Test DI44



Enfin, nous avons analysé les taux de couverture supérieure à 30X en les comparant gène par gène, patient par patient.

Pour l'ensemble des 6 patients, 31 gènes s'avèrent couverts à 100% avec le test DI117, contre seulement 5 gènes avec le test DI44 (**Tableau 5, Tableau 6**).

Pour le test DI117, 14 gènes sont mieux couverts expérimentalement qu'avec la couverture théorique déduite du design (*ANKRD11, ATRX, DYRK1A, GRIN2A, GRIN2B, IQSEQ2, MED13L, RAI1, SCN2A, SCN8A, SLC6A8, SMARCA2, TBR1, UPF3B*), 6 sont moins bien couverts (*ARID1B, FOXP1, KMT2A, SHANK3, SLC9A6, SYNGAP1*), et 24 sont couverts à 100% comme attendu par la couverture théorique (**Tableau 6**).

Pour le test DI44, seuls les 5 gènes couverts à 100% (*CDKL5, DLG3, GATAD2B, PQB1, TCF4*) ont une couverture conforme à la couverture théorique, les 39 autres gènes ayant une couverture expérimentale inférieure à la couverture théorique estimée (**Tableau 5**).

Tableau 5 : Représentation des taux de couverture > 30X par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 6 patients, sur les 44 gènes (CDS ± 25) pour le Test DI44

Tableau des taux de couverture >30X par gène, patient par patient, sur les 44 gènes (CDS ± 25)

Test DI44

Gènes	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Moyenne (écart-type) pour l'ensemble des patients
ANKRD11	99,60%	96,90%	97,00%	99,90%	99,90%	99,60%	98,82% (1,45%)
ARID1B	86,50%	81,90%	84,10%	86,50%	87,70%	81,90%	84,77% (2,51%)
ATRX	99,90%	96,80%	95,40%	95,70%	97,10%	100,00%	97,48% (2,02%)
CASK	96,70%	96,70%	93,10%	93,10%	93,10%	96,70%	94,90% (1,97%)
CDKL5	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CTNNB1	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	99,98% (0,04%)
CUL4B	100,00%	100,00%	95,80%	100,00%	100,00%	100,00%	99,30% (1,71%)
DLG3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
DYRK1A	100,00%	99,80%	100,00%	100,00%	100,00%	100,00%	99,97% (0,08%)
EP300	96,70%	94,30%	94,30%	96,20%	96,70%	96,70%	95,82% (1,19%)
FOXG1	56,00%	56,00%	56,00%	56,00%	56,00%	56,00%	56,00% (0,00%)
FOXP1	100,00%	99,80%	100,00%	100,00%	100,00%	100,00%	99,97% (0,08%)
GATAD2B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GRIA3	100,00%	99,90%	99,90%	99,90%	99,90%	99,90%	99,92% (0,04%)
GRIN1	96,60%	92,20%	96,60%	96,50%	96,70%	96,80%	95,90% (1,82%)
GRIN2A	94,60%	94,60%	94,60%	94,70%	94,90%	94,60%	94,67% (0,12%)
GRIN2B	94,70%	94,40%	94,70%	94,70%	94,70%	94,70%	94,65% (0,12%)
IL1RAPL1	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	99,95% (0,12%)
IQSEC2	87,10%	86,20%	81,90%	85,00%	85,30%	87,00%	85,42% (1,92%)
KDM5C	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70% (0,00%)
KMT2A	93,80%	93,80%	93,80%	95,60%	95,60%	95,60%	94,70% (0,99%)
MECP2	100,00%	99,80%	99,40%	99,80%	100,00%	100,00%	99,83% (0,23%)
MED13L	98,50%	98,50%	98,50%	98,50%	98,50%	98,50%	98,50% (0,00%)
NAA10	93,60%	93,60%	93,60%	93,60%	93,60%	93,60%	93,60% (0,00%)
PQBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PTCHD1	91,80%	91,80%	91,80%	91,80%	91,80%	91,80%	91,80% (0,00%)
RAI1	99,30%	99,30%	99,30%	99,30%	99,30%	99,30%	99,30% (0,00%)
SATB2	91,90%	91,90%	91,90%	91,90%	91,90%	91,90%	91,90% (0,00%)
SCN2A	99,60%	99,60%	99,60%	99,60%	99,70%	99,70%	99,63% (0,05%)
SCN8A	96,30%	96,30%	96,30%	96,30%	96,30%	96,30%	96,30% (0,00%)
SETBP1	91,50%	91,50%	91,50%	91,70%	91,50%	91,50%	91,53% (0,08%)
SHANK3	71,30%	69,00%	73,60%	72,40%	73,60%	73,60%	72,25% (1,84%)
SLC16A2	99,80%	99,80%	99,80%	99,80%	99,80%	99,80%	99,80% (0,00%)
SLC2A1	92,90%	92,30%	92,30%	92,00%	96,60%	91,90%	93,00% (1,80%)
SLC6A8	62,20%	62,20%	62,20%	62,20%	69,20%	62,40%	63,40% (2,84%)
SLC9A6	95,30%	91,20%	91,20%	91,20%	91,20%	95,30%	92,57% (2,12%)
SMARCA2	94,00%	94,00%	94,00%	94,00%	94,10%	94,00%	94,02% (0,04%)
SMC1A	92,80%	92,70%	92,70%	92,70%	92,70%	96,70%	93,38% (1,63%)
STXBP1	99,90%	99,90%	99,90%	99,90%	99,90%	99,90%	99,90% (0,00%)
SYNGAP1	88,80%	88,00%	89,00%	89,10%	88,60%	89,10%	88,77% (0,42%)
TBR1	90,10%	90,10%	90,10%	90,10%	90,10%	90,10%	90,10% (0,00%)
TCF4	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
UPF3B	99,80%	99,80%	99,80%	99,80%	99,80%	99,80%	99,80% (0,00%)
WDR45	89,90%	89,90%	89,90%	89,90%	89,90%	89,90%	89,90% (0,00%)
En moyenne (écart-type)	94,40% (9,48%)	93,60% (9,65%)	93,60% (9,38%)	94,20% (9,42%)	94,50% (8,85%)	94,60% (9,47%)	94,15% (0,45%)

Tableau 6 : Représentation des taux de couverture > 30X par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 6 patients, sur les 44 gènes (CDS ± 25) pour le Test DI117

Tableau des taux de couverture >30X par gène, patient par patient, sur les 44 gènes (CDS ± 25)

Test DI117

Gènes	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Moyenne (écart-type) pour l'ensemble des patients	
ANKRD11	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
ARID1B	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	99,98%	(0,04%)
ATRX	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
CASK	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
CDKL5	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
CTNNB1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
CUL4B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
DLG3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
DYRK1A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
EP300	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
FOXP1	91,60%	86,20%	91,60%	91,30%	87,20%	90,70%	89,77%	(2,42%)
FOXP2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
GATAD2B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
GRIA3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
GRIN1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
GRIN2A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
GRIN2B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
IL1RAPL1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
IQSEC2	99,40%	99,50%	98,40%	98,60%	99,30%	99,50%	99,12%	(0,49%)
KDM5C	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
KMT2A	99,70%	99,50%	100,00%	100,00%	99,20%	99,60%	99,67%	(0,31%)
MECP2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
MED13L	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
NAA10	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
PQBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
PTCHD1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
RAI1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SATB2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SCN2A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SCN8A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SETBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SHANK3	98,60%	98,70%	99,10%	99,20%	98,90%	99,10%	98,93%	(0,24%)
SLC16A2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SLC2A1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SLC6A8	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SLC9A6	100,00%	100,00%	100,00%	99,80%	99,80%	100,00%	99,93%	(0,10%)
SMARCA2	98,80%	98,50%	100,00%	99,90%	98,80%	99,60%	99,27%	(0,64%)
SMC1A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
STXBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
SYNGAP1	97,70%	99,30%	97,70%	98,60%	99,10%	99,60%	98,67%	(0,82%)
TBR1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
TCF4	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
UPF3B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
WDR45	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	(0,00%)
En moyenne (écart-type)	99,80% (1,32%)	99,70% (2,09%)	99,80% (1,32%)	99,80% (1,34%)	99,70% (1,94%)	99,80% (1,40%)	99,77%	(0,05%)

ii. *Sur les régions codantes avec une marge de 6 pb, et nombre attendu de « Sanger » pour compléter la couverture de cette région critique*

Nous avons également étudié la couverture en pratique minimale (avec les *gaps max*) et maximale (avec les *gaps min*) à une profondeur de 30X des régions codantes des 44 gènes, avec une marge de 6 pb (CDS ± 6), c'est-à-dire incluant uniquement les exons codants et les régions d'épissage. Nous en avons déduit le nombre de séquençages « Sanger » complémentaires nécessaires (**Tableau 7**).

De façon remarquable, pour parvenir à une couverture de 100% sur CDS ± 6 , 16 à 25 séquençages « Sanger » complémentaires seraient en pratique nécessaires pour le test DI117, contre 53 à 71 pour le test DI44 (**Figure 37, Tableau 7**).

En conclusion, le test DI117 présente indéniablement une meilleure performance que le test DI44, en ce qui concerne la couverture réelle déduite de la pratique expérimentale sur les régions CDS ± 25 et CDS ± 6 des 44 gènes en commun. Le nombre de séquençages « Sanger » nécessaires pour une couverture complète sur CDS ± 6 est deux à quatre fois inférieur avec le test DI117 par rapport au test DI44. L'apport du test DI117 en pratique est donc bien plus important que ce qui était perçu avec la couverture théorique.

Tableau 7 : Représentation des taux de couverture réelle minimale et maximale par gène sur les régions CDS ± 6 pb et CDS ± 25 pb des tests DI44 et DI117, avec indication du nombre de "Sanger" nécessaires pour compléter la couverture sur CDS ± 6 pb

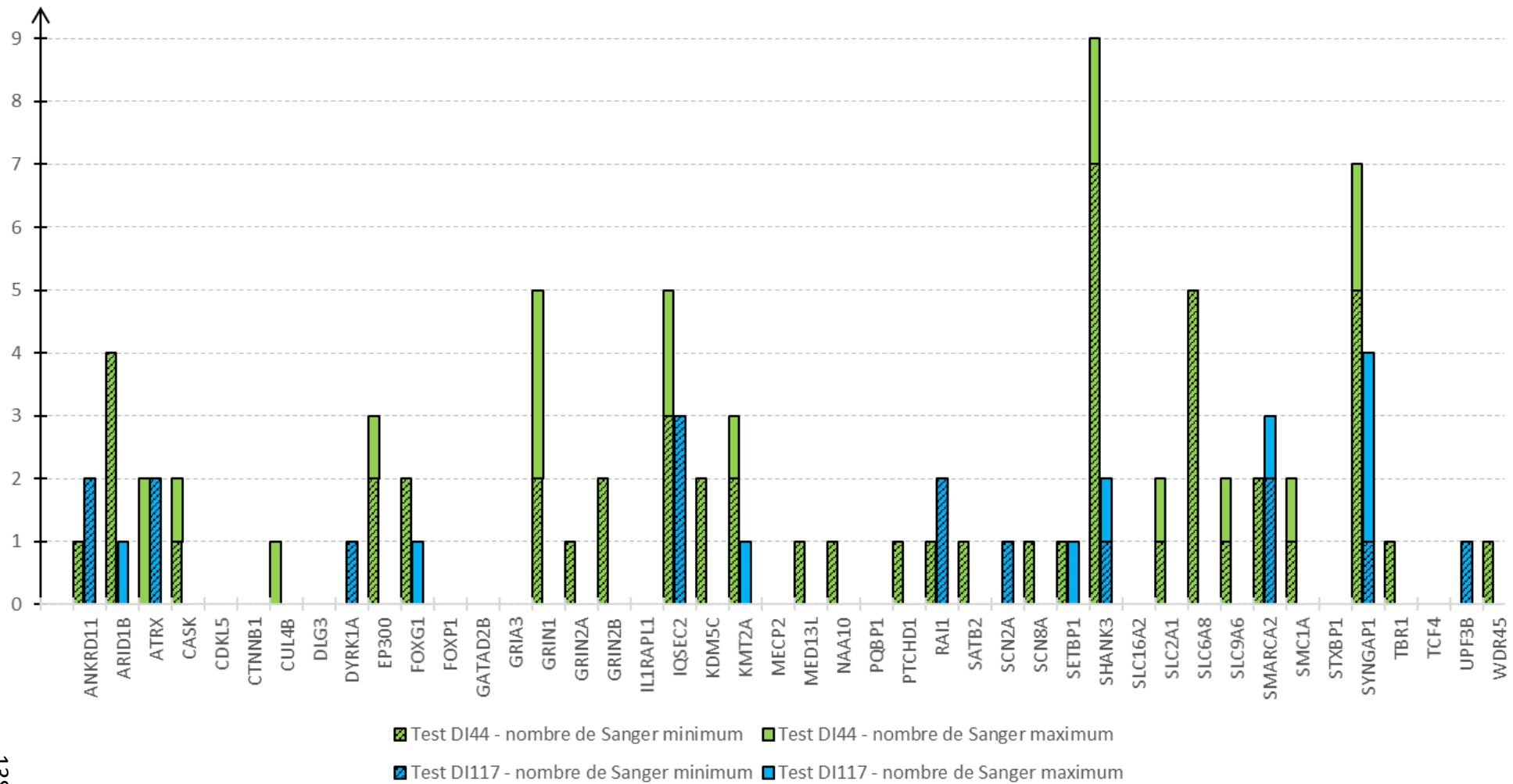
	Taux de couverture par gène								nbre de "Sanger" complémentaire(s)			
	Test DI44				Test DI117							
	Min		Max		Min		Max		Test DI44		Test DI117	
	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	Min	Max	Min	Max
ANKRD11	99,86%	99,88%	96,94%	96,97%	99,10%	99,10%	99,10%	99,10%	1	1	2	2
ARID1B	87,70%	88,19%	81,87%	82,36%	100,00%	100,00%	99,90%	99,90%	4	4	0	1
ATRX	100,00%	100,00%	95,37%	95,75%	99,01%	99,10%	99,01%	99,10%	0	2	2	2
CASK	96,74%	97,67%	93,10%	94,95%	100,00%	100,00%	100,00%	100,00%	1	2	0	0
CDKL5	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
CTNNA1	100,00%	100,00%	99,93%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
CUL4B	100,00%	100,00%	95,75%	96,76%	100,00%	100,00%	100,00%	100,00%	0	1	0	0
DLG3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
DYRK1A	100,00%	100,00%	99,75%	100,00%	99,58%	99,58%	99,58%	99,58%	0	0	1	1
EP300	96,76%	97,19%	94,14%	95,17%	100,00%	100,00%	100,00%	100,00%	2	3	0	0
FOXP1	55,72%	57,17%	55,72%	57,17%	99,28%	100,00%	85,46%	86,18%	2	2	0	1
FOXP1	100,00%	100,00%	99,75%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
GATAD2B	100,00%	100,00%	99,96%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
GRIA3	100,00%	100,00%	99,88%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
GRIN1	96,78%	97,77%	92,22%	94,21%	100,00%	100,00%	100,00%	100,00%	2	5	0	0
GRIN2A	94,72%	95,59%	94,58%	95,59%	99,92%	100,00%	99,92%	100,00%	1	1	0	0
GRIN2B	94,68%	95,92%	94,46%	95,71%	99,60%	100,00%	99,60%	100,00%	2	2	0	0
IL1RAPL1	100,00%	100,00%	99,69%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
IQSEC2	87,12%	87,90%	81,77%	82,96%	98,54%	98,81%	97,05%	97,57%	3	5	3	3
KDM5C	94,74%	96,56%	94,74%	96,56%	100,00%	100,00%	100,00%	100,00%	2	2	0	0
KMT2A	95,62%	95,96%	93,86%	94,51%	100,00%	100,00%	99,18%	99,18%	2	3	0	1

nbre de "Sanger" complémentaire(s) = nombre d'amplicon(s) séquencé(s) par la technique de Sanger pour compléter la couverture à 100%

	Taux de couverture par gène								nbre de "Sanger" complémentaire(s)			
	Test DI44				Test DI117							
	Min		Max		Min		Max		Test DI44		Test DI117	
	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	Min	Max	Min	Max
MECP2	100,00%	100,00%	99,44%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
MED13L	98,51%	98,97%	98,51%	98,97%	99,82%	100,00%	99,82%	100,00%	1	1	0	0
NAA10	93,59%	97,02%	93,59%	97,02%	100,00%	100,00%	100,00%	100,00%	1	1	0	0
PQBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
PTCHD1	91,65%	92,34%	91,65%	92,34%	100,00%	100,00%	100,00%	100,00%	1	1	0	0
RAI1	99,27%	99,59%	99,27%	99,59%	99,39%	99,39%	99,39%	99,39%	1	1	2	2
SATB2	91,82%	93,08%	91,82%	93,08%	100,00%	100,00%	100,00%	100,00%	1	1	0	0
SCN2A	99,70%	100,00%	99,63%	100,00%	99,66%	99,66%	99,66%	99,66%	0	0	1	1
SCN8A	96,27%	96,62%	96,26%	96,62%	99,94%	100,00%	99,94%	100,00%	1	1	0	0
SETBP1	91,73%	92,10%	91,49%	91,87%	100,00%	100,00%	99,74%	99,74%	1	1	0	1
SHANK3	73,16%	75,59%	68,60%	71,93%	99,32%	99,79%	98,40%	99,43%	7	9	1	2
SLC16A2	99,79%	100,00%	99,79%	100,00%	100,00%	100,00%	99,74%	100,00%	0	0	0	0
SLC2A1	96,56%	98,48%	91,86%	94,74%	100,00%	100,00%	100,00%	100,00%	1	2	0	0
SLC6A8	61,26%	66,50%	54,26%	60,95%	99,37%	100,00%	99,10%	100,00%	5	5	0	0
SLC9A6	95,29%	96,94%	91,16%	93,46%	100,00%	100,00%	99,72%	100,00%	1	2	0	0
SMARCA2	94,05%	94,97%	93,97%	94,97%	98,19%	98,19%	96,68%	96,98%	2	2	2	3
SMC1A	96,67%	97,52%	92,73%	94,41%	100,00%	100,00%	100,00%	100,00%	1	2	0	0
STXBP1	99,93%	100,00%	99,93%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
SYNGAP1	84,66%	86,57%	79,77%	82,44%	99,16%	99,64%	95,06%	96,67%	5	7	1	4
TBR1	90,12%	90,12%	90,08%	90,08%	99,66%	100,00%	99,66%	100,00%	1	1	0	0
TCF4	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	0	0	0	0
UPF3B	99,75%	100,00%	99,75%	100,00%	96,10%	96,10%	96,10%	96,10%	0	0	1	1
WDR45	89,91%	92,31%	89,91%	92,31%	100,00%	100,00%	100,00%	100,00%	1	1	0	0
pour les 44 gènes	94,68%	95,36%	92,84%	93,80%	99,65%	99,73%	99,24%	99,39%	53	71	16	25

nbre de "Sanger" complémentaire(s) = nombre d'amplicon(s) séquencé(s) par la technique de Sanger pour compléter la couverture à 100%

Figure 37 : Graphique représentant le nombre de "Sanger" nécessaires pour compléter la couverture réelle des 44 gènes (sur CDS ± 6) : Test DI44 comparé au Test DI117



b. Profondeur de lecture

Nous avons comparé les profondeurs de lecture obtenues sur les régions CDS ± 25 des 44 gènes communs aux 2 tests et chez les 6 patients analysés conjointement. L'objectif étant que cette comparaison de la profondeur moyenne gène par gène, patient par patient, apporte de nouvelles données objectives sur les performances des 2 tests NGS. Comme souligné dans l'introduction (paragraphe II.2.a, page 76) de ce travail, la profondeur de lecture du séquençage NGS est en effet un paramètre crucial pour la qualité du test diagnostique, car sa précision et sa fiabilité en dépendent (**Annexe 9**).

Pour chacun des patients, nous avons calculé la profondeur de lecture moyenne (PLM) sur les régions CDS ± 25 de l'ensemble des 44 gènes communs aux tests DI117 et DI44. Une PLM trop faible remet en cause la qualité de la librairie du patient concerné et peut rendre ininterprétable les résultats du séquençage. Le seuil habituellement fixé par notre laboratoire pour le rendu de résultats est de 200X.

Le **Tableau 8** montre que les 2 tests semblent avoir une bonne profondeur de lecture. En effet, la PLM est au minimum de 500X, pour les 2 tests.

Nous ne constatons pas de différence significative lorsque nous comparons la PLM de chaque patient entre les 2 tests (**Tableau 8**).

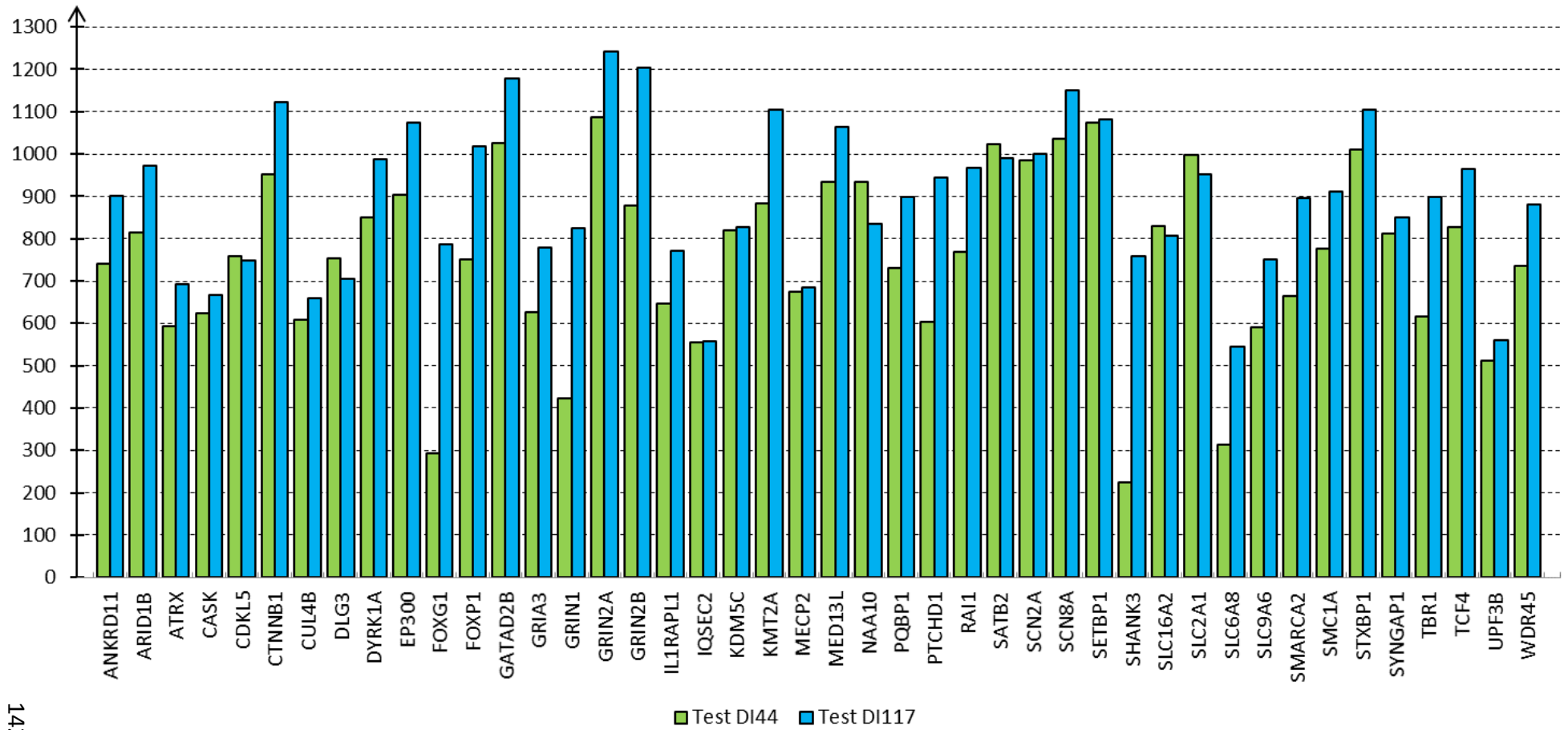
Si nous nous intéressons aux profondeurs de lecture moyennes calculées pour chaque gène, nous remarquons que pour les deux tests DI44 et DI117, toutes sont supérieures à 200X (**Figure 38**). Pour la quasi-totalité des gènes, aucun écart notable n'est observé entre les 2 tests. Toutefois, quelques gènes ont une profondeur de lecture bien meilleure avec le test DI117 (**Figure 38**). Par exemple le gène *SHANK3* a une profondeur de lecture moyenne calculée pour l'ensemble des patients de 758X avec le test DI117 contre seulement 224X avec le test DI44 (**Annexe 9**). D'ailleurs, avec le test DI44, la profondeur de lecture moyenne des gènes *FOXG1* et *SHANK3* est inférieure à 200X pour le patient 2 (**Annexe 9**).

Globalement, les PLM de tous les patients paraissent satisfaisantes pour les deux tests. Néanmoins, pour le test DI44 nous avons observé chez un patient une profondeur de lecture moyenne inférieure à 200X pour 2 gènes, alors que cela n'a été le cas pour aucun gène et pour aucun patient avec le test DI117. Le test DI117 semble donc être légèrement plus performant que le test DI44 en ce qui concerne la profondeur de lecture sur les régions CDS ± 25 des 44 gènes communs.

Tableau 8 : Tableau récapitulant la PLM (profondeur de lecture moyenne sur les régions CDS \pm 25 de l'ensemble des 44 gènes) de chaque patient pour chacun des 2 tests, DI44 et DI117

	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Moyenne (écart-type) pour l'ensemble des patients	
	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117
Moyenne (écart-type) pour les 44 gènes	919 (238)	1006 (148)	520 (133)	862 (121)	623 (222)	731 (252)	801 (303)	874 (290)	874 (317)	817 (270)	918 (239)	1240 (179)	776 (167)	922 (180)

Figure 38 : Profondeurs de lecture moyennes par gène pour les 44 gènes (CDS ± 25) : Test DI44 comparé au Test DI117 (n = 6 patients)



c. Performance diagnostique

Nous nous sommes ensuite intéressés à la capacité des 2 tests à détecter les variations de séquence pathogènes des patients atteints de DI (vrais positifs) mais aussi les fausses variations ; l'objectif étant d'évaluer les performances du séquençage en termes de sensibilité et de spécificité, deux paramètres cruciaux pour un test diagnostique.

i. Capacité à détecter un variant pathogène connu

Nous avons analysé la capacité des 2 tests DI117 et DI44 à mettre en évidence les variations réelles de séquence. Pour cela, nous nous sommes intéressés aux variants pathogènes préalablement identifiés par des laboratoires référents chez les patients inclus dans notre cohorte et qui sont considérés comme causaux de leur DI. Conformément aux bonnes pratiques, le diagnostic par NGS de ces mutations avait été confirmé par un séquençage avec la méthode de Sanger.

Pour l'ensemble des patients, les mutations responsables de leur DI ont toutes été retrouvées avec les 2 tests après analyse bio-informatique par les 2 VC utilisés, HaplotypeCaller et VarScan 2 (cf. partie introduction, paragraphe II.2.d, page 79). Les critères pris en compte pour le filtrage des données analytiques ont été les suivants : une localisation dans la région cible CDS ± 25 des 44 gènes en commun, une profondeur de lecture supérieur à 30X, une VAF* (*variant allele frequency*) supérieure à 35%, et une présence du variant chez moins de la moitié des patients au sein du *run* de séquençage (**Tableau 9**).

Il n'existe donc pas de différence notable entre les tests DI117 et DI44 pour le diagnostic des variants pathogènes des 6 patients témoins étudiés. Les 2 tests mettent en évidence les variations cibles avec une profondeur supérieure à 100X, et une VAF cohérente avec une présence à l'état hétérozygote*.

Tableau 9 : Analyse du variant confirmé en séquençage Sanger causal de chaque patient

Patients	Variants		Test DI44			Test DI117				
			1 ou 2 VC	VarScan 2 HaplotypeCaller	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2 HaplotypeCaller	VAF	Profondeur de lecture
Patient 1	ANKRD11 NM_001256182.1:c.7535G>C	mutation faux-sens p.(Arg2512Pro)	2 VC	VarScan 2 HaplotypeCaller	46,87 49,22	702 386	2 VC	VarScan 2 HaplotypeCaller	43,38 45,27	816 296
Patient 2	CDKL5 NM_003159.2:c.719G>C	mutation faux-sens p.(Ser240Thr)	2 VC	VarScan 2 HaplotypeCaller	47,84 47,49	439 259	2 VC	VarScan 2 HaplotypeCaller	50,71 50,00	777 288
Patient 3	/	/	/	/	/	/	/	/	/	/
Patient 4	RAI1 NM_030665.3:c.1958_1959delTG	mutation frameshift p.(Val653Glyfs35*)	2 VC	VarScan 2 HaplotypeCaller	53,61 53,71	748 283	2 VC	VarScan 2 HaplotypeCaller	45,43 47,03	1358 202
Patient 5	SCN8A NM_014191.3:c.3943G>C	mutation faux-sens p.(Val1315Leu)	2 VC	VarScan 2 HaplotypeCaller	51,33 52,17	1541 414	2 VC	VarScan 2 HaplotypeCaller	49,75 48,67	810 339
Patient 6	MECP2 NM_004992.3:c.806delG	mutation frameshift p.(Gly269Alafs*19)	2 VC	VarScan 2 HaplotypeCaller	44,57 47,33	902 600	2 VC	VarScan 2 HaplotypeCaller	44,19 50,75	1550 266

ii. Nombre de variations de séquence en fonction de différents filtres bio-informatiques

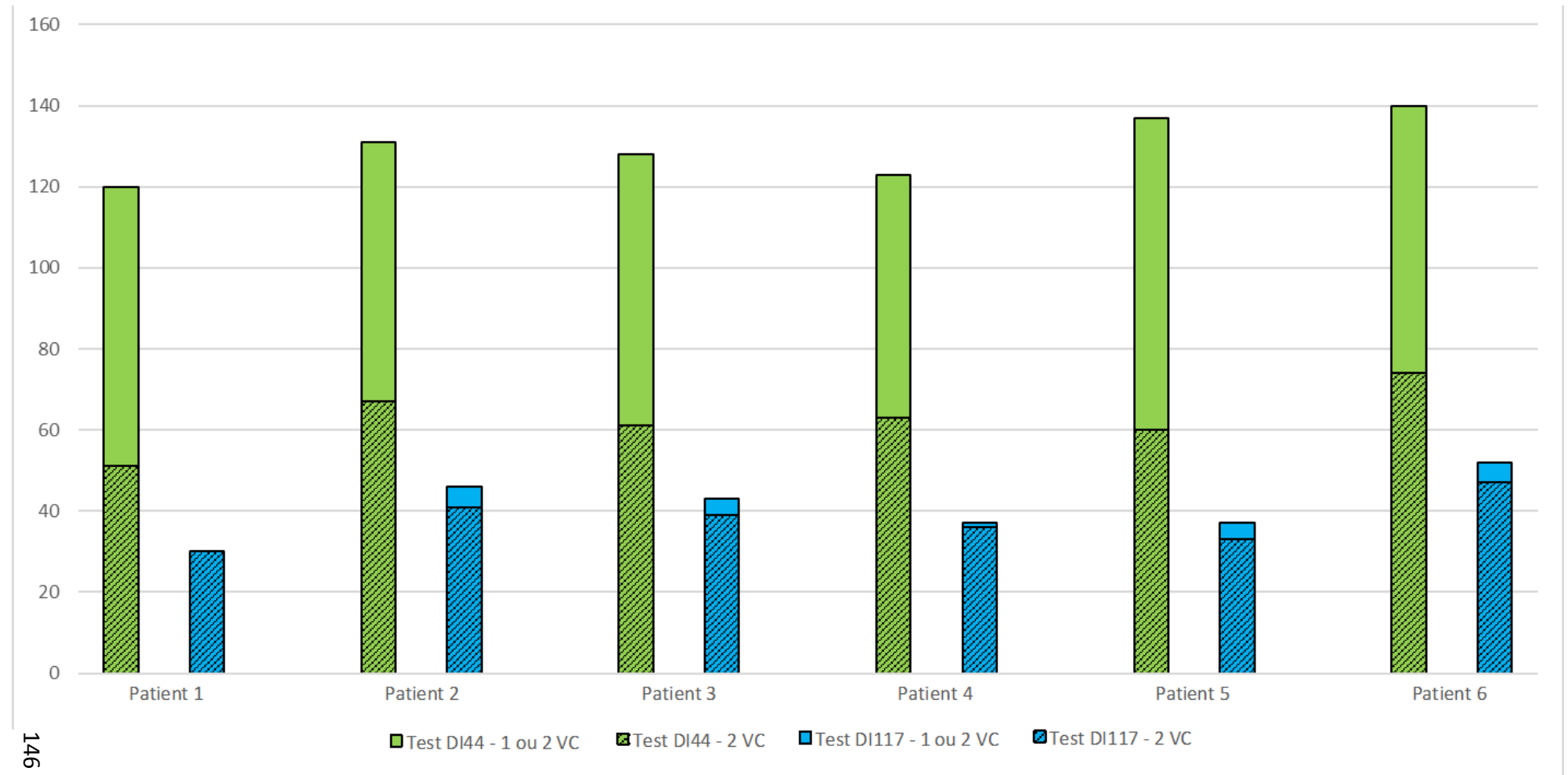
Pour poursuivre, nous avons étudié toutes les variations de séquence que chaque test met en évidence. Nous avons mis en œuvre plusieurs filtres bio-informatiques pour déterminer le nombre de variants pertinents et de variants faux-positifs détectés par chaque test.

Nous avons commencé par lister le nombre de variants retenus avec un premier filtre bio-informatique appelé « Filtre 1 » (filtre proche de celui utilisé au laboratoire pour l'analyse des résultats des patients). Pour retenir un variant, les critères de ce « Filtre 1 » sont les suivants : une localisation dans la région cible CDS ± 25 des 44 gènes en commun, une profondeur de lecture supérieure à 30X, et une VAF supérieure à 35%. Nous avons comparé les résultats obtenus avec la présence ou non du critère suivant : variant retenu par les 2 *Variants Callers* (VC), HaploTypeCaller et VarScan 2 (**Annexe 10**).

Le nombre de variants retenus avec le « Filtre 1 » et trouvés par au moins un VC (« Filtre 1 - 1 ou 2 VC ») apparaît plus élevé, environ 3 à 4 fois plus important, pour le test DI44 que pour le test DI117. Il en est de même pour le nombre de variants retenus avec le « Filtre 1 » et trouvés par les 2 VC (« Filtre 1 - 2 VC »), environ 1,5 à 2 fois plus nombreux pour le test DI44 que pour le test DI117. Le nombre de variants avec le « Filtre 1 - 2 VC » du test DI44 est d'ailleurs toujours supérieur à celui obtenu avec le « Filtre 1 - 1 ou 2 VC » du test DI117 (**Figure 39, Annexe 10**).

De plus, le nombre de variants retenus varie peu entre le « Filtre 1 – 1 ou 2 VC » et le « Filtre 1 – 2VC » pour le test DI117, alors qu'il diminue d'environ de moitié pour le test DI44 (**Figure 39, Annexe 10**).

Figure 39 : Nombre de variants retenus sur les 44 gènes avec le Filtre 1 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%) retrouvés par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117



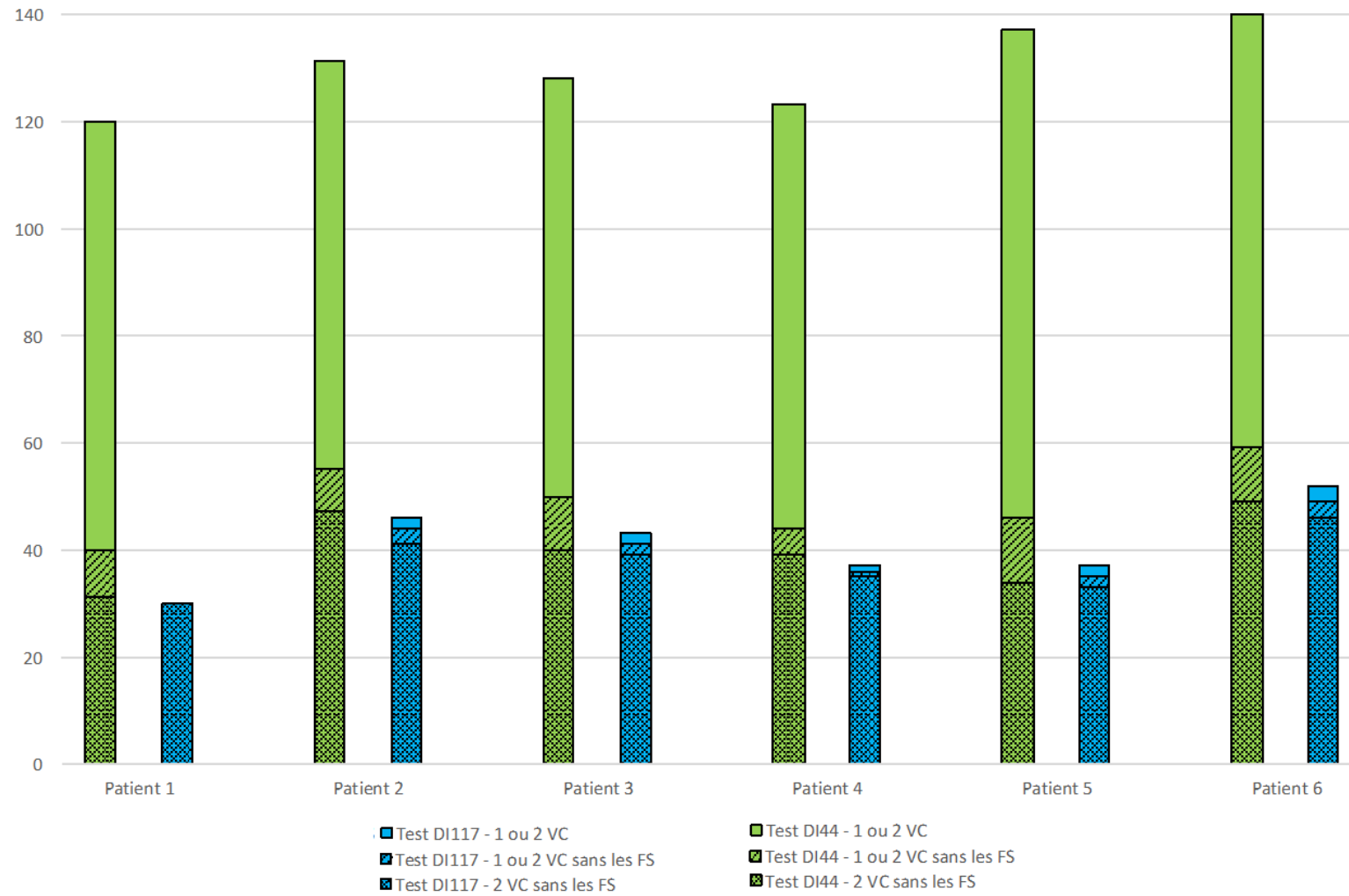
Devant ces résultats, il était pertinent d'évaluer le nombre de variants faux-positifs, c'est à dire reflétant de possibles erreurs de séquençage, retenus par le « Filtre 1 » pour les 2 tests. Pour cela, nous avons comptabilisé les variants *frameshift** (entraînant un décalage du cadre de lecture) retrouvés par les deux tests. Les variants décalant le cadre de lecture conduisent le plus souvent à un codon stop prématuré et donc à la synthèse d'une protéine tronquée ou à l'absence de synthèse de protéine ; ce sont par conséquent dans la grande majorité des cas des variants pathogènes. Or nous connaissons déjà les variants pathogènes de chaque patient existant dans les 44 gènes étudiés. Nous avons par conséquent émis l'hypothèse que les variants *frameshift* supplémentaires mis en évidence par les deux tests sont des variants faux-positifs (**Annexe 10**).

Pour le test DI44, le nombre de variants *frameshift* retenus par patient est supérieur à 75 avec le « Filtre 1 – 1 ou 2 VC » et encore au moins égal à 20 avec le « Filtre 1 – 2 VC », alors qu'il ne dépasse pas trois pour le test DI117 (**Annexe 10**).

En effet, pour le test DI44, les variants *frameshift* représentent plus de la moitié, et même parfois jusqu'au 2/3, des variants retenus par le « Filtre 1 – 1 ou 2 VC », et environ 30 à 40 % de ceux retenus par le « Filtre 1 – 2VC » (**Annexe 10**).

En revanche, en excluant les variants *frameshift*, l'analyse du nombre de variants retenus par le « Filtre 1 – 1 ou 2 VC » montre assez peu de différences entre le test DI44 et le test DI117. Cependant, ce nombre reste un peu plus élevé pour le test DI44 (8 à 11 variants supplémentaires comptabilisés selon les patients) (**Figure 40, Annexe 10**). En ce qui concerne le nombre de variants retenus par le « Filtre 1 – 2VC » en excluant les variants *frameshift*, la différence entre les 2 tests s'amenuise. En effet, selon les patients, le test DI44 compte de 1 à 6 variants supplémentaires par rapport au test DI117 (**Figure 40, Annexe 10**).

Figure 40 : Nombre de variants retenus sur les 44 gènes, en excluant les variants *frameshift*, avec le Filtre 1 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%) par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117

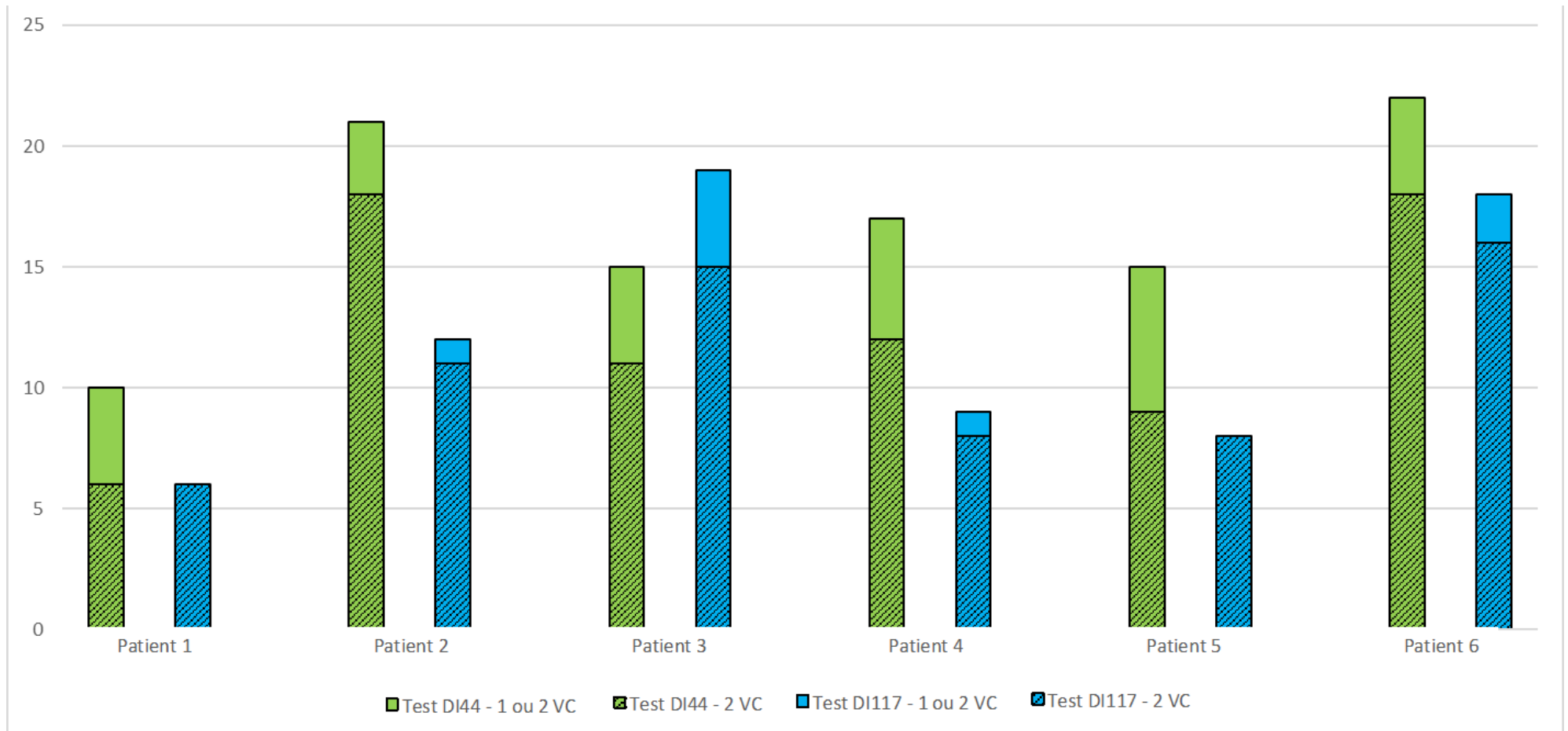


A partir des critères du « Filtre 1 », le critère suivant a été ajouté pour définir le « Filtre 2 » : une présence du variant chez moins de la moitié des patients au sein du *run* de séquençage. L'objectif de ce critère est d'éliminer les variants dus à une erreur de séquençage, dits « faux-positifs », récurrents dans un *run*, ainsi que les SNV fréquents et bénins.

Les données de la **Figure 41** n'indiquent pas de différence notable entre les 2 tests concernant le nombre de variants retenus avec le « Filtre 2 » et trouvés par au moins un VC (« Filtre 2 - 1 ou 2 VC »). Pour la majorité des patients (5 patients sur 6), ce nombre est un peu plus élevé pour le test DI44 que pour le test DI117. De même, il n'existe pas non plus de différence très marquée entre les 2 tests pour le nombre de variants retenus avec le « Filtre 2 » et trouvés par les 2 VC (« Filtre 2 - 2 VC ») (**Figure 41, Annexe 10**).

Ces résultats suggèrent que le test DI44 engendre un bien plus grand nombre de variants faux-positifs que le test DI117. En appliquant des filtres dédiés à l'élimination des variants faux-positifs, cet écart diminue, et le nombre de variants une fois filtré avec le « Filtre 2 » s'équilibre entre les deux tests. Ainsi, les résultats obtenus tendent à laisser penser que le test DI44 est plus sensible aux erreurs de séquençage liées aux technologies utilisées, ce qui expose au risque potentiel de ne pas détecter un variant pathogène masqué par ces erreurs.

Figure 41 : Nombre de variants retenus sur les 44 gènes, avec le Filtre 2 (Localisation dans régions CDS \pm 25 , Profondeur de lecture > 30X, VAF > 35%, et fréquence < 4 / run), retrouvés par 1 ou 2 VC ou par 2 VC pour le Test DI44 comparé au Test DI117



iii. Analyse des variants retenus par le Filtre 2 avec 2 VC

Le filtre bio-informatique « Filtre 2 - 2 VC » est celui habituellement utilisé en première intention au laboratoire pour l'analyse des variants candidats. Les variants conservés par ce filtre sont le plus souvent les plus pertinents, et potentiellement pathogènes (à noter que ces variants n'ont été ni confirmés ni infirmés par un séquençage Sanger). Le laboratoire s'est fixé l'objectif d'une vérification en séquençage Sanger de ces variants. Nous avons par conséquent analysé, patient par patient, les variants retenus par le « Filtre 2 - 2 VC ».

1. Analyse des variants du patient 1

Pour le patient 1, on dénombre 6 variants pour le test DI44 et 6 variants pour le test DI117.

5 variants sont communs aux 2 tests, parmi lesquels le variant pathogène du patient situé dans le gène *ANKRD11* (NM_001256182.1: c.7535G>C) (**Tableau 10**).

Un variant situé dans le gène *ARID1B* (NM_020732.3:c.228delC) n'est retrouvé que dans le test DI44, et, à l'inverse, un variant du gène *GRIN2B* (NM_000834.3:c.1341C>T) n'est décelé que dans le test DI117 (**Tableau 10**).

Le variant du gène *ARID1B* (NM_020732.3:c.228delC) est probablement un variant faux-positif du test DI44 (**Figure 42**). C'est un variant *frameshift*, situé dans un homopolymère constitué de 5 cytosines, avec une faible profondeur de lecture (inférieure à 40X) et une VAF limite à environ 35%, retrouvé 2 fois au sein du *run*. De plus, ce variant n'est pas retrouvé dans le fichier VCF des données de séquençage du laboratoire de biologie moléculaire des hôpitaux de Marseille (APHM), ayant auparavant séquencé ce patient sur leur test dédié à la DI de 142 gènes (DI142), incluant le gène *ARID1B*.

Le variant du gène *GRIN2B* (NM_000834.3:c.1341C>T) est un faux-négatif* du test DI44 (**Figure 43**). En effet, il se situe dans une région non couverte par le dessin du test DI44 (**Tableau 4, page 111**). De plus, ce variant du test DI117 est retrouvé par les 2 VC avec une bonne profondeur de lecture (supérieure à 300X) et une VAF cohérente avec une présence à l'état hétérozygote (environ 49%). Il est également détecté par le test DI142 de l'APHM.

Tableau 10 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 1

Patient 1

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117				
			1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture
MED13L	chr12:116403920-G>A	variant synonyme	2 VC	VarScan 2	51,22	1107	2 VC	VarScan 2	52,32	1011
	NM_015335.4:c.6354C>T	p.(Pro2118=)		HaplotypeCaller	48,95			HaplotypeCaller	52,78	
SETBP1	chr18:42529996-G>C	variant faux-sens	2 VC	VarScan 2	48,55	966	2 VC	VarScan 2	49,07	1349
	NM_015559.2:c.691G>C	p.(Val231Leu)		HaplotypeCaller	48,52			HaplotypeCaller	55,13	
SMARCA2	chr9:2073292-A>G	variant synonyme	2 VC	VarScan 2	52,14	1260	2 VC	VarScan 2	45,81	1229
	NM_003070.4:c.1827A>G	p.(Pro609=)		HaplotypeCaller	52,16			HaplotypeCaller	49,39	
TCF4	chr18:53070914-G>A	variant intronique	2 VC	VarScan 2	51,33	754	2 VC	VarScan 2	49,54	983
	NM_001083962.1:c.305-165C>T			HaplotypeCaller	51,59			HaplotypeCaller	49,59	
ARID1B	chr6:157099286-GC>G	variant frameshift	2 VC	VarScan 2	35,90	39				
	NM_020732.3:c.228delC	p.(Ala77Profs*34)		HaplotypeCaller	35,29					
GRIN2B	chr12:13768586-G>A	variant synonyme					2 VC	VarScan 2	49,00	749
	NM_000834.3:c.1341C>T	p.(Asp447=)						HaplotypeCaller	49,84	

Figure 42 : Visualisation sur le logiciel IGV du variant d'ARID1B (NM_020732.3:c.228delC) retenu par le Filtre 2 avec 2 VC pour le test DI44, et non retrouvé avec le test DI117

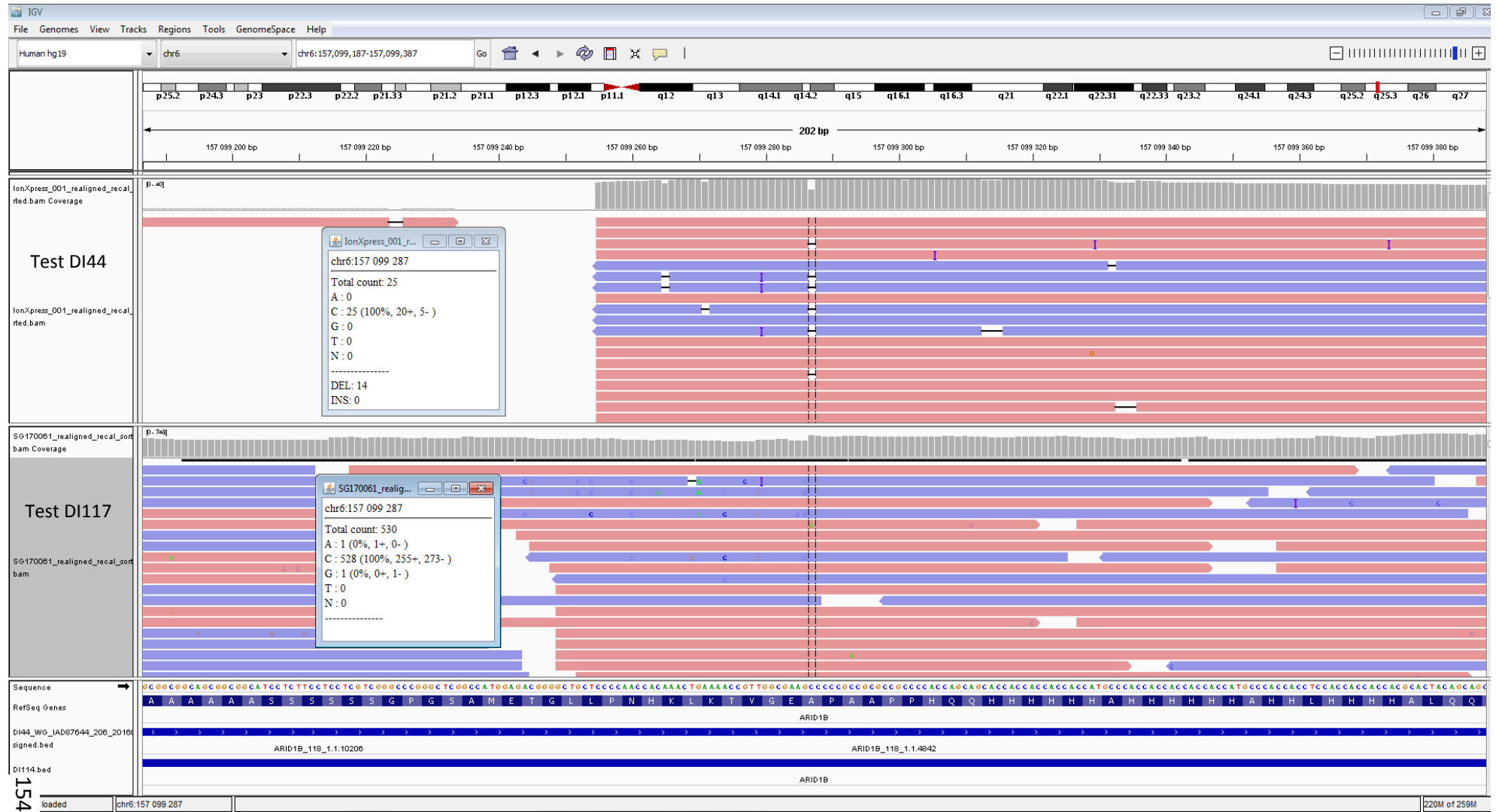
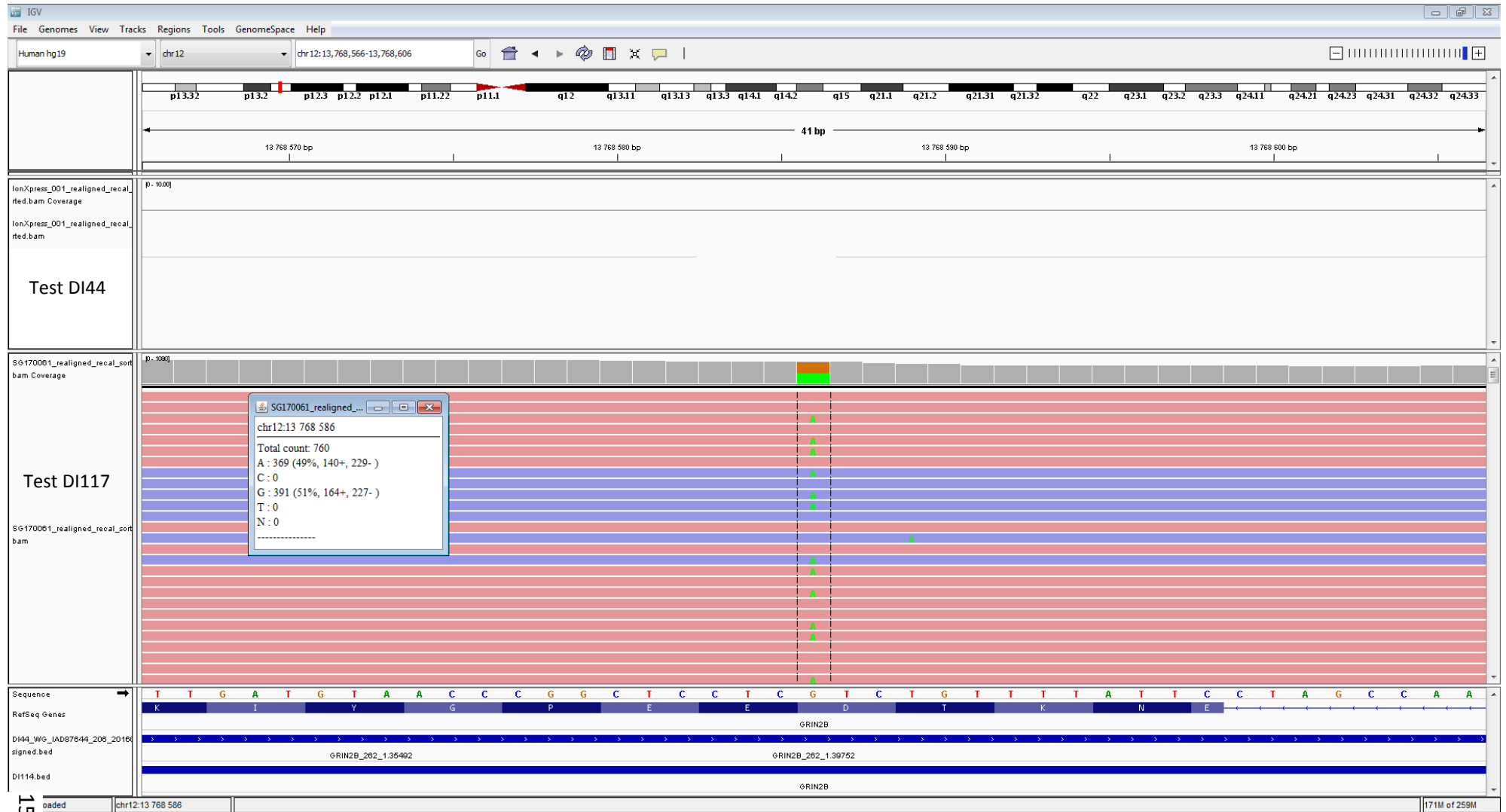


Figure 43 : Représentation sur IGV du variant de *GRIN2B* (NM_000834.3:c.1341C>T) retenu par le Filtre 2 avec 2 VC pour le test DI117



2. Analyse des variants du patient 2

Pour le patient 2, un total de 16 variants est retrouvé par les 2 tests, dont le variant pathogène du patient situé dans le gène *CDKL5* (NM_003159.2:c.719G>C), et 2 variants uniquement décelés par le test DI44 situés dans les gènes *ANKRD11* (NM_001256182.1:c.6787C>T) et *RAI1* (NM_030665.3:c.864_872delGCAGCAGCA) (**Tableau 11**).

Parmi les variants en commun, le variant du gène *SMARCA2* (NM_003070.4:c.683A>C) est détecté par le test DI117 (**Figure 44**), bien que situé dans une région non couverte par la couverture théorique de ce test (**Tableau 4, page 111**).

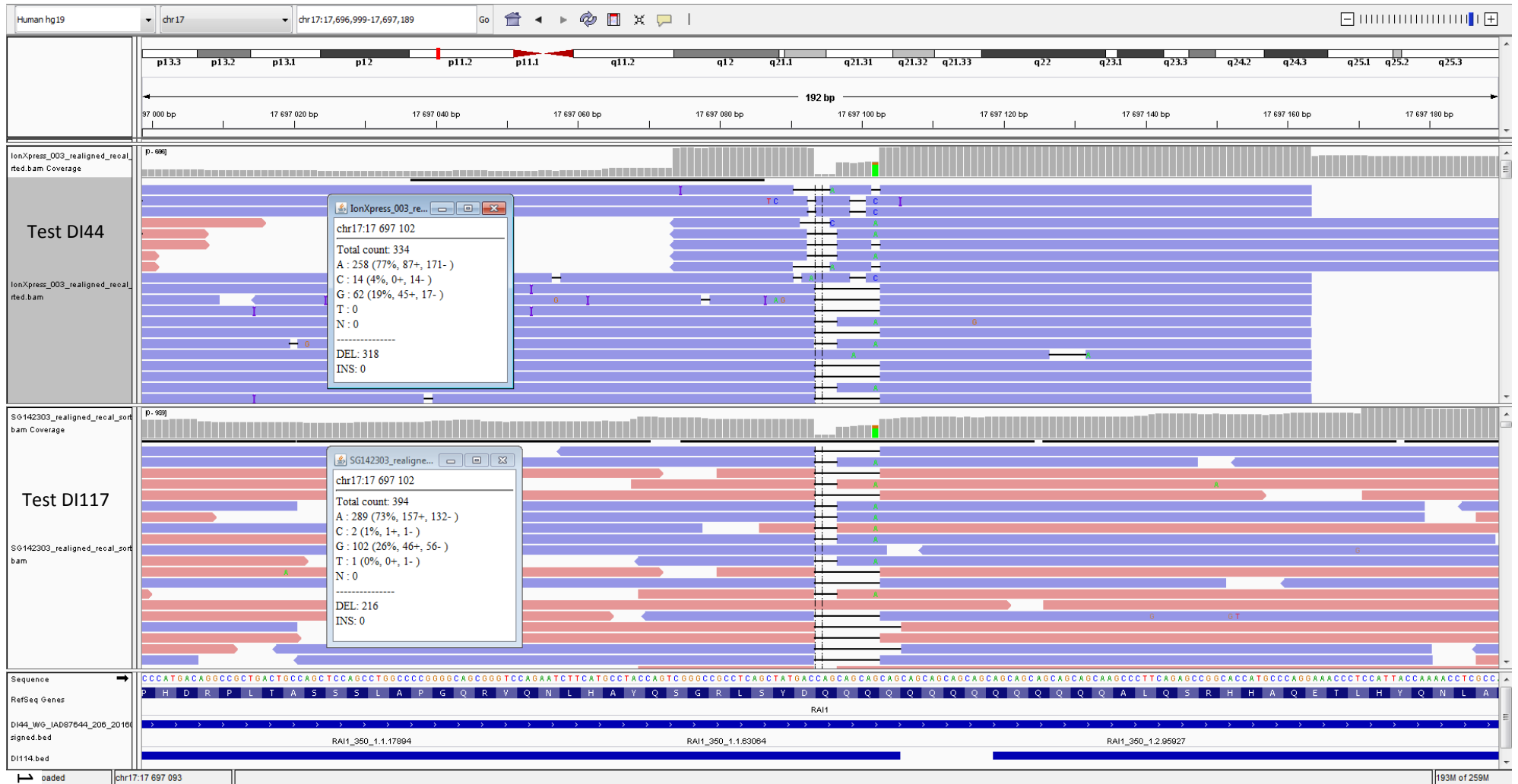
Le variant du gène *ANKRD11* (NM_001256182.1:c.6787C>T) est probablement un vrai positif* du test DI44 (**Figure 45**). En effet, il est retrouvé par les 2 VC avec une VAF cohérente avec une présence à l'état hétérozygote (environ 45%), et est présent une seule fois au sein du *run*. Cependant, la profondeur de lecture reste limitée (35 à 44X). En retirant les filtres, il est également décelé par le test DI117 par les 2 VC avec une profondeur de lecture correcte (plus de 100X) mais avec une VAF limitée (environ 33% et 40%). Il est, de la même façon, présent une seule fois au sein du *run* (**Tableau 11**).

Le variant du gène *RAI1* (NM_030665.3:c.864_872delGCAGCAGCA) est un probable vrai positif du test DI44 (**Figure 46**). En effet, il est retenu par les 2 VC avec une profondeur de lecture élevée (supérieure à 400X) et une VAF cohérente avec une présence à l'état hétérozygote (environ 43-44%). Il est fréquent au sein du *run* (présent 3 fois sur 7), et se trouve dans une région compliquée à séquencer composée de 13 motifs répétés CAG. Il est également retrouvé par le test DI117 avec une profondeur de lecture importante (supérieure à 500X) et une VAF à environ 43%, mais seulement par un seul VC et mal annoté par ce VC.

Tableau 11 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 2

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117		
			1 ou 2 VC	VAF	Profondeur de lecture	1 ou 2 VC	VAF	Profondeur de lecture
ANKRD11	chr16:89346883-C>G	variant faux-sens	2 VC	58,90	163	2 VC	54,22	166
	NM_001256182.1:c.6067G>C	p.(Ala2023Pro)		59,46	148		55,56	72
ANKRD11	chr16:89348018-C>T	variant synonyme	2 VC	52,20	887	2 VC	47,05	1203
	NM_001256182.1:c.4932G>A	p.(Gly1644=)		53,11	595		43,61	266
ANKRD11	chr16:89349128-G>A	variant synonyme	2 VC	47,35	775	2 VC	49,96	1323
	NM_001256182.1:c.3822C>T	p.(Ala1274=)		42,70	281		48,22	253
ANKRD11	chr16:89349377-G>C	variant synonyme	2 VC	49,01	808	2 VC	48,91	552
	NM_001256182.1:c.3573C>G	p.(Ala1191=)		48,29	321		51,21	207
CDKL5	chrX:18638082-A>C	variant faux-sens	2 VC	50,80	272	2 VC	44,25	400
	NM_003159.2:c.2372A>C	p.(Gln791Pro)		51,10	311		44,47	398
EP300	chr22:41548008-A>G	variant faux-sens	2 VC	54,07	381	2 VC	51,47	612
	NM_001429.3:c.2989A>G	p.(Ile997Val)		54,23	284		49,53	212
FOXP1	chr3:71021303-T>C	variant intronique	2 VC	58,19	299	2 VC	48,74	911
	NM_032682.5:c.1652+403A>G			58,51	241		53,56	295
GRIN2A	chr16:9857699-G>A	variant synonyme	2 VC	50,00	492	2 VC	48,14	966
	NM_000833.4:c.3702C>T	p.(Phe1234=)		50,31	489		51,83	191
RAI1	chr17:17700152-C>T	variant faux-sens	2 VC	45,38	260	2 VC	46,87	862
	NM_030665.3:c.3890C>T	p.(Thr1297Ile)		46,63	208		43,02	258
TCF4	chr18:53303150-T>C	variant intronique	2 VC	46,72	792	2 VC	53,82	667
	NM_001083962.1:c.-47902A>G			44,41	322		54,04	235
SLC2A1	chr1:43408966-G>A	variant synonyme	2 VC	46,97	577	2 VC	47,65	1106
	NM_006516.2:c.45C>T	p.(Ala15=)		46,96	296		47,44	527
GRIN2B	chr12:13717508-G>A	variant synonyme	2 VC	51,22	695	2 VC	47,74	1286
	NM_000834.3:c.2664C>T	p.(Thr888=)		51,51	299		50,44	343
SMARCA2	chr9:2039793-A>C	variant faux-sens	2 VC	52,94	221	NA	NA	NA
	NM_003070.4:c.683A>C	p.(Gln228Pro)		56,98	179		NA	NA
SMARCA2	chr9:2191309-C>G	variant faux-sens	2 VC	94,64	653	2 VC	100,00	487
	NM_001289396.1:c.4638C>G	p.(Asp1546Glu)		98,26	288		100,00	442
PTCHD1	chrX:23398214-T>C	variant synonyme	2 VC	48,51	435	2 VC	49,01	225
	NM_173495.2:c.858T>C	p.(Cys286=)		49,64	274		52,00	1116
ANKRD11	chr16:89346163-G>A	variant faux-sens	2 VC	45,45	44	2 VC	33,09	136
	NM_001256182.1:c.6787C>T	p.(Pro2263Ser)		45,71	35		40,52	116
RAI1	chr17:17697093-CCAGCAGCAG>C	variant frameshift	2 VC	43,06	648	1 VC	43,35	579
	NM_030665.3:c.864_872delGCAGCAGCA	p.(Gln289_Gln291del)		44,28	402			

Figure 46 : Représentation sur IGV du variant de *RAI1* (NM_030665.3:c.864_872delGCAGCAGCA) retenu par le Filtre 2 avec 2 VC pour le test DI44, et retrouvé mal annoté dans le Filtre 2 avec 1 VC pour le test DI117



3. Analyse des variants du patient 3

Pour le patient 3, le **Tableau 12** dénombre 11 variants retrouvés par les 2 tests, et 2 variants uniquement détectés par le test DI117 situés dans les gènes *GRIN1* (NM_007327.3:c.1045G>T) et *EP300* (NM_001429.3:c.2242-4delT).

Le variant du gène *GRIN1* (NM_007327.3:c.1045G>T) est un vrai positif du test DI117 (**Figure 47**). En effet, il est retenu par les 2 VC avec une profondeur de lecture élevée (supérieure à 300X) et une VAF cohérente avec une présence à l'état hétérozygote (52-55%), et il est détecté une seule fois au sein du *run*. Après retrait des filtres, ce variant est également décelé par le test DI44 par les 2 VC avec une profondeur de lecture élevée (plus de 500X) mais avec une VAF faible (inférieure à 35%). Il est, de la même façon, présent une seule fois au sein du *run*. Il est aussi mis en évidence avec le test DI142 de l'APHM.

Le variant du gène *EP300* (NM_001429.3:c.2242-4delT) est probablement un faux-négatif du test DI44 (**Figure 48**). Il est en effet détecté avec le test DI117 par les 2 VC avec une profondeur de lecture élevée (supérieure à 300X), une VAF cohérente avec un état hétérozygote, et n'est présent qu'une seule fois au sein du *run*. La visualisation sur IGV (**Figure 48**) révèle qu'il se trouve au sein d'un homopolymère de T, région difficile à séquencer avec la technologie PGM™ Ion Torrent™.

Tableau 12 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 3

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117				
			1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture
GRIN2B	chr12:13720043-G>A NM_000834.3:c.2514C>T	variant synonyme p.(Cys838=)	2 VC	VarScan 2	48,26	893	2 VC	VarScan 2	48,88	1117
				HaplotypeCaller	43,87	579		HaplotypeCaller	49,53	535
MED13L	chr12:116408538-A>G NM_015335.4:c.5928T>C	variant synonyme p.(Thr1976=)	2 VC	VarScan 2	100,00	94	2 VC	VarScan 2	99,89	908
				HaplotypeCaller	100,00	104		HaplotypeCaller	100,00	603
MED13L	chr12:116429689-A>G NM_015335.4:c.3070T>C	variant synonyme p.(Leu1024=)	2 VC	VarScan 2	99,93	1349	2 VC	VarScan 2	100,00	799
				HaplotypeCaller	100,00	342		HaplotypeCaller	100,00	318
MED13L	chr12:116446445-C>T NM_015335.4:c.1773G>A	variant synonyme p.(Gln591=)	2 VC	VarScan 2	99,80	1468	2 VC	VarScan 2	100,00	1179
				HaplotypeCaller	99,82	550		HaplotypeCaller	100,00	256
SETBP1	chr18:42532606-G>A NM_015559.2:c.3301G>A	variant faux-sens p.(Val1101Ile)	2 VC	VarScan 2	47,83	207	2 VC	VarScan 2	47,69	1428
				HaplotypeCaller	47,83	207		HaplotypeCaller	49,69	324
SCN2A	chr2:166152389-G>A NM_021007.2:c.56G>A	variant faux-sens p.(Arg19Lys)	2 VC	VarScan 2	50,39	639	2 VC	VarScan 2	48,02	933
				HaplotypeCaller	48,60	286		HaplotypeCaller	48,45	452
DYRK1A	chr21:38850536-C>T NM_001396.3:c.261C>T	variant synonyme p.(Asp87=)	2 VC	VarScan 2	56,50	354	2 VC	VarScan 2	49,80	996
				HaplotypeCaller	57,06	326		HaplotypeCaller	48,48	623
SMARCA2	chr9:2073292-A>G NM_003070.4:c.1827A>G	variant synonyme p.(Pro609=)	2 VC	VarScan 2	53,30	1227	2 VC	VarScan 2	47,38	1051
				HaplotypeCaller	53,31	272		HaplotypeCaller	48,39	560
CDKL5	chrX:18638082-A>C NM_003159.2:c.2372A>C	variant faux-sens p.(Gln791Pro)	2 VC	VarScan 2	100,00	272	2 VC	VarScan 2	99,38	160
				HaplotypeCaller	100,00	244		HaplotypeCaller	99,37	158
GRIN2A	chr16:9916204-C>G NM_000833.4:c.2085G>C	variant synonyme p.(Arg695=)	2 VC	VarScan 2	50,55	548	2 VC	VarScan 2	50,46	989
				HaplotypeCaller	49,62	397		HaplotypeCaller	52,45	286
EP300	chr22:41548008-A>G NM_001429.3:c.2989A>G	variant faux-sens p.(Ile997Val)	2 VC	VarScan 2	99,61	509	2 VC	VarScan 2	100,00	804
				HaplotypeCaller	100,00	325		HaplotypeCaller	100,00	246
GRIN1	chr9:140051238-A>G NM_007327.3:c.789A>G	variant synonyme p.(Pro284=)	2 VC	VarScan 2	47,65	554	2 VC	VarScan 2	49,24	794
				HaplotypeCaller	45,97	472		HaplotypeCaller	47,37	304
GRIN1	chr9:140051376-G>A NM_007327.3:c.855G>A	variant synonyme p.(Val306=)	2 VC	VarScan 2	50,15	329	2 VC	VarScan 2	50,53	1039
				HaplotypeCaller	51,27	314		HaplotypeCaller	51,67	329
RAI1	chr17:17696531-G>C NM_030665.3:c.269G>C	variant faux-sens p.(Gly90Ala)	2 VC	VarScan 2	100,00	399	2 VC	VarScan 2	99,58	480
				HaplotypeCaller	100,00	335		HaplotypeCaller	100,00	223
RAI1	chr17:17696755-C>A NM_030665.3:c.493C>A	variant faux-sens p.(Pro165Thr)	2 VC	VarScan 2	99,14	1389	2 VC	VarScan 2	99,32	1024
				HaplotypeCaller	99,25	265		HaplotypeCaller	98,96	289
EP300	chr22:41545024-GT>G NM_001429.3:c.2242-4delT	variant d'épissage					2 VC	VarScan 2	52,60	597
								HaplotypeCaller	61,28	390
GRIN1	chr9:140052907-G>T NM_007327.3:c.1045G>T	variant faux-sens p.(Ala370Ser)	2 VC	VarScan 2	34,42	1014	2 VC	VarScan 2	54,69	916
				HaplotypeCaller	22,55	510		HaplotypeCaller	52,96	389

Figure 47 : Représentation sur IGV du variant de *GRIN1* (NM_007327.3:c.1045G>T) retenu par le Filtre 2 avec 2 VC pour le test DI117, et décelé par le test DI44 mais éliminé par le Filtre 1 à cause d'une VAF < 35%

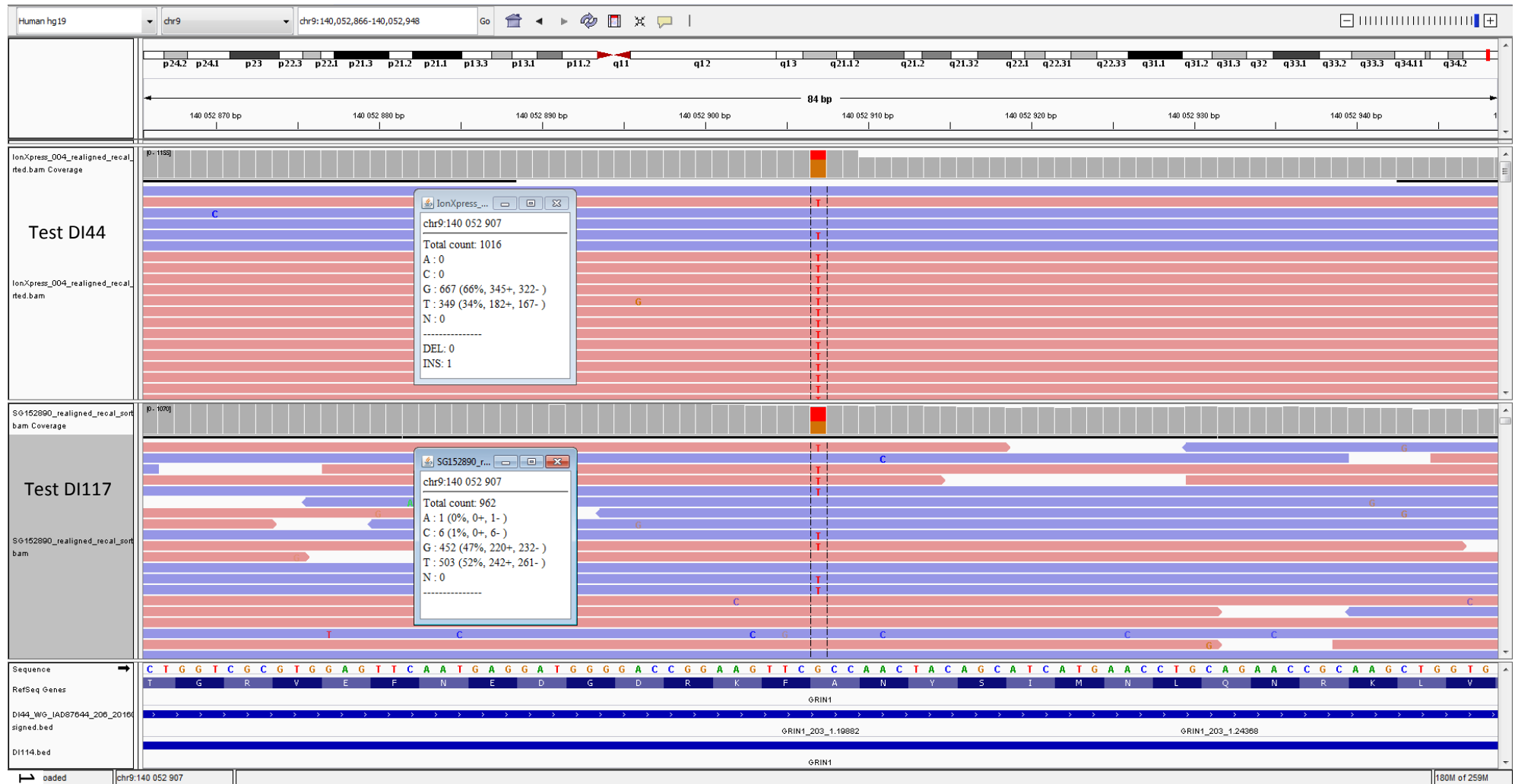
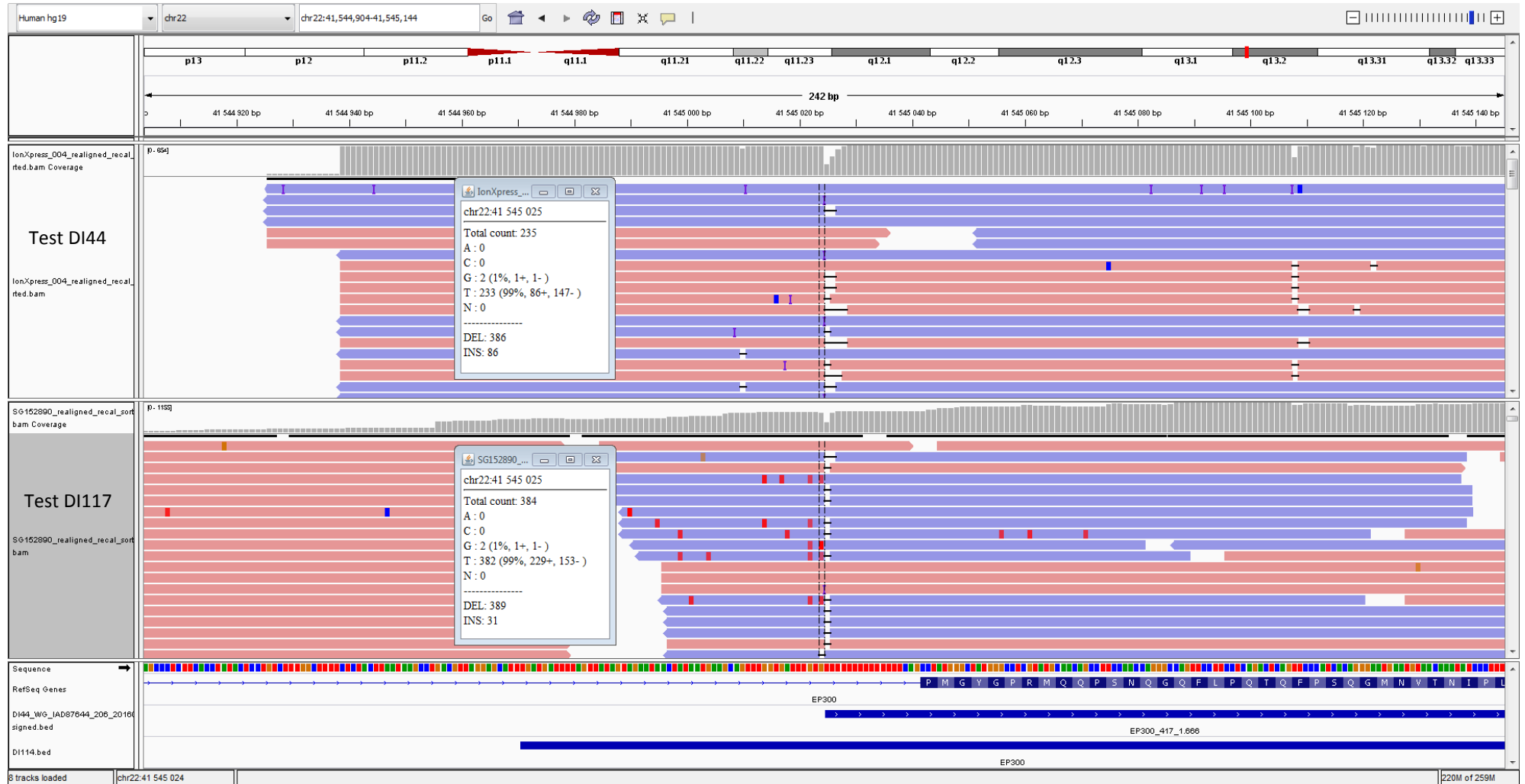


Figure 48 : Représentation sur IGV du variant d'EP300 (NM_001429.3:c.2242-4delT) retenu par le Filtre 2 avec 2 VC pour le test DI117



4. Analyse des variants du patient 4

Pour le patient 4, on énumère 12 variants communs aux 2 tests, dont le variant pathogène du patient situé dans le gène *RAI1* (NM_030665.3:c.1958_1959delTG), et un variant uniquement retenu par le test DI117 situé dans le gène *GRIN2A* (NM_000833.4:c.1275G>A) (**Tableau 13**).

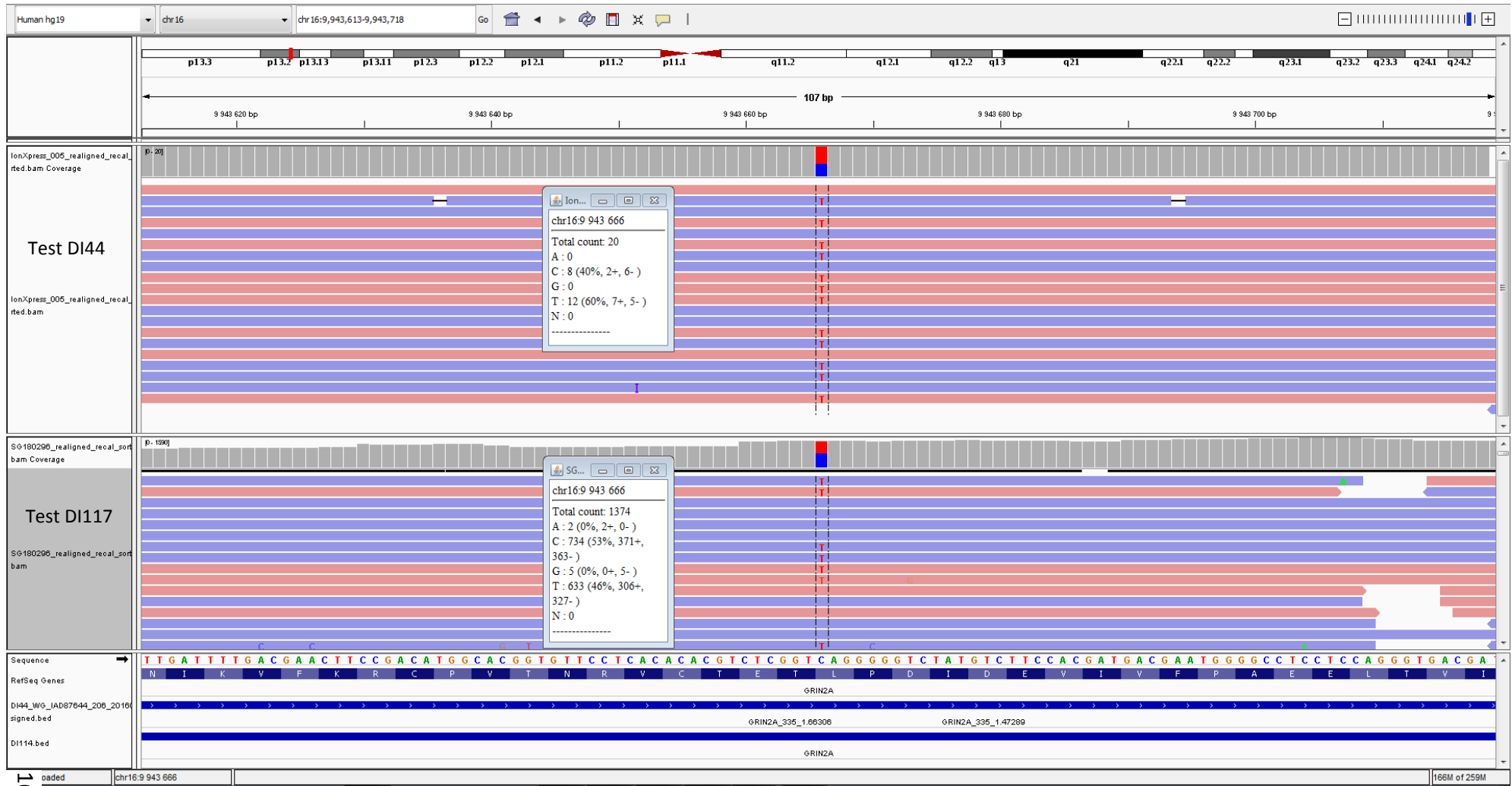
Le variant du gène *GRIN2A* (NM_000833.4:c.1275G>A) est un faux-négatif du test DI44 (**Figure 49**). En effet, avec le test DI117, il est détecté par les 2 VC avec une profondeur de lecture élevée (supérieure à 500X) et une VAF cohérente avec une présence à l'état hétérozygote (45% environ). De plus, il est situé dans une région de *Gap 30X* du test DI44 (**Annexe 7**). Par ailleurs, en l'absence de filtre, on le retrouve avec les 2 VC à une profondeur de lecture d'environ 20X et avec une VAF d'environ 60%.

Tableau 13 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 4

Patient 4

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117				
			1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture
MED13L	chr12:116421133-A>T NM_015335.4:c.4744T>A	variant faux-sens p.(Ser1582Thr)	2 VC	VarScan 2	47,22	1476	2 VC	VarScan 2	43,63	1123
				HaplotypeCaller	48,19	637		HaplotypeCaller	41,92	229
MED13L	chr12:116422049-G>A NM_015335.4:c.4467C>T	variant synonyme p.(Ser1489=)	2 VC	VarScan 2	53,88	1522	2 VC	VarScan 2	49,00	1304
				HaplotypeCaller	51,42	282		HaplotypeCaller	50,10	489
SETBP1	chr18:42532693-C>A NM_015559.2:c.3388C>A	variant faux-sens p.(Pro1130Thr)	2 VC	VarScan 2	43,35	263	2 VC	VarScan 2	53,15	1319
				HaplotypeCaller	43,46	260		HaplotypeCaller	57,74	239
SHANK3	chr22:51153371-G>A NM_033517.1:c.2119G>A	variant faux-sens p.(Ala707Thr)	2 VC	VarScan 2	50,47	212	2 VC	VarScan 2	54,87	811
				HaplotypeCaller	50,00	206		HaplotypeCaller	53,93	356
GRIA3	chrX:122536945-G>A NM_000828.4:c.1181G>A	variant faux-sens p.(Arg394Gln)	2 VC	VarScan 2	95,65	207	2 VC	VarScan 2	100,00	312
				HaplotypeCaller	100,00	178		HaplotypeCaller	100,00	300
SLC2A1	chr1:43395635-C>T NM_006516.2:c.588G>A	variant synonyme p.(Pro196=)	2 VC	VarScan 2	53,93	573	2 VC	VarScan 2	55,50	955
				HaplotypeCaller	53,18	1619		HaplotypeCaller	53,30	364
SLC2A1	chr1:43396414-G>A NM_006516.2:c.399C>T	variant synonyme p.(Cys133=)	2 VC	VarScan 2	48,21	112	2 VC	VarScan 2	51,41	673
				HaplotypeCaller	47,27	110		HaplotypeCaller	47,01	251
SLC2A1	chr1:43408966-G>A NM_006516.2:c.45C>T	variant synonyme p.(Ala15=)	2 VC	VarScan 2	53,99	1128	2 VC	VarScan 2	46,97	1205
				HaplotypeCaller	55,24	353		HaplotypeCaller	46,82	519
GRIN2B	chr12:13717508-G>A NM_000834.3:c.2664C>T	variant synonyme p.(Thr888=)	2 VC	VarScan 2	51,72	1514	2 VC	VarScan 2	47,67	1462
				HaplotypeCaller	53,85	338		HaplotypeCaller	48,97	341
GRIN2B	chr12:14018777-G>C NM_000834.3:c.366C>G	variant synonyme p.(Pro122=)	2 VC	VarScan 2	50,56	89	2 VC	VarScan 2	45,46	1014
				HaplotypeCaller	52,87	87		HaplotypeCaller	46,19	210
SMARCA2	chr9:2191309-C>G NM_003070.4:c.4638C>G	variant faux-sens p.(Asp1546Glu)	2 VC	VarScan 2	48,49	1555	2 VC	VarScan 2	46,91	663
				HaplotypeCaller	44,27	384		HaplotypeCaller	47,71	436
SYNGAP1	chr6:33408680-G>A NM_006772.2:c.1851G>A	variant synonyme p.(Glu617=)	2 VC	VarScan 2	52,67	2343	2 VC	VarScan 2	100,00	146
				HaplotypeCaller	55,39	919		HaplotypeCaller	100,00	142
GRIN2A	chr16:9943666-C>T NM_000833.4:c.1275G>A	variant synonyme p.(Leu425=)	2 VC	VarScan 2	60,00	20	2 VC	VarScan 2	45,91	1333
				HaplotypeCaller	63,16	19		HaplotypeCaller	46,44	519

Figure 49 : Représentation sur IGV du variant de *GRIN2A* (NM_000833.4:c.1275G>A) retenu par le Filtre 2 avec 2 VC pour le test DI117, et présent dans une région de *Gap 30X* pour le test DI44



5. Analyse des variants du patient 5

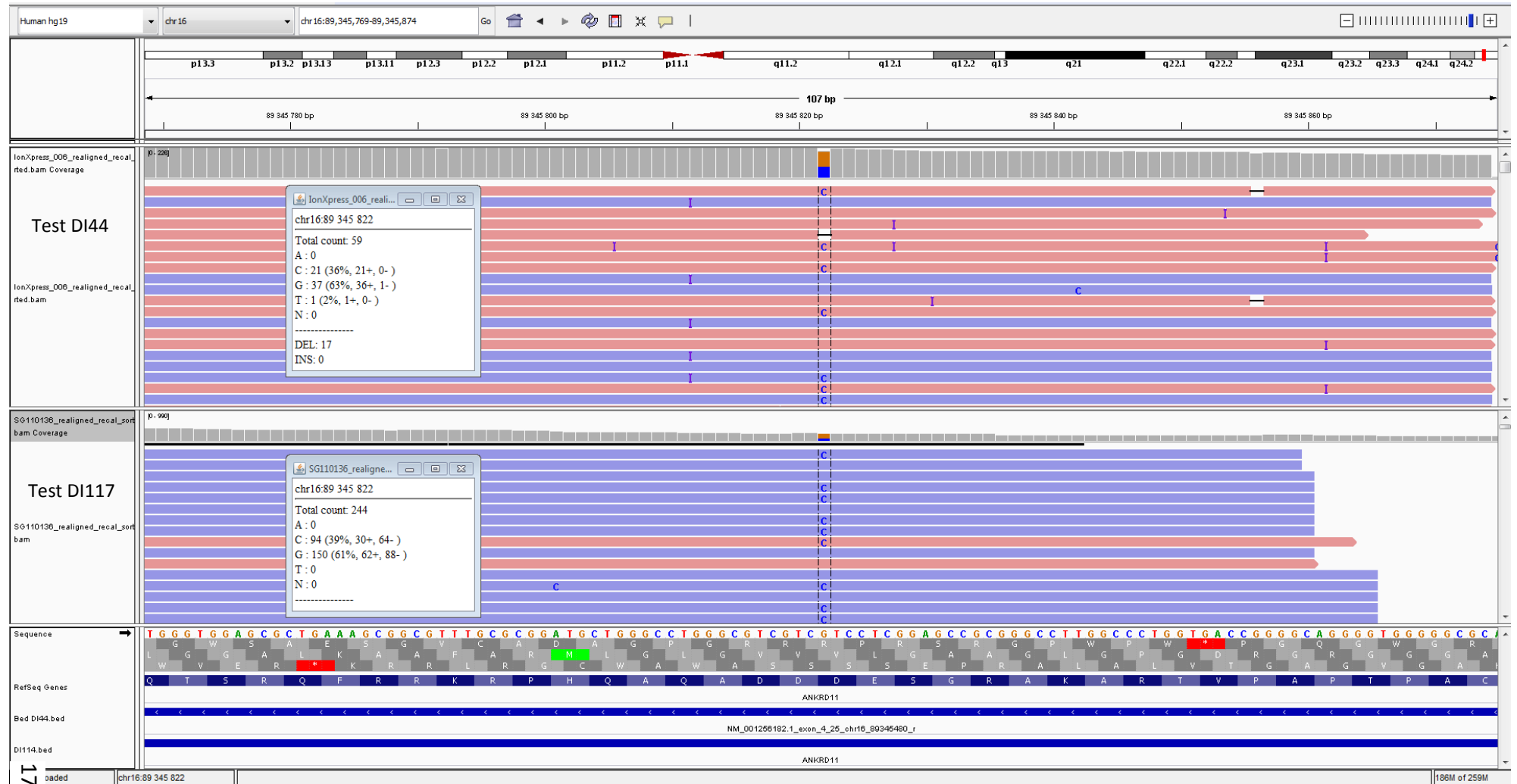
Pour le patient 5, 9 variants sont communs aux 2 tests, dont le variant pathogène du patient situé dans le gène *SCN8A* (NM_014191.3 : c.3943G>C) (**Tableau 14**).

Un variant du gène *ANKRD11* (NM_001256182.1:c.7128C>G) n'est décelé que par le test DI117 ; c'est probablement un vrai positif de ce test (**Figure 50**). En effet, il est retrouvé par les 2 VC avec une bonne profondeur de lecture (supérieure à 100X), une VAF cohérente avec une présence à l'état hétérozygote (environ 46 à 48%), et est présent une seule fois au sein du *run*. En retirant les filtres, il est également décelé par le test DI44 par les 2 VC avec une profondeur de lecture correcte (plus de 100X) mais avec une VAF limite (environ 33% et 38%), et il est, de la même façon, présent une seule fois au sein du *run* (**Tableau 14**).

Tableau 14 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 5

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117				
			1 ou 2 VC	VAF	Profondeur de lecture	1 ou 2 VC	VAF	Profondeur de lecture		
RAI1	chr17:17700788-A>G	variant faux-sens	2 VC	VarScan 2	49,32	730	2 VC	VarScan 2	42,37	616
	NM_030665.3:c.4526A>G	p.(Gln1509Arg)		HaplotypeCaller	47,55	286		HaplotypeCaller	40,71	253
SETBP1	chr18:42618533-A>C	variant faux-sens	2 VC	VarScan 2	48,46	1009	2 VC	VarScan 2	45,93	860
	NM_015559.2:c.4084A>C	p.(Lys1362Gln)		HaplotypeCaller	46,13	323		HaplotypeCaller	49,12	452
SMARCA2	chr9:2073292-A>G	variant synonyme	2 VC	VarScan 2	48,92	1433	2 VC	VarScan 2	43,92	1259
	NM_003070.4:c.1827A>G	p.(Pro609=)		HaplotypeCaller	48,75	279		HaplotypeCaller	42,34	418
SLC9A6	chrX:135095545-A>C	variant synonyme	2 VC	VarScan 2	95,92	368	2 VC	VarScan 2	100,00	576
	NM_001042537.1:c.1185A>C	p.(Ala395=)		HaplotypeCaller	96,15	260		HaplotypeCaller	100,00	570
GRIN2B	chr12:13717508-G>A	variant synonyme	2 VC	VarScan 2	52,47	1799	2 VC	VarScan 2	43,76	1467
	NM_000834.3:c.2664C>T	p.(Thr888=)		HaplotypeCaller	54,95	364		HaplotypeCaller	44,10	390
GRIN2B	chr12:14018777-G>C	variant synonyme	2 VC	VarScan 2	42,17	83	2 VC	VarScan 2	40,61	990
	NM_000834.3:c.366C>G	p.(Pro122=)		HaplotypeCaller	42,50	80		HaplotypeCaller	36,84	285
SCN8A	chr12:52174435-C>T	variant synonyme	2 VC	VarScan 2	52,73	1631	2 VC	VarScan 2	41,47	258
	NM_014191.3:c.3822C>T	p.(Val1274=)		HaplotypeCaller	52,46	509		HaplotypeCaller	44,26	183
TCF4	chr18:53070914-G>A	variant faux-sens	2 VC	VarScan 2	48,61	973	2 VC	VarScan 2	53,06	1029
	NM_001083962.1:c.305-165C>T	p.(Ala11Val)		HaplotypeCaller	49,64	413		HaplotypeCaller	51,67	240
ANKRD11	chr16:89345822-G>C	variant faux-sens	2 VC	VarScan 2	38,35	156	2 VC	VarScan 2	48,41	157
	NM_001256182.1:c.7128C>G	p.(Asp2376Glu)		HaplotypeCaller	32,67	150		HaplotypeCaller	46,22	119

Figure 50 : Représentation sur IGV du variant d'ANKRD11 (NM_001256182.1:c.7128C>G) retenu par le Filtre 2 avec 2 VC pour le test DI117 et par le Filtre 2 avec 1 VC pour le test DI44



6. Analyse des variants du patient 6

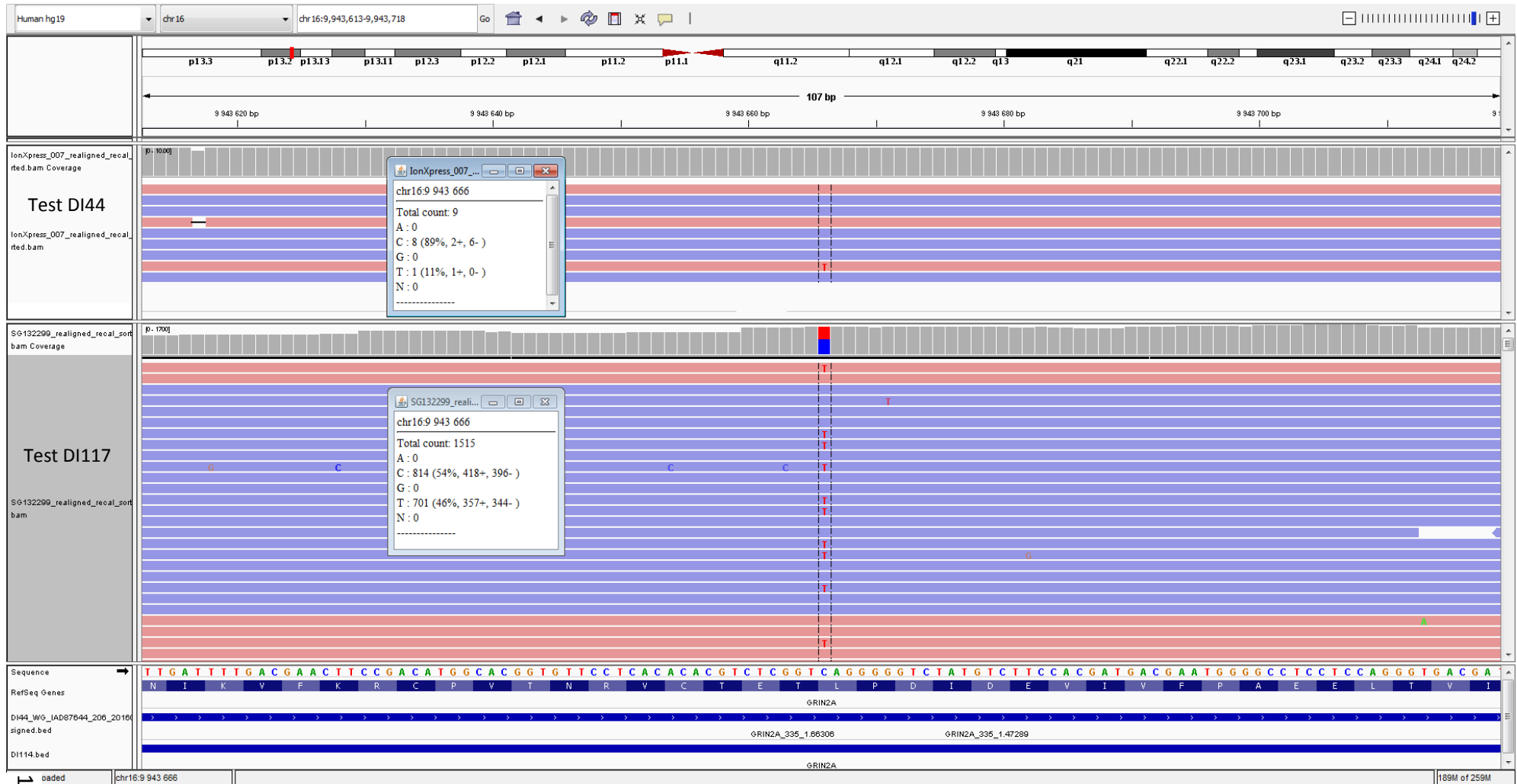
Pour le patient 6, 18 variants sont retrouvés par les 2 tests, dont le variant pathogène du patient situé dans le gène *MECP2* (NM_004992.3:c.806delG) (**Tableau 15**).

Un variant du gène *GRIN2A* (NM_000833.4:c.1275G>A) est détecté seulement avec le test DI117 ; c'est un probable faux-négatif du test DI44 (**Figure 51**). Il est décelé, avec le test DI117, par les 2 VC avec une profondeur de lecture élevée (supérieure à 400X) et une VAF cohérente avec un état hétérozygote (46 à 52%), présent 2 fois au sein du *run*. Il est par ailleurs situé dans une région de *gap 30X* du test DI44.

Tableau 15 : Liste des variants retenus par le « Filtre 2 - 2 VC » du patient 6

Variants retenus avec le "Filtre 2 - 2 VC" (autre que variant causal confirmé en Sanger)			Test DI44			Test DI117				
			1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2	VAF	Profondeur de lecture
SCN8A	NM_014191.3:c.2997C>T chr12:52162744-C>T	variant synonyme p.(Ile999=)	2 VC	VarScan 2	49,96	1173	2 VC	VarScan 2	51,20	1789
				HaplotypeCaller	50,48	523		HaplotypeCaller	50,18	283
MED13L	NM_015335.4:c.1095G>A chr12:116452994-C>T	variant synonyme p.(Ser365=)	2 VC	VarScan 2	53,03	758	2 VC	VarScan 2	49,64	1646
				HaplotypeCaller	57,41	324		HaplotypeCaller	49,52	624
GRIN2A	NM_000833.4:c.2085G>C chr16:9916204-C>G	variant synonyme p.(Arg695=)	2 VC	VarScan 2	50,43	690	2 VC	VarScan 2	43,06	1391
				HaplotypeCaller	51,97	456		HaplotypeCaller	43,40	371
ANKRD11	NM_001256182.1:c.4344T>C chr16:89348606-A>G	variant synonyme p.(Tyr1448=)	2 VC	VarScan 2	53,32	2170	2 VC	VarScan 2	46,82	1023
				HaplotypeCaller	53,26	768		HaplotypeCaller	46,36	302
ANKRD11	NM_001256182.1:c.1020G>A chr16:89351930-C>T	variant synonyme p.(Thr340=)	2 VC	VarScan 2	49,77	1300	2 VC	VarScan 2	52,12	1581
				HaplotypeCaller	49,46	463		HaplotypeCaller	50,66	229
ANKRD11	NM_001256182.1:c.890C>T chr16:89352449-G>A	variant faux-sens p.(Thr297Met)	2 VC	VarScan 2	51,40	1568	2 VC	VarScan 2	47,17	1255
				HaplotypeCaller	51,51	299		HaplotypeCaller	47,72	417
SETBP1	NM_015559.2:c.691G>C chr18:42529996-G>C	variant faux-sens p.(Val231Leu)	2 VC	VarScan 2	53,01	996	2 VC	VarScan 2	45,87	1635
				HaplotypeCaller	51,68	596		HaplotypeCaller	47,02	302
SETBP1	NM_015559.2:c.3301G>A chr18:42532606-G>A	variant faux-sens p.(Val1101Ile)	2 VC	VarScan 2	49,79	239	2 VC	VarScan 2	47,93	1911
				HaplotypeCaller	50,64	233		HaplotypeCaller	53,94	241
SCN2A	NM_021007.2:c.5326C>T chr2:166245642-C>T	variant synonyme p.(Leu1776=)	2 VC	VarScan 2	50,97	1911	2 VC	VarScan 2	47,65	2061
				HaplotypeCaller	50,17	301		HaplotypeCaller	41,95	205
EP300	NM_001429.3:c.3348G>A chr22:41553259-G>A	variant synonyme p.(Gln1116=)	2 VC	VarScan 2	50,44	456	2 VC	VarScan 2	46,30	1337
				HaplotypeCaller	52,38	777		HaplotypeCaller	51,09	411
ARID1B	NM_020732.3:c.5922C>T chr6:157528197-C>T	variant synonyme p.(Ile1974=)	2 VC	VarScan 2	53,02	1690	2 VC	VarScan 2	48,44	1860
				HaplotypeCaller	52,89	380		HaplotypeCaller	48,34	302
SMARCA2	NM_003070.4:c.3672G>A chr9:2116037-G>A	variant synonyme p.(Glu1224=)	2 VC	VarScan 2	51,16	1902	2 VC	VarScan 2	45,80	941
				HaplotypeCaller	47,62	273		HaplotypeCaller	48,63	292
IQSEC2	NM_001111125.2:c.4104G>A chrX:53263764-C>T	variant synonyme p.(Leu1368=)	2 VC	VarScan 2	54,41	816	2 VC	VarScan 2	46,89	450
				HaplotypeCaller	55,64	275		HaplotypeCaller	49,41	425
IQSEC2	NM_001111125.2:c.1641G>A chrX:53280117-C>T	variant synonyme p.(Ala547=)	2 VC	VarScan 2	49,00	702	2 VC	VarScan 2	45,10	204
				HaplotypeCaller	48,39	467		HaplotypeCaller	56,06	751
SLC2A1	NM_006516.2:c.45C>T chr1:43408966-G>A	variant synonyme p.(Ala15=)	2 VC	VarScan 2	48,88	1119	2 VC	VarScan 2	51,16	1640
				HaplotypeCaller	48,77	326		HaplotypeCaller	49,92	593
GRIN2B	NM_000834.3:c.366C>G chr12:14018777-G>C	variant synonyme p.(Pro122=)	2 VC	VarScan 2	43,94	66	2 VC	VarScan 2	47,88	1130
				HaplotypeCaller	43,75	64		HaplotypeCaller	48,15	243
PTCHD1	NM_173495.2:c.858T>C chrX:23398214-T>C	variant synonyme p.(Cys286=)	2 VC	VarScan 2	49,08	868	2 VC	VarScan 2	53,76	1622
				HaplotypeCaller	50,96	312		HaplotypeCaller	52,00	250
GRIN2A	NM_000833.4:c.1275G>A chr16:9943666-C>T	variant synonyme p.(Leu425=)	2 VC	VarScan 2			2 VC	VarScan 2	46,08	1504
				HaplotypeCaller				HaplotypeCaller	51,31	497

Figure 51 : Représentation sur IGV du variant de *GRIN2A* (NM_000833.4:c.1275G>A) retenu par le Filtre 2 avec 2 VC pour le test DI117, et présent dans une région de *Gap 30X* pour le test DI44



Au total et comme l'atteste le **Tableau 16**, les variants détectés dans le « Filtre 2 – 2 VC » pour ces 6 patients comportent un nombre plus élevé de probables faux-négatifs et faux-positifs avec le test DI44 comparé au test DI117. En effet, 6 variants probablement réels ne sont pas décelés avec le test DI44, contre 2 avec le test DI117. Malgré ce filtre, un variant faux-positif est retrouvé avec le test DI44.

Ces éléments tendent à démontrer que la performance diagnostique du test DI117 est supérieure à celle du test DI44 pour les 6 patients étudiés.

Tableau 16 : Tableau récapitulant les variants retenus par le « Filtre 2 - 2 VC » qui diffèrent entre les 2 tests, Test DI44 et Test DI117, et les hypothèses émises pour expliquer ces écarts

Variants retenus avec le "Filtre 2 - 2 VC" différents entre les 2 tests			Test DI44			Test DI117			Probable Faux-positif ou Faux-négatif ?	Hypothèses émises		
			1 ou 2 VC	VarScan 2 HaplotypeCaller	VAF	Profondeur de lecture	1 ou 2 VC	VarScan 2 HaplotypeCaller			VAF	Profondeur de lecture
ARID1B (Patient 1)	chr6:157099286-GC>G NM_020732.3:c.228delC	variant frameshift p.(Ala77Profs*34)	2 VC	VarScan 2 HaplotypeCaller	35,90 35,29	39 34				Faux-positif du test DI44	homopolymère	
GRIN2B (Patient 1)	chr12:13768586-G>A NM_000834.3:c.1341C>T	variant synonyme p.(Asp447=)					2 VC	VarScan 2 HaplotypeCaller	49,00 49,84	749 311	Faux-négatif du test DI44	absence de couverture
ANKRD11 (Patient 2)	chr16:89346163-G>A NM_001256182.1:c.6787C>T	variant faux-sens p.(Pro2263Ser)	2 VC	VarScan 2 HaplotypeCaller	45,45 45,71	44 35	2 VC	VarScan 2 HaplotypeCaller	33,09 40,52	136 116	Faux-négatif du test DI117	VAF faible
RAI1 (Patient 2)	chr17:17697093-CCAGCAGCAG>C NM_030665.3:c.864_872delGCAGCAGCA	variant frameshift p.(Gln289_Gln291del)	2 VC	VarScan 2 HaplotypeCaller	43,06 44,28	648 402	1 VC	VarScan 2	43,35	579	Faux-négatif du test DI117	région difficile à aligner
EP300 (Patient 3)	chr22:41545024-GT>G NM_001429.3:c.2242-4delT	variant d'épissage					2 VC	VarScan 2 HaplotypeCaller	52,60 61,28	597 390	Faux-négatif du test DI44	homopolymère
GRIN1 (Patient 3)	chr9:140052907-G>T NM_007327.3:c.1045G>T	variant faux-sens p.(Ala370Ser)	2 VC	VarScan 2 HaplotypeCaller	34,42 22,55	1014 510	2 VC	VarScan 2 HaplotypeCaller	54,69 52,96	916 389	Faux-négatif du test DI44	VAF faible
GRIN2A (Patient 4)	chr16:9943666-C>T NM_000833.4:c.1275G>A	variant synonyme p.(Leu425=)	2 VC	VarScan 2 HaplotypeCaller	60,00 63,16	20 19	2 VC	VarScan 2 HaplotypeCaller	45,91 46,44	1333 519	Faux-négatif du test DI44	Gap 30X
ANKRD11 (Patient 5)	chr16:89345822-G>C NM_001256182.1:c.7128C>G	variant faux-sens p.(Asp2376Glu)	2 VC	VarScan 2 HaplotypeCaller	38,35 32,67	156 150	2 VC	VarScan 2 HaplotypeCaller	48,41 46,22	157 119	Faux-négatif du test DI44	VAF faible
GRIN2A (Patient 6)	chr16:9943666-C>T NM_000833.4:c.1275G>A	variant synonyme p.(Leu425=)					2 VC	VarScan 2 HaplotypeCaller	46,08 51,31	1504 497	Faux-négatif du test DI44	absence de couverture

3. Coût, temps technique

Nous avons comparé le coût des réactifs et le temps technique nécessaires pour les tests DI117 et DI44 (**Annexe 11**, **Annexe 12**). Les séquenceurs Illumina sont réputés pour être plus onéreux à l'achat que les technologies Ion Torrent™. Cependant, il est utile de comparer le coût des tests par patient lors de leur mise en œuvre.

La technique d'enrichissement des bibliothèques par capture SureSelectQXT (*Agilent*) revêt un coût réactif un peu plus élevé que la technique d'enrichissement par amplification Ampliseq™ (*Thermo Fisher Scientific*). Toutefois, cet écart s'amenuise lorsque le nombre de patients testés en même temps augmente. Ainsi, pour un patient, le coût réactif calculé est de 202,47€ pour la technologie SureSelectQXT (*Agilent*) contre 137,15€ pour la technique Ampliseq™ (*Thermo Fisher Scientific*). Pour une série de 15 patients, le coût réactif calculé par patient diminue à 176,18€ pour le système SureSelectQXT (*Agilent*) et s'abaisse à 113,79€ pour le procédé Ampliseq™ (*Thermo Fisher Scientific*).

La technique d'enrichissement des bibliothèques par capture demande plus de temps technique que celle d'enrichissement par amplification, 8 à 10 heures estimées de temps de travail d'un technicien contre 5 à 8 heures.

Le coût de l'étape d'amplification de la bibliothèque associé à celui de l'étape de séquençage proprement dite est supérieure pour un patient avec le MiSeq Illumina (933,72€) comparé au PGM™ Ion Torrent™ (890,02€). Néanmoins, pour une série de 15 patients, le coût cumulé de ces 2 étapes par patient devient inférieur avec le MiSeq Illumina (62,25€) par rapport au PGM™ Ion Torrent™ (118,24€).

Le temps de travail d'un technicien estimé pour ces 2 étapes est plus faible avec le MiSeq Illumina qu'avec le PGM™ Ion Torrent™, 1 heure estimée de temps technique pour l'un contre 2 heures pour l'autre.

En matière de coûts, les données des tableaux **Annexe 11** et **Annexe 12** prouvent que les deux tests sont globalement du même ordre de prix. Le temps technique nécessaire diffère également peu entre les 2 tests. Le choix du test diagnostique mis en œuvre dans notre laboratoire n'est donc pas guidé par ces arguments économiques.

III. Résultats du test NGS DI117 appliqué à 15 patients atteints de DI

Afin de pouvoir évaluer le test NGS DI117 à des fins de diagnostic, nous avons constitué une cohorte de 16 patients atteints de DI (cf. partie Matériel et méthodes, paragraphe I, page 81). Leur diagnostic génétique est connu et confirmé, à l'exception du « patient 3 ». Ce diagnostic est décrit dans le **Tableau 2**. Sur ces 16 patients, un des échantillons n'avait pas la qualité requise pour être séquençé. Ainsi 15 patients de notre cohorte ont été séquençés avec notre test DI117, qui couvre 117 gènes impliqués dans la DI. Les régions codantes de ces 117 gènes ont été couvertes, avec une marge d'au minimum 10 paires de bases (inclusion des jonctions intron-exon). Cette couverture doit permettre en théorie de détecter la quasi-totalité des variations pathogènes de type SNV et petits indels. En effet, plus de 98% de ces mutations se situent dans les régions codantes des gènes ou dans les régions d'épissage localisées dans les 6 premières bases des jonctions intron-exon (62).

La librairie, cible du séquençage NGS, a été obtenue pour chaque patient par capture, et le séquençage effectué à l'aide de la technologie Illumina (cf. partie Introduction, paragraphe II.1.b.i.3, page 61).

Les librairies ont été réalisées en une seule fois, avec en parallèle un contrôle négatif (de l'eau). Le séquençage des 15 patients a été effectué sur un séquenceur MiSeq Illumina en 2 *runs* : un de 7 patients avec le contrôle négatif, et un de 8 patients. Un seul séquençage par patient a été réalisé. Aucune trace d'ADN n'a été identifiée dans le contrôle négatif, ce qui permet de s'assurer de l'absence de contamination inter-échantillon.

Nous avons évalué les caractéristiques suivantes de notre test DI117 pour les 15 patients séquençés :

- La profondeur de lecture du séquençage réalisé
- La couverture réelle des régions ciblées
- La détection effective des variants pathogènes déjà identifiés et confirmés pour chaque patient

Ces critères sont en effet cruciaux pour apprécier la sensibilité et la précision du test avant des applications à visée diagnostique.

1. Profondeur de lecture du séquençage réalisé

La profondeur de lecture moyenne (*PLM*) sur l'ensemble des régions cibles du test DI117 a été calculée pour chaque patient. Comme expliqué dans le paragraphe II.2.b (page 139), la *PLM* doit être au-dessus d'un seuil fixé par le laboratoire, pour notre laboratoire 200X, pour que les données issues du séquençage soient estimées de qualité suffisante pour être interprétables. Sa valeur est un reflet de l'enrichissement clonal de la librairie du patient.

Le **Tableau 17** atteste que la *PLM* est supérieure à 600X pour l'ensemble des patients.

La profondeur de lecture moyenne par gène et par patient n'est jamais inférieure à 200X, et ce pour aucun patient, comme le montrent la **Figure 52** et l'**Annexe 13**.

La profondeur de lecture moyenne par gène calculée pour l'ensemble des patients est comprise entre 300X et près de 1400X (**Figure 52, Annexe 13**).

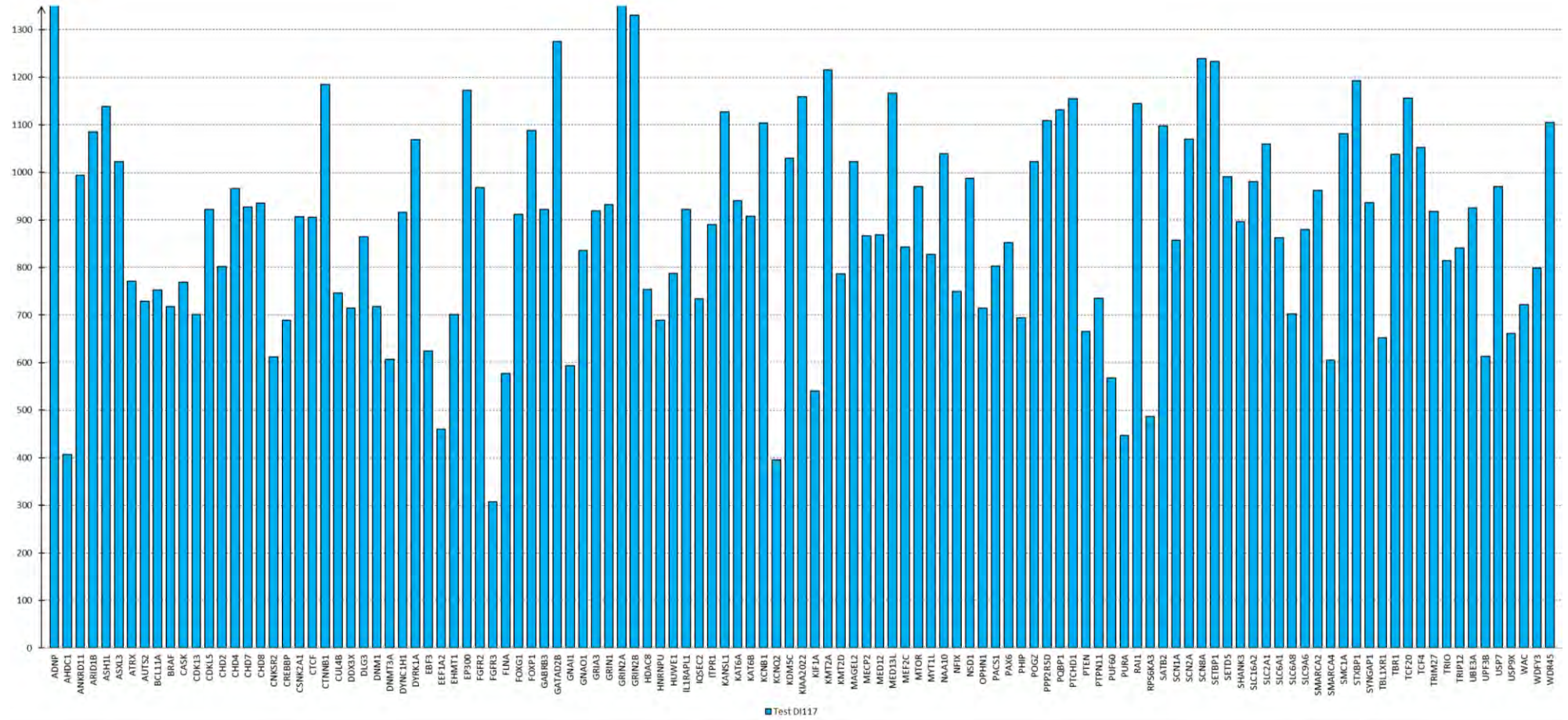
La profondeur de lecture d'un test NGS est étroitement liée à sa sensibilité diagnostique : plus la profondeur de lecture est élevée, plus le test sera à même de détecter un variant ayant une faible occurrence. La performance du test DI117 en ce qui concerne la *PLM* est plus que correcte, permettant de prédire une bonne sensibilité diagnostique.

Tableau 17 : Tableau récapitulatif de la *PLM* (profondeur de lecture moyenne sur les régions CDS \pm 25 de l'ensemble des 117 gènes) de chaque patient pour le Test DI117

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10
Moyenne (écart-type) pour les 117 gènes	885 (208)	759 (176)	643 (217)	807 (255)	755 (242)	1088 (255)	893 (211)	916 (217)	732 (236)

	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
Moyenne (écart-type) pour les 117 gènes	1790 (469)	754 (255)	825 (217)	979 (252)	735 (195)	877 (305)	831 (272)

Figure 52 : Profondeurs de lecture moyennes par gène du test DI117 pour les 117 gènes (CDS ± 25) (n = 15 patients)



2. Couverture réelle des régions ciblées (CDS+ /- 25)

Nous avons observé de façon globale le pourcentage de couverture des régions CDS ± 25 des 117 gènes à une profondeur de 30X et de 100X, ce qui a conduit à la réalisation des **Figure 53** et **Figure 54**.

Pour chaque patient, le pourcentage de couverture est supérieur à 99% à 30X, et au-delà de 96% à 100X.

Par conséquent, la couverture des régions CDS ± 25 des 117 gènes est adaptée aux recommandations de l'ANPGM (cf. paragraphe II.2.a.i, page 123), proche de 100% à 30X, et au-dessus de 95% à 100X.

Dans un deuxième temps, nous avons analysé les pourcentages de couverture avec une profondeur de lecture supérieure à 30X pour chaque gène et pour chaque patient. Il s'avère que les taux de couverture supérieure à 30X de chaque gène varient peu entre les différents patients (au maximum de quelques pourcents, et ce pour une minorité de gènes, cf. **Annexe 14**).

En étudiant la moyenne, par gène, des taux de couverture supérieure à 30X pour l'ensemble des 15 patients, le **Tableau 18** présente 68 gènes à 100%, 23 gènes entre 99 et 100%, 20 gènes entre 95 et 99%, et 6 gènes entre 90 et 95%. Nous ne recensons aucun gène ayant en moyenne un pourcentage de couverture à 30X inférieur à 90%.

Ces taux de couverture sur les régions CDS ± 25 sont convenables pour les trois-quarts des 117 gènes ciblés par le test NGS DI117, couverts à 99% minimum avec une profondeur de lecture de 30X, plus de la moitié étant couverte à 100%. Cependant 20 gènes sont couverts entre 95 et 99%, et 6 entre 90 et 95%. Ce n'est pas pleinement satisfaisant ; les *gaps 30X* associés, pouvant héberger une variation pathogène, seraient à séquencer avec la méthode Sanger si aucune mutation n'était retrouvée par ailleurs avec le test DI117.

Figure 53 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le 1^{er} run de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina

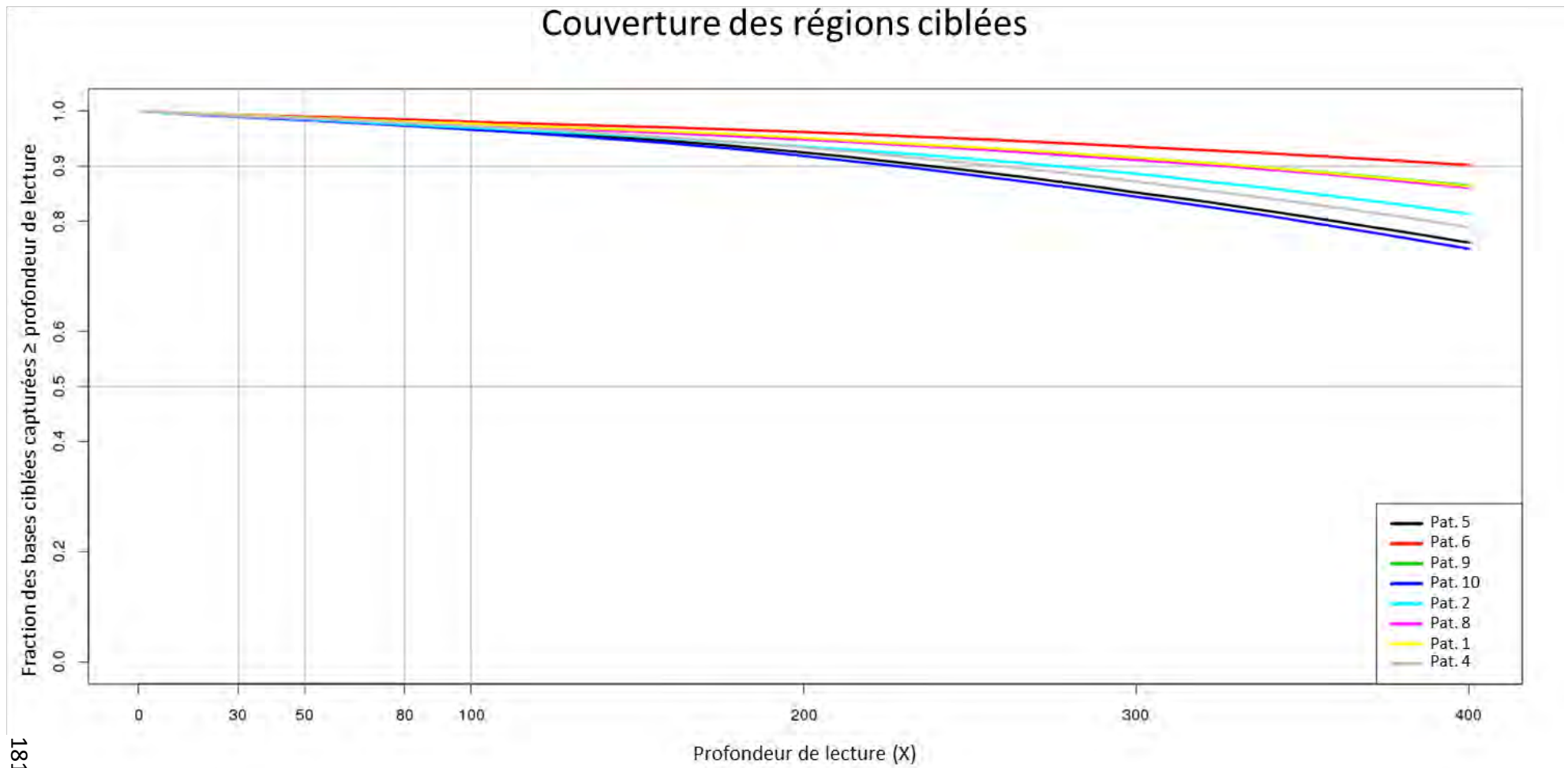


Figure 54 : Graphique représentant la fraction couverte des bases ciblées en fonction de la profondeur de lecture pour le 2^{ème} run de séquençage du Test DI117 réalisé sur le séquenceur MiSeq Illumina

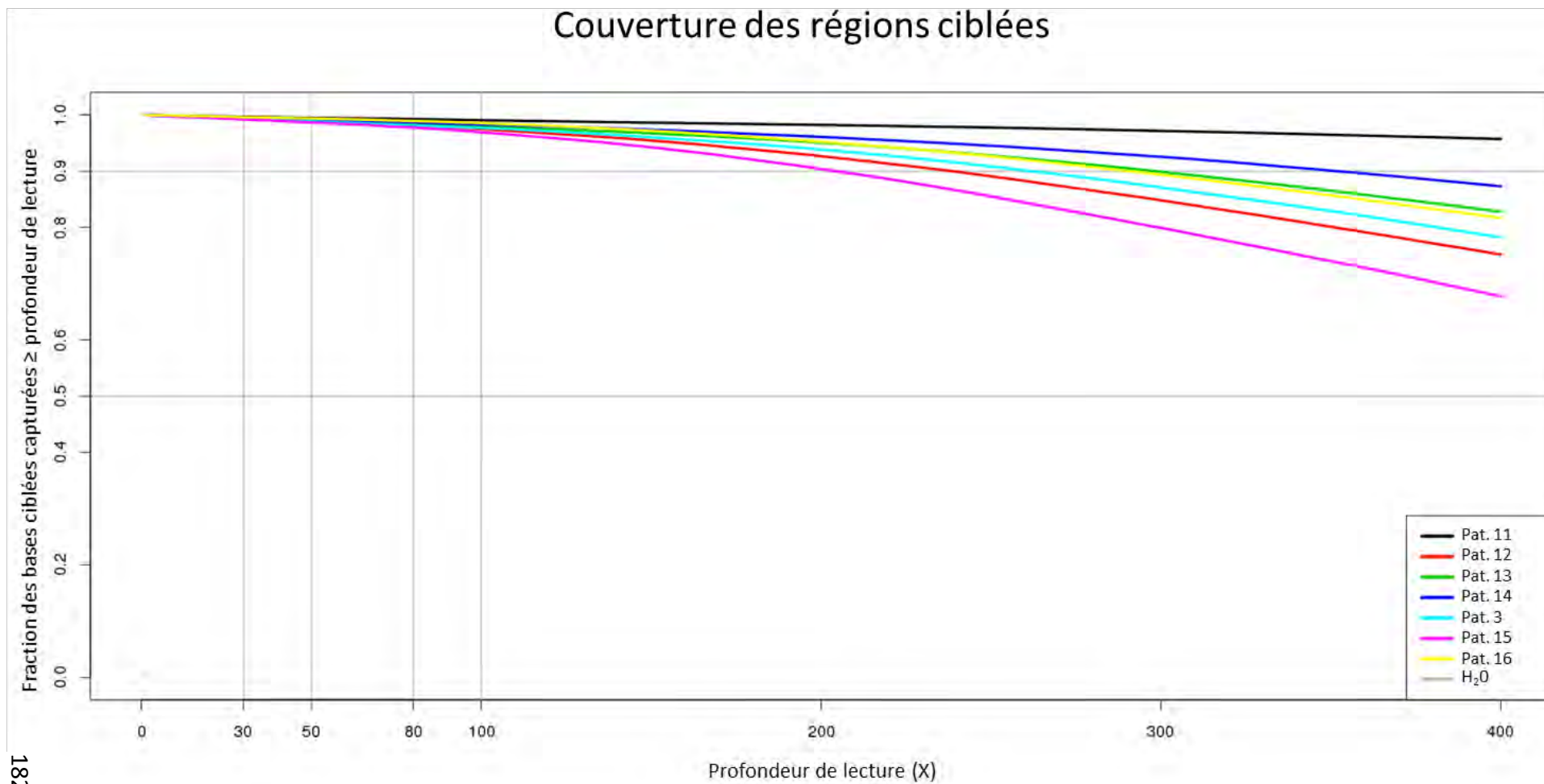


Tableau 18 : Tableau des moyennes des taux de couverture > 30X du Test DI117 pour les 117 gènes (CDS ± 25) (n = 15 patients)

Tableau des moyennes des taux de couverture >30X pour les 117 gènes (CDS ± 25)

Moyenne (écart-type) par gène des taux de couverture > 30X pour l'ensemble des 15 patients					
ADNP	100,00%	(0,00%)	TCF20	100,00%	(0,00%)
ANKRD11	100,00%	(0,00%)	TRIM27	100,00%	(0,00%)
ASH1L	100,00%	(0,00%)	TRIP12	100,00%	(0,00%)
ATRX	100,00%	(0,00%)	UBE3A	100,00%	(0,00%)
CASK	100,00%	(0,00%)	UPF3B	100,00%	(0,00%)
CDKL5	100,00%	(0,00%)	USP7	100,00%	(0,00%)
CHD2	100,00%	(0,00%)	USP9X	100,00%	(0,00%)
CHD4	100,00%	(0,00%)	WDFY3	100,00%	(0,00%)
CHD7	100,00%	(0,00%)	WDR45	100,00%	(0,00%)
CHD8	100,00%	(0,00%)	ARID1B	99,99%	(0,03%)
CSNK2A1	100,00%	(0,00%)	ASXL3	99,99%	(0,03%)
CTCF	100,00%	(0,00%)	TCF4	99,99%	(0,03%)
CTNNB1	100,00%	(0,00%)	PAX6	99,99%	(0,04%)
CUL4B	100,00%	(0,00%)	CNKSR2	99,98%	(0,08%)
DDX3X	100,00%	(0,00%)	SETBP1	99,98%	(0,08%)
DLG3	100,00%	(0,00%)	SLC16A2	99,98%	(0,08%)
DYRK1A	100,00%	(0,00%)	SLC6A8	99,98%	(0,08%)
EP300	100,00%	(0,00%)	PUF60	99,95%	(0,14%)
FGFR2	100,00%	(0,00%)	SLC9A6	99,95%	(0,10%)
FOXP1	100,00%	(0,00%)	MED12	99,85%	(0,28%)
GATAD2B	100,00%	(0,00%)	FLNA	99,83%	(0,25%)
GNAO1	100,00%	(0,00%)	KAT6A	99,79%	(0,27%)
GRIA3	100,00%	(0,00%)	AUTS2	99,77%	(0,18%)
GRIN1	100,00%	(0,00%)	KMT2A	99,77%	(0,28%)
GRIN2A	100,00%	(0,00%)	MYT1L	99,72%	(0,40%)
GRIN2B	100,00%	(0,00%)	SMARCA4	99,66%	(0,11%)
HDAC8	100,00%	(0,00%)	DYNC1H1	99,61%	(0,39%)
HUWE1	100,00%	(0,00%)	EBF3	99,54%	(0,75%)
IL1RAPL1	100,00%	(0,00%)	SMARCA2	99,53%	(0,57%)
ITPR1	100,00%	(0,00%)	IQSEC2	99,30%	(0,44%)
KANSL1	100,00%	(0,00%)	SHANK3	99,10%	(0,29%)
KAT6B	100,00%	(0,00%)	SYNGAP1	99,06%	(0,81%)
KDM5C	100,00%	(0,00%)	GNAI1	98,97%	(1,52%)
KIAA2022	100,00%	(0,00%)	AHDC1	98,79%	(0,68%)
KMT2D	100,00%	(0,00%)	PTEN	98,77%	(2,00%)
MAGEL2	100,00%	(0,00%)	HNRNPU	98,70%	(0,82%)
MECP2	100,00%	(0,00%)	EHMT1	98,29%	(0,36%)
MED13L	100,00%	(0,00%)	KCNQ2	98,21%	(1,22%)
MEF2C	100,00%	(0,00%)	CREBBP	97,79%	(0,32%)
MTOR	100,00%	(0,00%)	PHIP	97,63%	(0,26%)
NAA10	100,00%	(0,00%)	KIF1A	97,63%	(1,14%)
NSD1	100,00%	(0,00%)	KCNB1	97,51%	(0,71%)
OPHN1	100,00%	(0,00%)	PTPN11	97,50%	(0,00%)
POGZ	100,00%	(0,00%)	DNMT3A	97,01%	(0,80%)
PPP2R5D	100,00%	(0,00%)	PACS1	97,01%	(1,45%)
PQBP1	100,00%	(0,00%)	TRIO	96,90%	(1,09%)
PTCHD1	100,00%	(0,00%)	WAC	96,60%	(0,00%)
RAI1	100,00%	(0,00%)	BCL11A	96,10%	(0,00%)
SATB2	100,00%	(0,00%)	RPS6KA3	96,00%	(1,57%)
SCN1A	100,00%	(0,00%)	NFIX	95,70%	(0,00%)
SCN2A	100,00%	(0,00%)	GABRB3	95,33%	(0,91%)
SCN8A	100,00%	(0,00%)	FGFR3	95,24%	(1,52%)
SETD5	100,00%	(0,00%)	BRAF	94,57%	(1,23%)
SLC2A1	100,00%	(0,00%)	DNM1	92,70%	(0,00%)
SLC6A1	100,00%	(0,00%)	FOXP1	91,85%	(4,20%)
SMC1A	100,00%	(0,00%)	CDK13	91,03%	(4,28%)
STXBP1	100,00%	(0,00%)	PURA	90,60%	(5,15%)
TBL1XR1	100,00%	(0,00%)	EEF1A2	90,49%	(1,84%)
TBR1	100,00%	(0,00%)	Pour les 117 gènes	99,33%	(0,17%)

■ Gène inclus dans le test DI44

3. Détection des variations « vraies positives »

Il s'agit ici d'analyser la capacité du test DI117 à mettre en évidence des mutations de type SNV chez des patients témoins. Les variations de séquence de type CNV ne sont actuellement pas recherchées par le pipeline d'analyse bio-informatique du test DI117.

Sur les 15 patients étudiés, 13 sont des sujets témoins porteurs d'une mutation de type SNV identifiée comme responsable de leur DI, un patient témoin possède deux variations pathogènes de type CNV, et un seul n'a pas de diagnostic génétique moléculaire confirmé. Les variations pathogènes de type SNV de ces patients ont été confirmées par un séquençage Sanger dans un LBM français.

Pour tous les patients ayant une mutation de type SNV (13 patients sur 15), les variations de séquence pathogènes sont retrouvées par les 2 *Variant Caller* (VC) avec une profondeur supérieure à 100X. La fréquence allélique du variant (VAF) calculée est cohérente avec le diagnostic génétique posé au préalable : une présence à l'état hétérozygote pour 12 des 13 patients, et une mosaïque* avec une fréquence allélique de 30% pour le patient 11 (**Tableau 19**).

Les variations non causales détectées auparavant par séquençage Sanger ont toutes été retrouvées, comme le montre le **Tableau 19**.

Par ailleurs, le gène MECP2 fait l'objet d'un séquençage par la méthode de Sanger au CHU de Toulouse. Ainsi, 6 patients de notre cohorte (patients 2, 6, 11, 12, 13 et 14) ont précédemment bénéficié d'un séquençage Sanger pour ce gène. Nous avons comparé ces résultats à ceux obtenus par le test DI117. Aucun variant n'a été détecté dans le gène MECP2 chez les 4 patients (patients 2, 12, 13 et 14) pour qui le séquençage Sanger de ce gène révélait une séquence identique à la séquence de référence. Aucun variant supplémentaire dans le gène MECP2 n'a été décelé pour le patient 6 en dehors du variant pathogène confirmé ; cela coïncide avec le résultat du séquençage Sanger de ce gène. Ces conclusions identiques entre les résultats de notre test DI117 et ceux du séquençage Sanger pour le gène MECP2 témoignent d'une bonne spécificité de notre test DI117 concernant ce gène.

La profondeur de lecture du test NGS DI117 est au-dessus des seuils fixés par notre LBM. Le pourcentage de couverture à 30X est acceptable pour les trois-quarts des 117 gènes, cependant il est inférieur à 99% pour 26 d'entre eux. Cela impliquerait des analyses Sanger complémentaires des gaps 30X en cas d'absence de mutation détectée par le test DI117.

Le test DI117 mis en œuvre a montré une fiabilité sur la détection des mutations de type SNV des gènes étudiés. Le nombre de gènes étudiés est toutefois limité, avec 13 patients porteurs de mutation SNV, dont certaines sur les mêmes gènes.

Tableau 19 : Analyse du variant confirmé causal de chaque patient et des variants non causaux détectés en séquençage Sanger

Test DI117

Analyse du variant confirmé en séquençage Sanger causal de chaque patient

Patients	Variants	1 ou 2 VC	VAF	Profondeur de lecture
Patient 1	ANKRD11 mutation faux-sens NM_001256182.1:c.7535G>C p.(Arg2512Pro)	2 VC	VarScan 2 43,38	816
			HaplotypeCaller 45,27	296
Patient 2	CDKL5 mutation faux-sens NM_003159.2:c.719G>C p.(Ser240Thr)	2 VC	VarScan 2 50,71	777
			HaplotypeCaller 50,00	288
Patient 3	/	/	/	/
Patient 4	RAI1 mutation frameshift NM_030665.3:c.1958_1959delTG p.(Val653Glyfs35*)	2 VC	VarScan 2 45,43	1358
			HaplotypeCaller 47,03	202
Patient 5	SCN8A mutation faux-sens NM_014191.3:c.3943G>C p.(Val1315Leu)	2 VC	VarScan 2 49,75	810
			HaplotypeCaller 48,67	339
Patient 6	MECP2 mutation frameshift NM_004992.3:c.806delG p.(Gly269Alafs*19)	2 VC	VarScan 2 44,19	1550
			HaplotypeCaller 50,75	266
Patient 8	ANKRD11 mutation frameshift NM_001256182.1:c.3066del p.(Thr1023Glnfs*295)	2 VC	VarScan 2 46,37	785
			HaplotypeCaller 46,06	241
Patient 9	ARID1B mutation frameshift NM_020732.3:c.4275T>A p.(Tyr1425*)	2 VC	VarScan 2 47,42	814
			HaplotypeCaller 41,84	239
Patient 10	EEF1A2 mutation faux-sens NM_001958.3:c.370G>A p.(Glu124Lys)	2 VC	VarScan 2 53,35	433
			HaplotypeCaller 52,21	249
Patient 11	MAGEL2 délétion en phase en mosaïque NM_019066.4:c.461_481del p.(Pro154_Ala160del)	2 VC	VarScan 2 29,85	1521
			HaplotypeCaller 40,07	292
Patient 12	SCN2A mutation faux-sens NM_021007.2:c.1199C>G p.(Thr400Arg)	2 VC	VarScan 2 47,53	749
			HaplotypeCaller 40,91	242
Patient 13	SYNGAP1 mutation frameshift NM_006772.2:c.1746delG p.(Asp583Thrfs*67)	2 VC	VarScan 2 52,57	972
			HaplotypeCaller 50,25	398
Patient 14	IQSEC2 mutation frameshift NM_001111125.2:c.2776C>T p.(Arg926*)	2 VC	VarScan 2 45,49	1308
			HaplotypeCaller 46,93	456
Patient 15	MECP2 CNVs réarrangement complexe (délétion associée à une duplication)	non analysé		
Patient 16	GRIN1 mutation faux-sens NM_007327.3:c.2452A>C p.(Met818Leu)	2 VC	VarScan 2 48,26	1467
			HaplotypeCaller 52,71	277

Analyse des variants non causaux détectés en séquençage Sanger

Patients	Variants	1 ou 2 VC	VAF	Profondeur de lecture
Patient 10	SLC2A1 variant en 5'UTR hérité NM_006516.2:c.-57C>A	2 VC	VarScan 2 52,17	621
			HaplotypeCaller 52,04	613
Patient 10	SLC2A1 variant intronique hérité NM_006516.2:c.972+39G>A	2 VC	VarScan 2 47,05	1016
			HaplotypeCaller 52,43	267
Patient 11	MAGEL2 SNP NM_019066.4:c.2886C>T p.(Ser962Ser)	2 VC	VarScan 2 49,86	1835
			HaplotypeCaller 51,41	177

IV. Analyse comparative des résultats du test DI117 face à d'autres tests NGS dédiés à la DI

Au sein de notre cohorte, 5 patients avaient déjà fait l'objet d'un séquençage haut débit réalisé sur un panel de gènes dédié à la DI dans deux autres laboratoires français de génétique moléculaire : celui du CHU de Marseille et celui du CHU de Strasbourg. Les données transmises par ces 2 laboratoires ont été utilisées à des fins de comparaison avec celles obtenues avec notre test DI117 (comparaison portant uniquement sur les gènes communs).

Le laboratoire de Strasbourg a mis en place un test de 456 gènes (TC-456), actualisé récemment en un test de 520 gènes (TC-520). Les tests TC-456 et TC-520 disposent respectivement de 92 et de 110 gènes en commun avec notre test DI117. Le CHU de Marseille réalise un test de 142 gènes (DI-142) qui possède 73 gènes en commun avec le DI117. Les gènes communs à l'ensemble des 4 tests (TC-456, TC-520, DI-142 et DI117) sont donc au nombre de 67 (**Annexe 1**).

Dans notre test, nous avons sélectionné les gènes les plus fréquemment impliqués dans la DI en France en y ajoutant des gènes du syndrome de Prader-Willi like (cf. Résultats I.2, page 106). Les différences dans les gènes sélectionnés entre les quatre tests s'expliquent par des écarts de prévalence constatée au moment de la mise en place des tests, mais aussi par d'autres choix propres aux différents laboratoires.

Les résultats du chapitre précédent ont pu montrer chez 13 patients la bonne sensibilité de notre test DI117 pour détecter les variations de séquence pathogènes de type SNV confirmées par la technique de séquençage Sanger.

La comparaison avec les autres tests de DI a comme objectif de prouver que notre test NGS DI117 est aussi juste que ces autres tests pour l'observation de l'ensemble des variations de séquences (mutations causales et polymorphismes*) au sein des 117 gènes de notre panel. Des hypothèses seront émises pour justifier les éventuels écarts observés. En étudiant si le test DI117 peut aussi détecter les variants non pathogènes, nous pourrions ainsi supputer les capacités du test DI117 à distinguer des mutations dans les

régions qu'il cible, même si ces mutations ne sont pas présentes dans notre cohorte de patients.

Parmi les 5 patients déjà soumis à un séquençage dans un autre centre, les tests et technologies utilisés sont les suivants (**Tableau 20**) :

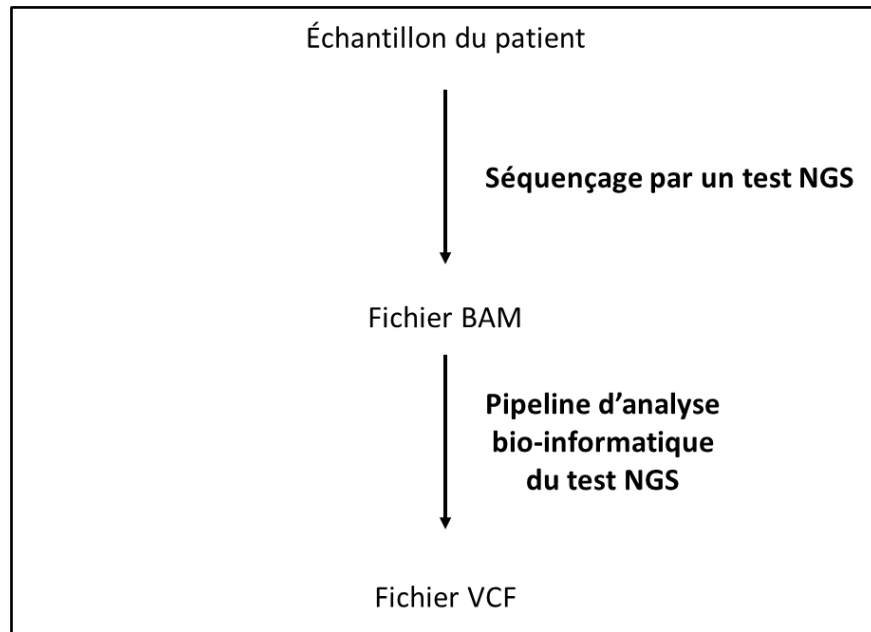
- 3 patients ont été séquencés via le test de 142 gènes du CHU de Marseille (DI-142). La librairie de ce test est obtenue par capture, pour un séquençage sur technologie Ion Torrent (patients 1, 3, et 9).
- 1 patient a été séquencé via le test de 456 gènes (TC-456) et 1 patient via le test de 520 gènes (TC-520) du CHU de Strasbourg. La librairie de ces tests est obtenue par capture pour un séquençage sur technologie Illumina (patients 4 et 10).

Tableau 20 : Tableau récapitulant les caractéristiques techniques des différents tests NGS réalisés chez les patients étudiés

Différents tests NGS de DI	Méthode d'enrichissement des librairies	Technologie de séquençage	Patients séquencés
Test TC-456 (CHU de Strasbourg)	Capture de séquence SureSelect XT2 (<i>Agilent</i>)	HiSeq, Illumina	Patient 4
Test TC-520 (CHU de Strasbourg)	Capture de séquence Seq Cap Ez (<i>Roche</i>)	NextSeq550, Illumina	Patient 10
Test DI-142 (CHU de Marseille)	Capture de séquence SureSelect XT Inherited Diseases Panel (<i>Agilent</i>)	Ion Proton, Thermo Fisher Scientific	Patients 1, 3 et 9

Comme présenté dans la partie introduction, paragraphe II.2.d (page 79), le séquençage haut débit génère des données de séquençage sous la forme de fichiers BAM. Ces fichiers BAM sont traités par un pipeline d'analyse bio-informatique dépendant de la technologie mise en œuvre et propre à chaque test. Le passage du fichier BAM dans un pipeline d'analyse permet de générer un fichier contenant les variations de séquence trouvées (fichier VCF) (**Figure 55**).

Figure 55 : Schéma représentant le moment d'obtention du fichier BAM et du fichier VCF lors d'un test NGS



Les CHU de Marseille et de Strasbourg nous ont transmis leurs données de séquençage secondaires (fichiers BAM et VCF), pour les 5 patients séquencés avec leurs tests. Nous disposons par conséquent de données extérieures suffisantes pour procéder à une comparaison minutieuse avec les données bio-informatiques de séquençage et le pipeline associé à notre propre test DI117.

Les écarts dans les variants retrouvés par chaque test peuvent s'expliquer par la présence de faux-positifs et de faux-négatifs propres à chaque test ou à leur pipeline bio-informatique. En les étudiant, il est ainsi possible de mieux qualifier la sensibilité et la spécificité du test NGS DI117.

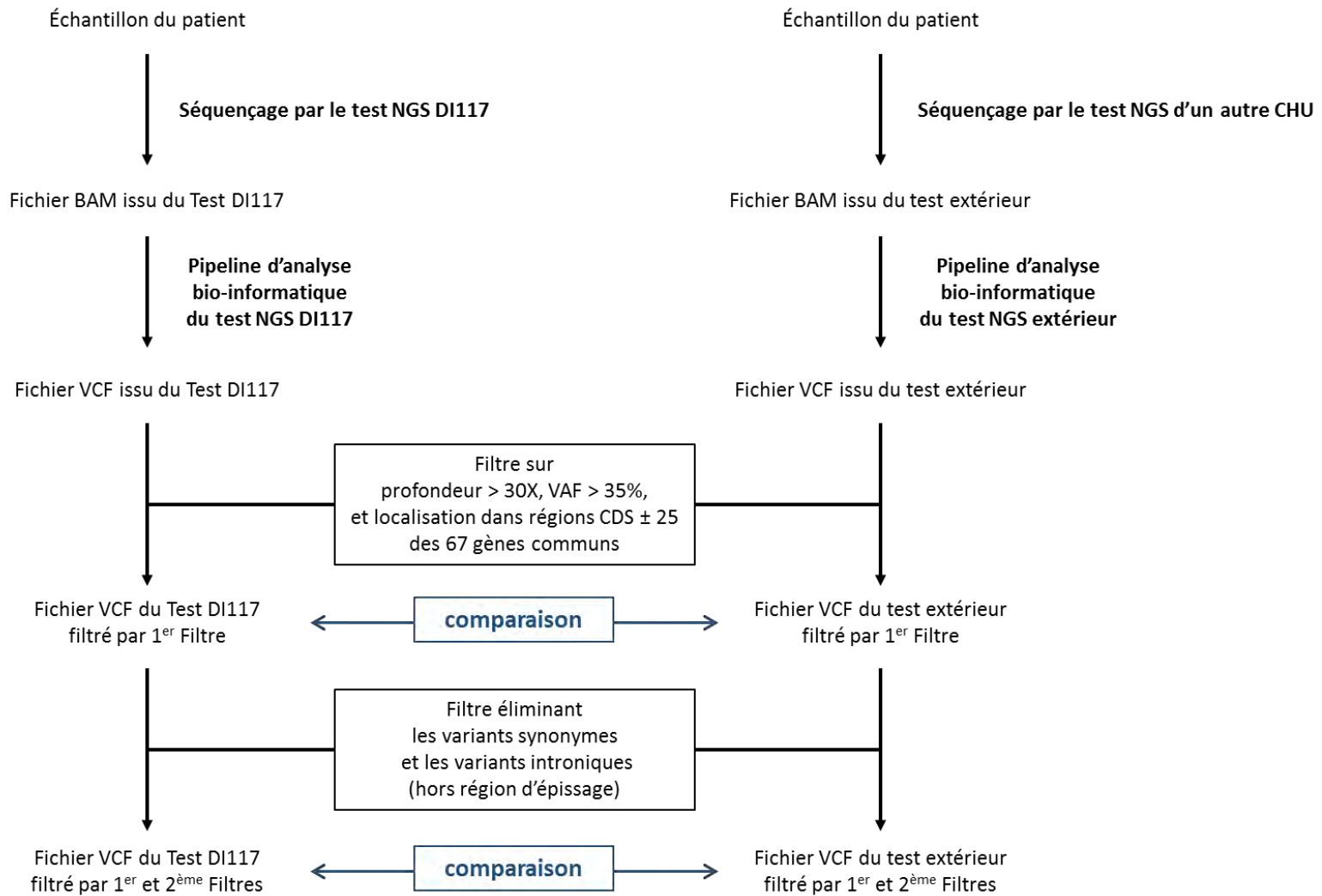
1. Comparaison des variants détectés par le test DI117 avec ceux détectés par les autres tests

Pour chaque patient, nous avons souhaité confronter les variations de séquence mises en évidence par notre test DI117 à celles détectées par le test réalisé dans un autre CHU.

Nous avons pour cela comparé les fichiers VCF issus de notre test DI117 à ceux générés par les tests dédiés à la DI des autres CHU.

Selon les recommandations de l'ANPGM, une profondeur minimale de 30X est conseillée pour l'analyse des variants SNV ; c'est le seuil fixé dans cette étude. Un seuil minimum de 35% pour la VAF (valeur de fréquence allélique) a été fixé au laboratoire de Toulouse pour filtrer les variants lors de l'analyse des résultats des patients. Ce critère permet d'éliminer une grande partie des faux-positifs dus à une erreur de séquençage. Le filtrage bio-informatique avec un seuil de profondeur de lecture à 30X et un seuil de VAF à 35% a été appliqué pour tous les fichiers VCF (**Figure 56**).

Figure 56 : Schéma expliquant les analyses et les filtrages réalisés lors de la comparaison des variants détectés par le test DI117 avec ceux décelés par les autres tests NGS de DI



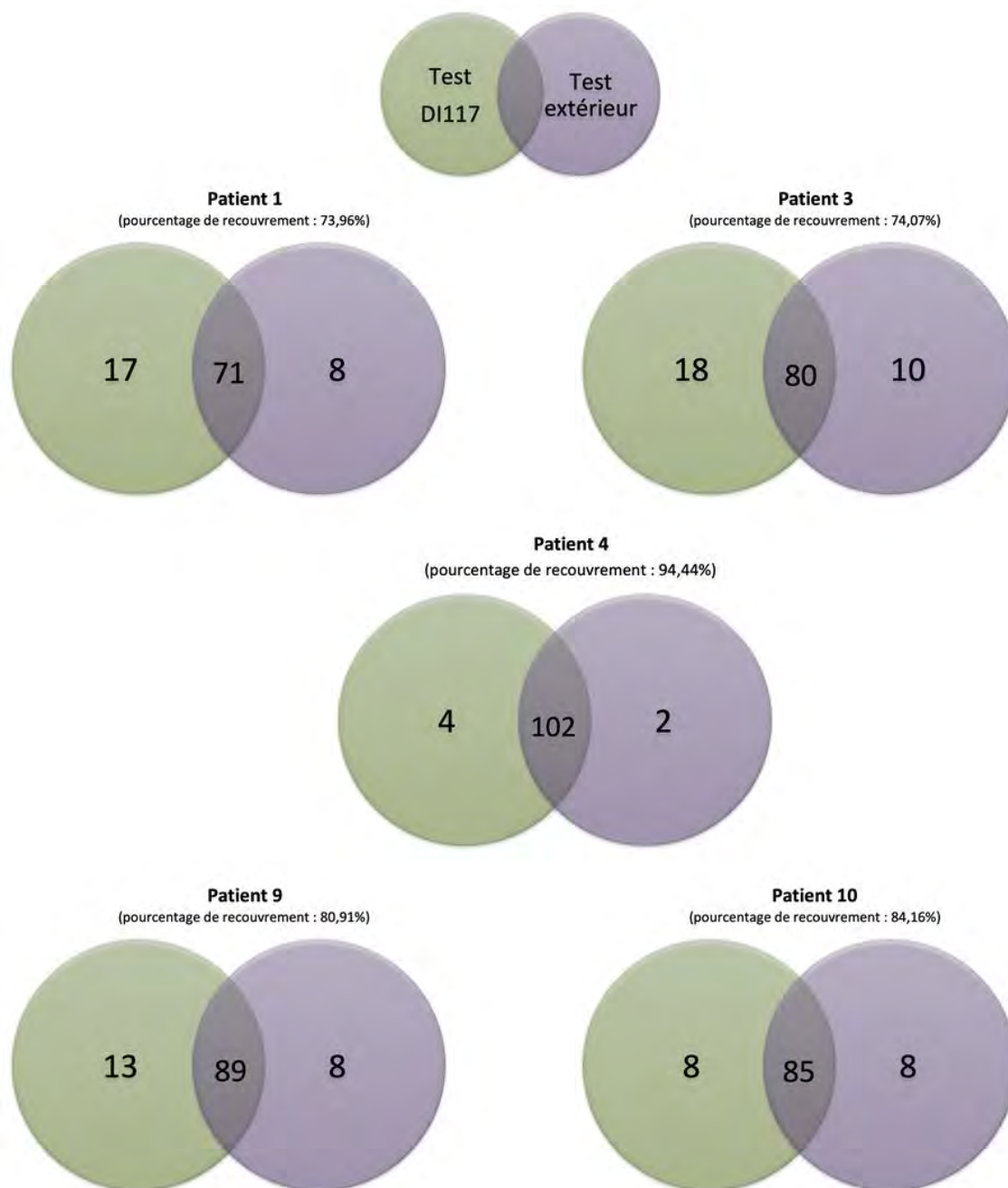
Nous avons concentré notre analyse sur la recherche de variations ayant une localisation dans la région cible CDS ± 25 des 67 gènes communs aux 4 tests avec les seuils de filtrage décrits (**Figure 56**).

Les données ainsi collectées montrent des variations totales détectées par patient pour l'ensemble des régions ciblées des 67 gènes communs ayant une profondeur minimale de 30X et une VAF d'au moins 35% (**Figure 57**).

- Patient 1 : 71 variants en commun, 8 variants en plus avec le test DI-142 du CHU de Marseille, et 17 en plus avec le test DI117 au CHU de Toulouse (73,96% de variations filtrées identiques)
- Patient 3 : 80 variants en commun, 10 variants en plus avec le test DI-142 du CHU de Marseille, et 18 en plus avec le test DI117 au CHU de Toulouse (74,01% de variations filtrées identiques)
- Patient 4 : 102 variants en commun, 2 en plus avec le test TC-456 du CHU de Strasbourg, et 4 en plus avec le test DI117 au CHU de Toulouse (94,44% de variations filtrées identiques)
- Patient 9 : 89 variants en commun, 8 variants en plus avec le test DI-142 du CHU de Marseille, et 13 en plus avec le test DI117 au CHU de Toulouse (80,91% de variations filtrées identiques)
- Patient 10 : 85 variants en commun, 8 variants en plus avec le test TC-520 du CHU de Strasbourg, et 8 en plus avec le test DI117 au CHU de Toulouse (84,16% de variations filtrées identiques)

La majorité des variations détectées sont communes aux tests extérieurs et au test DI117, avec un minimum de 73,96% de variations détectées identiques. Les tests du CHU de Strasbourg, utilisant une technologie analogue à celle utilisée dans le cadre du test DI117 du CHU de Toulouse, sont ceux dont les résultats sont les plus proches (94,44% et 84,16% de variations identiques sur deux patients). D'autres hypothèses peuvent expliquer les différences entre ces résultats. Par exemple, il est possible que les marges autour des régions ciblées soient différentes sur certains gènes entre notre test DI117 et le test DI-142 du CHU de Marseille, ce qui justifierait une partie des écarts constatés.

Figure 57 : Diagrammes de Venn représentant la comparaison entre les variants détectés par le test DI117 et ceux décelés par le test extérieur (selon les filtrages décrits)



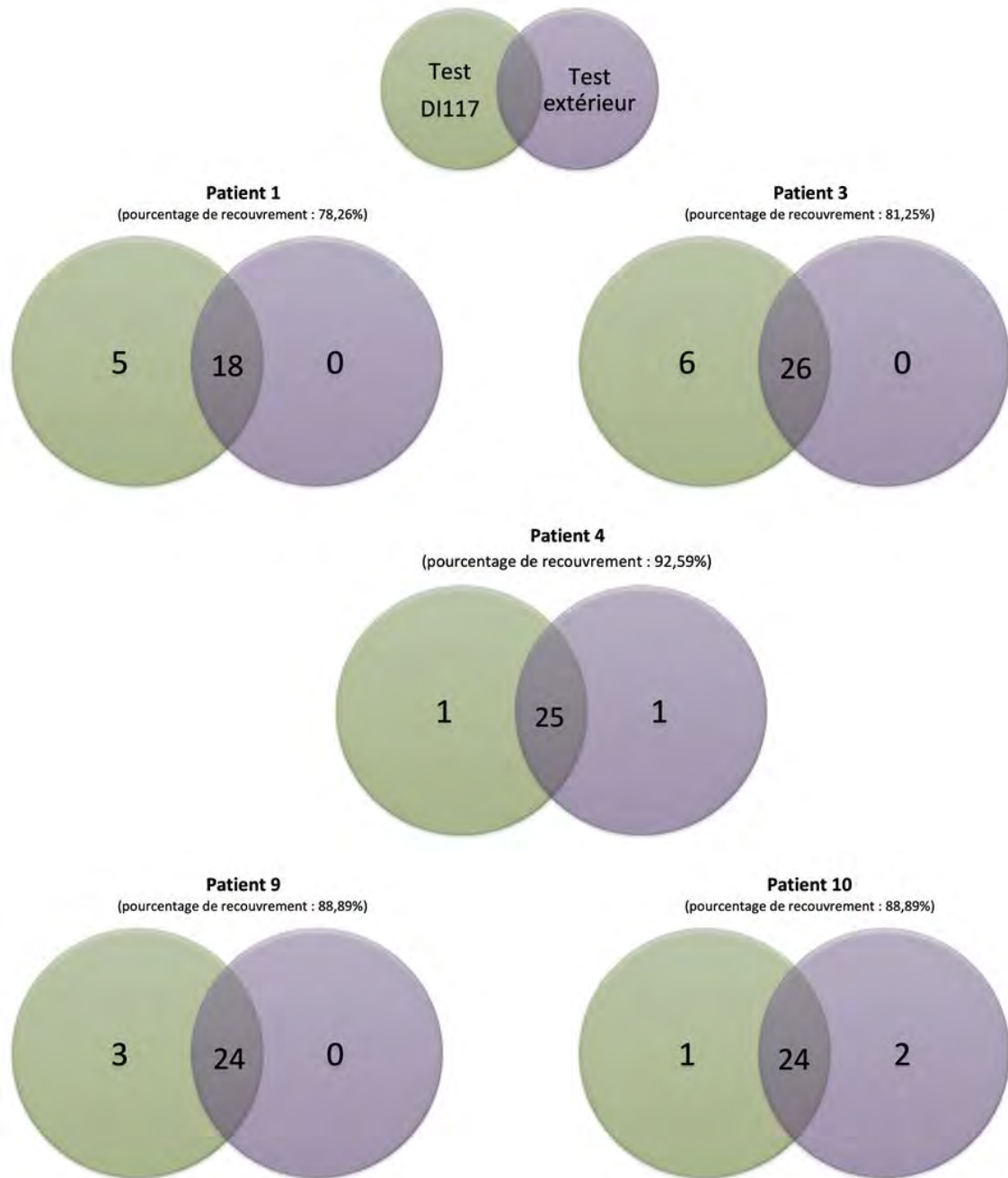
Pour analyser les écarts dans les variants découverts par les différents tests, l'étude peut être restreinte à un sous-ensemble excluant les variants synonymes et les variants introniques à l'exception des régions d'épissage (**Figure 56**). Ces variants filtrés n'étant pas pathogènes en général, leur intérêt médical est moindre et il est donc possible de les éliminer de cette analyse plus poussée.

Les données des tests peuvent ainsi indiquer par patient les variants non synonymes situés dans les régions codantes et les variants localisés dans les régions d'épissage, retrouvés au sein des 67 gènes en commun et avec les filtrages précédemment décrits : une profondeur de lecture > 30X et une VAF minimum de 35% (**Figure 58**).

- Patient 1 : 18 variants en commun, aucun en plus avec le test DI-142 du CHU de Marseille, et 5 en plus avec le test DI117 au CHU de Toulouse (78,26% de variations filtrées identiques)
- Patient 3 : 26 variants en commun, aucun en plus avec le test DI-142 du CHU de Marseille, et 6 en plus avec le test DI117 au CHU de Toulouse (81,25% de variations filtrées identiques)
- Patient 4 : 25 variants en commun, 1 en plus avec le test TC-456 du CHU de Strasbourg, et 1 en plus avec le test DI117 au CHU de Toulouse (92,59% de variations filtrées identiques)
- Patient 9 : 24 variants en commun, aucun en plus avec le test DI-142 du CHU de Marseille, et 3 en plus avec le test DI117 au CHU de Toulouse (88,89% de variations filtrées identiques)
- Patient 10 : 24 variants en commun, 2 en plus avec le test TC-520 du CHU de Strasbourg, et 1 en plus avec le test DI117 au CHU de Toulouse (88,89% de variations filtrées identiques)

Pour ce sous-ensemble, la majorité des variants détectés sont retrouvés dans les tests extérieurs comme dans le test DI117. Le taux minimal de variations de séquences identiques est de 78,26% pour un patient. Les données ayant été restreintes à des régions d'intérêt médical avéré, il est légitime de s'intéresser aux motifs de l'absence de détection de variants dans un test alors qu'ils sont détectés par un autre. Nous souhaitons trouver une explication légitime pour tous les variants distincts entre le test DI117 et les tests des autres CHU. Pour approfondir la question, le pipeline d'analyse bio-informatique que nous avons utilisé peut être mis à l'épreuve.

Figure 58 : Diagrammes de Venn représentant la comparaison entre les variants non synonymes situés dans les régions codantes et les variants localisés dans les régions d'épissage détectés par le test DI117 et ceux décelés par le test extérieur (selon les filtrages décrits)



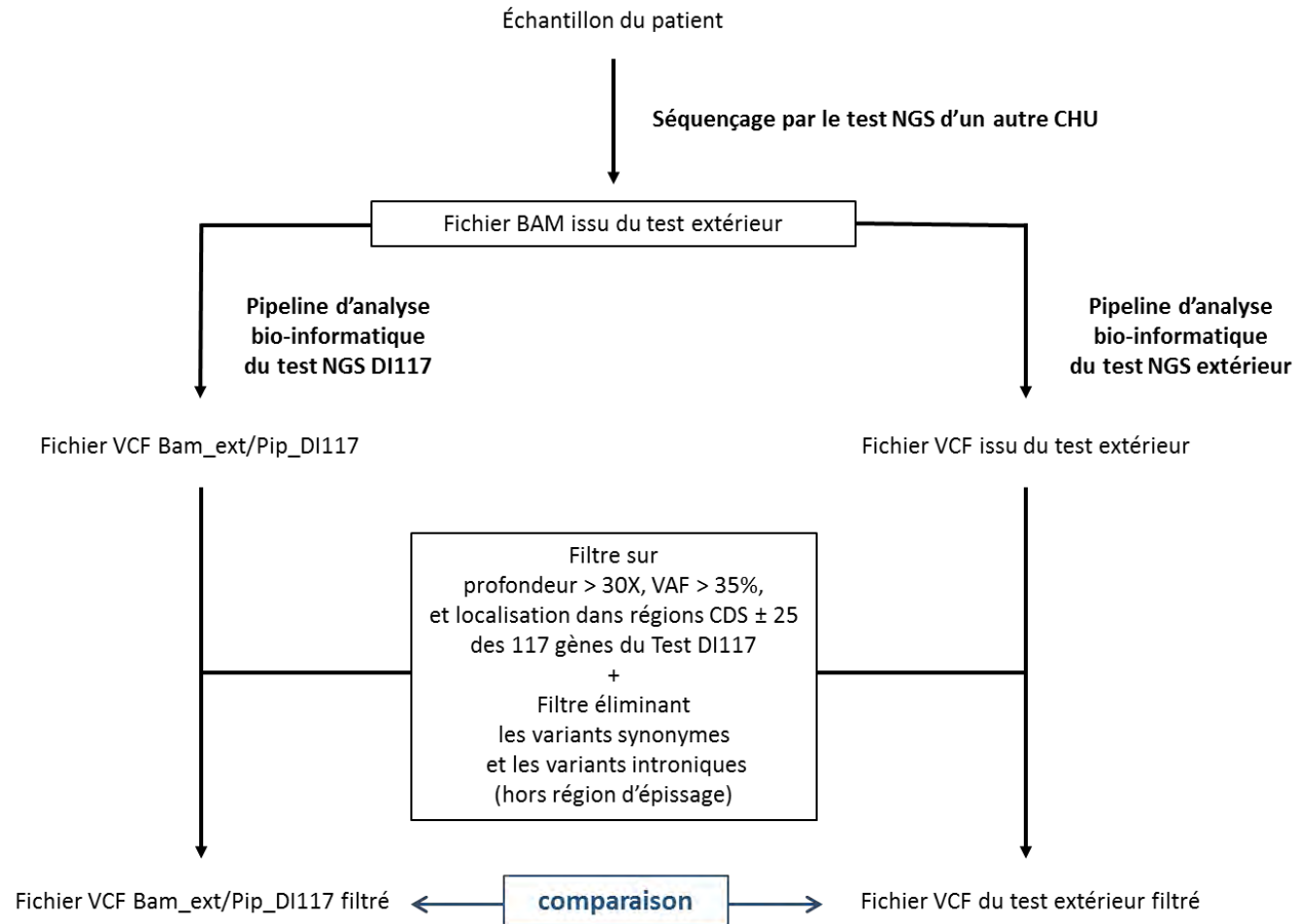
2. Comparaison des variants détectés avec le pipeline du test DI117 par rapport au pipeline des autres tests

Pour comprendre les différences de détection entre les tests, nous avons voulu tester notre pipeline d'analyse bio-informatique mis en œuvre dans le cadre du test DI117. Chaque pipeline bio-informatique est dépendant de nombreux éléments qui peuvent affecter le nombre de variants trouvés avec celui-ci : la méthode de séquençage, le ou les *Variant Caller* utilisés, les variants synonymes connus par ce pipeline, etc.

Notre pipeline a ainsi été mis en œuvre sur les données de séquençage (fichiers BAM) récupérées auprès des autres CHU, générées à l'issue du séquençage de leurs tests. Cette analyse a été réalisée avec les 117 gènes de notre test DI117 (**Figure 59**). Cela permet de comparer l'utilisation de notre pipeline pour le plus grand nombre de gènes communs possibles entre le test NGS DI117 et chacun des autres tests : respectivement 92, 110 et 73 gènes pour les tests TC-456, TC-520 et DI-142.

On obtient par ce biais un nouveau fichier VCF par patient, qui a été confronté au fichier VCF transmis par chacun des CHU (**Figure 59**). Les données finales ont été filtrées avec un seuil de profondeur de lecture à 30X et un seuil de VAF à 35%. Le filtre a conservé uniquement les variants non synonymes des régions codantes et les variants des régions d'épissage.

Figure 59 : Schéma expliquant les analyses et les filtrages réalisés lors de la comparaison des variants détectés avec le pipeline du test DI117 par rapport à ceux décelés par le pipeline des autres tests



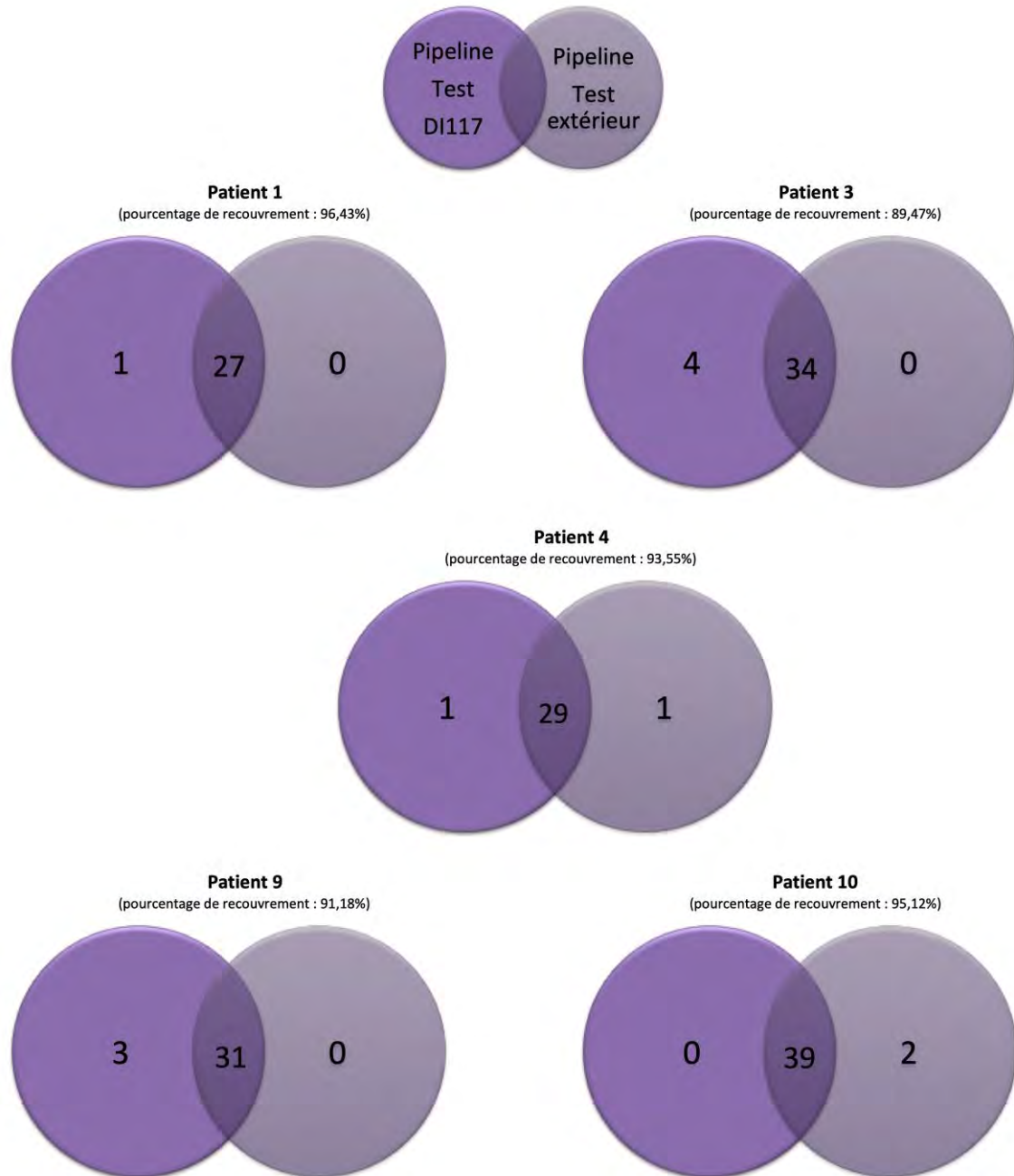
La comparaison entre les variants décelés avec le filtrage présenté pour les différents pipelines est ainsi possible (**Figure 60**) :

- Patient 1 : 27 variants en commun, aucun en plus avec le pipeline du test DI-142 du CHU de Marseille, et 1 en plus avec le pipeline du test DI117 au CHU de Toulouse (96,43% de variations filtrées identiques)
- Patient 3 : 34 variants en commun, aucun en plus avec le pipeline du test DI-142 du CHU de Marseille, 4 en plus avec le pipeline du test DI117 au CHU de Toulouse (89,47% de variations filtrées identiques)
- Patient 4 : 29 variants en commun, 1 en plus avec le pipeline du test TC-520 du CHU de Strasbourg, et 1 avec le pipeline du test DI117 au CHU de Toulouse (93,55% de variations filtrées identiques)
- Patient 9 : 31 variants en commun, aucun en plus avec le pipeline du test DI-142 du CHU de Marseille, et 3 en plus avec le pipeline du test DI117 au CHU de Toulouse (91,18% de variations filtrées identiques)
- Patient 10 : 39 variants en commun, 2 en plus avec le pipeline du test TC-456 du CHU de Strasbourg, et aucun en plus avec le pipeline du test DI117 au CHU de Toulouse (95,12% de variations filtrées identiques)

La plupart des variants est retrouvée à la fois par le pipeline du test DI117 et par le pipeline du test extérieur, avec un recouvrement de 89,47% variants identiques au minimum. Toutefois, en partant du même fichier BAM en entrée, certains variants sont détectés uniquement par le pipeline du test DI117 du CHU de Toulouse, et d'autres sont décelés uniquement par les pipelines des tests extérieurs. Par conséquent, il est possible de considérer qu'une partie des écarts observés dans la partie précédente est causée par les pipelines des tests.

Avec l'apport de ces données relatives aux pipelines de tests, il est possible d'expliquer la quasi-totalité des écarts de détection de variants entre les tests extérieurs et notre test DI117.

Figure 60 : Diagrammes de Venn représentant la comparaison entre les variants non synonymes situés dans les régions codantes et les variants localisés dans les régions d'épissage détectés par le pipeline du test DI117 et ceux décelés par le pipeline du test extérieur (selon les filtrages décrits)



3. Hypothèses expliquant la différence de variants observés entre le test DI117 et les tests des autres CHU

Grâce aux éléments obtenus dans les paragraphes précédents, nous pouvons décrire les écarts entre les variants détectés par le test DI117 du CHU de Toulouse et par les tests des autres CHU selon 4 situations détaillées ci-dessous :

- Cas d'un variant :
 - Détecté par un test extérieur,
 - Non détecté par le test DI117,
 - Non détecté par l'utilisation du pipeline du test DI117 sur les données de séquençage (fichier BAM) générées dans le cadre du test extérieur.

Ce cas montre une absence de détection par le pipeline du test DI117. Il peut également s'agir d'un faux-positif lié au séquençage réalisé dans le cadre du test extérieur.

- Cas d'un variant :
 - Non détecté par un test extérieur,
 - Détecté par le test DI117,
 - Détecté par l'utilisation du pipeline du test DI117 sur les données de séquençage (fichier BAM) générées dans le cadre du test extérieur.

Ce cas montre une absence de détection par le pipeline du test extérieur. Cette absence peut être expliquée si le pipeline comporte un filtrage différent des données par rapport au pipeline du test DI117. Par exemple, il peut être choisi de filtrer les variants bénins fréquents dans la population générale (polymorphismes) ou les variants dus à une erreur de séquençage (faux-positifs).

- Cas d'un variant :
 - Détecté par un test extérieur,
 - Non détecté par le test DI117,
 - Détecté par l'utilisation du pipeline du test DI117 sur les données de séquençage (fichier BAM) générées dans le cadre du test extérieur.

Ce cas montrerait une absence d'observation du variant avec le séquençage réalisé dans le cadre du test DI117. Aucun variant ne se trouve dans cette situation. Cela tend à

démontrer que le séquençage mis en œuvre dans le cadre du test DI117 est exhaustif sur les régions ciblées.

- Cas d'un variant :
 - Non détecté par un test extérieur,
 - Détecté par le test DI117,
 - Non détecté par l'utilisation du pipeline du DI117 sur les données de séquençage (fichier BAM) générées dans le cadre du test extérieur.

Ce cas montrerait une absence de détection d'un variant par le séquençage réalisé dans le cadre du test extérieur, dont les données sont filtrées à une profondeur de lecture de 30X. Cette situation peut aussi s'expliquer par la présence d'un faux-positif dû à une erreur de séquençage dans le cadre du test DI117.

L'ensemble des variants qui diffèrent entre le test DI117 et les tests externes est récapitulé dans le **Tableau 21** ci-après.

Tableau 21 : Tableau récapitulant les variants décelés qui diffèrent et les hypothèses émises pour expliquer ces écarts de détection

Variants différents entre le test DI117 et les tests des autres CHU	Variant recensé dans les bases de données	Patients concernés	Variant retrouvé dans :			Hypothèses émises
			VCF du Test extérieur	VCF du Test DI117	VCF Bam_ext/Pip_DI117	
EHMT1 chr9:140637804-CT>C NM_024757.4:c.824-3delT	non	Patient 4	oui	non	non	probable Faux-positif de séquençage du test extérieur (région difficile à séquençer : homopolymère de T)
ARID1B chr6:157100005-CGGA>C NM_020732.3:c.954_956delIAGG p.(Gly319del)	oui, SNP	Patient 4	non	oui	oui	polymorphisme dans une région difficile à aligner
KMT2D chr12:49424177-T>G NM_003482.3:c.13885A>C p.(Thr4629Pro)	non	Patient 10	oui	non	non	probable Faux-positif de séquençage du test extérieur (probable mauvaise qualité de l'ADN et région difficile à séquençer)
DLG3 chrX:69665253-A>C NM_021120.3:c.202A>C p.(Thr68Pro)	non	Patient 10	oui	non	non	probable Faux-positif de séquençage du test extérieur (probable mauvaise qualité de l'ADN et région difficile à séquençer)
SLC6A8 chrX:152954025-A>G NM_005629.3:c.-5A>G	oui, SNP	Patients 1, 3, et 10	non	oui	non	profondeur de lecture < 30 X sur le test extérieur
KAT6B chr10:76781905-GGAA>G NM_001256468.1:c.2761_2763delIGAA p.(Glu921del)	oui, SNP	Patient 9	non	oui	oui	polymorphisme dans une région difficile à aligner
SCN8A chr12:52159438-G>GT NM_014191.3:c.2545-7dupT	oui, SNP	Patients 1, 3, et 9	non	oui	oui	probable Faux-positif de séquençage sur les 2 tests (région difficile à séquençer : homopolymère de T)
SMARCA2 chr9:2029018-G>A NM_003070.4:c.-5G>A	oui, SNP	Patient 9	non	oui	oui	profondeur de lecture proche de 30 X sur le test extérieur
PTEN chr10:89720633-C>CT NM_000314.4:c.802-3dupT	non	Patient 1	non	oui	non	probable Faux-positif de séquençage du test DI117 (région difficile à séquençer : homopolymère de T)
SLC16A2 chrX:73641569-T>C NM_006517.4:c.97T>C p.(Ser33Pro)	oui, SNP	Patient 1	non	oui	non	profondeur de lecture < 30 X sur le test extérieur
EHMT1 chr9:140637804-CT>C NM_024757.4:c.824-3delT	non	Patient 1	non	oui	non	probable Faux-positif de séquençage du test DI117 (région difficile à séquençer : homopolymère de T)
RAI1 chr17:17697093-CCAG>C NM_030665.3:c.870_872delIGCA p.(Gln291del)	oui, SNP	Patient 3	non	oui	oui	polymorphisme dans une région difficile à aligner
RAI1 chr17:17697098-AGC>A NM_030665.3:c.837_838delIGC p.(Gln280fs)	oui, SNP	Patient 3	non	oui	oui	polymorphisme dans une région difficile à aligner
RAI1 chr17:17697101-AG>A NM_030665.3:c.840delG p.(Gln280fs)	oui, SNP	Patient 3	non	oui	oui	polymorphisme dans une région difficile à aligner
EP300 chr22:41545024-GT>G NM_001429.3:c.2242-4delT	non	Patient 3	non	oui	non	probable Faux-positif de séquençage du test DI117 (région difficile à séquençer : homopolymère de T)

SNP (single-nucleotide polymorphism) = variant polymorphique, c'est-à-dire dont la fréquence est $\geq 1\%$ dans la population étudiée

Pour chacun des variants du **Tableau 21**, il est possible de déterminer pourquoi ils sont retrouvés dans notre test DI117 et non dans le test externe, ou inversement. Nous pouvons donc les classer en 4 catégories :

a. Profondeur de lecture inférieure à 30X

Une absence de détection d'un variant par un test peut être expliquée par un séquençage de la région du variant avec une profondeur de lecture inférieure à 30X.

En effet, le filtre appliqué est un seuil de profondeur de lecture supérieure à 30X sur l'ensemble des données. Il est ainsi possible qu'un variant ait été initialement décelé avec une profondeur de lecture inférieure par le test externe, et donc non retrouvé avec un seuil à 30X. Le seuil de profondeur de lecture requis pour la détection d'un variant reste discuté, certains laboratoires choisissent de le placer à 20X.

Par exemple, le variant bénin du gène *SLC6A8* (NM_005629.3:c.-5A>G) des patients 1, 3 et 10, est situé dans une région avec une profondeur de lecture inférieure à 30X dans les tests extérieurs de ces patients (**Figure 61**). Il a par conséquent été filtré lors de l'analyse des données de séquençage extérieur. Lors du test DI117, la profondeur de lecture dans cette région est supérieure à 100X. C'est pour cela qu'il ressort comme étant détecté par le test DI117, et non décelé par les tests extérieurs.

b. Anomalies liées à la qualité du séquençage

Des anomalies survenant au cours des différentes étapes du séquençage peuvent entraîner une mauvaise qualité des *reads* obtenus.

Une défaillance peut avoir lieu lors de l'étape de préparation des librairies ou lors de l'étape du séquençage proprement dite. L'utilisation d'un échantillon d'ADN dégradé peut également être à l'origine de la mauvaise qualité de la librairie ou des *reads*.

Les variants retrouvés chez le patient 10 dans les gènes *KMT2D* (NM_003482.3 :c.13885A>C) et *DLG3* (NM_021120.3 :c.202A>C) sont probablement des faux-positifs, secondaires à une erreur de séquençage (**Figure 62** et **Figure 63**). De nombreuses variations de séquence incohérentes sont en effet visibles dans l'ensemble des *reads* situés dans les régions de ces 2 variants. Les régions en question sont réputées difficiles à séquencer, du fait de la répétition de la même base 4 à 6 fois d'affilée. La visualisation des *reads* issus du test DI117 ne montre pas ces variations de séquences incohérentes. L'ADN utilisé pour le test TC-520 est donc probablement de mauvaise qualité. Couplé à la difficulté de séquençage de ces régions, cela a généré deux variants faux-positifs avec une VAF > 35%.

c. Régions difficiles à séquencer

La répétition d'une même base plusieurs fois d'affilée est appelée homopolymère. Ces régions sont réputées être difficile à séquencer. Le taux d'erreur est d'autant plus important que la taille de l'homopolymère augmente. Ainsi, le nombre de bases identiques successives est compliqué à déterminer avec précision lors du séquençage, et des variants faux-positifs (délétion ou insertion) peuvent ainsi être générés. Ces variants faux-positifs ont pour caractéristiques d'être retrouvés chez tous les patients du *run*, et d'avoir une VAF peu élevée.

Ces variants faux-positifs peuvent être filtrés par certains pipelines bio-informatiques. C'est le cas notamment du pipeline du Ion Torrent du CHU de Marseille. En effet, la technologie Ion Torrent comptabilise difficilement le nombre de bases identiques incorporées à la suite au-delà de 3. Par exemple, le variant du gène *SCN8A* (NM_014191.3:c.2545-7dupT) est un variant faux-positif détecté par le pipeline du test DI117, mais pas par le pipeline du test DI-142, sur les données générées lors du séquençage du test DI-142 au CHU de Marseille, chez les patients 1, 3 et 9 (**Figure 64**).

d. Régions difficiles à aligner

Les régions constituées de triplets de nucléotides répétés constituent des régions difficiles à aligner et par conséquent à annoter par le pipeline. Une variation du nombre de triplets nucléotidiques (délétion ou insertion d'un ou plusieurs triplets) est délicate à analyser par le logiciel d'alignement, qui peine à déterminer la position de la variation. Le logiciel génère des alignements différents, conduisant à des annotations différentes pour une même variation. Ces variants sont la plupart du temps des polymorphismes. Il est possible de paramétrer le pipeline bio-informatique pour filtrer ces variants s'ils ne sont pas pathogènes.

Par exemple, le variant du gène *ARID1B* (NM_020732.3:c.954_956delAGG) est un polymorphisme retrouvé chez le patient 4 par le pipeline du test DI117, mais pas par le pipeline du test TC-456, sur les données générées lors du séquençage du test TC-456 au CHU de Strasbourg (**Figure 65**).

Pour conclure, la comparaison et l'analyse du test NGS DI117 face aux autres tests NGS de LBM extérieurs a prouvé que la majorité des variations de séquences détectées par les tests extérieurs sont aussi retrouvées par le test DI117. Pour la détection de ces variants, le test DI117 est équivalent aux tests extérieurs.

Les écarts de détection constatés à partir des mêmes patients ont pu être expliqués par des hypothèses suite à la compilation des données de nos deux analyses comparatives, comme le révèle le **Tableau 21** (page 202). Les trois variants non retrouvés par le test DI117 sont de probables faux-positifs des autres tests, la sensibilité de notre test DI117 n'est par conséquent pas remise en cause par l'absence de leur détection. En revanche, sa spécificité est plus discutable ; il semblerait que trois variants faux-positifs dus à une erreur de séquençage ont été générés par le test NGS DI117, alors que le test extérieur n'a pas détecté ces variations de séquence. Il est plus délicat de conclure sur les autres écarts, les polymorphismes détectés par notre test DI117 étant par exemple vraisemblablement filtrés par les pipelines bio-informatique des autres tests, car ils n'apportent aucune plus-value dans le cadre d'un diagnostic étiologique.

Figure 61 : Représentation sur IGV du variant du gène *SLC6A8* (NM_005629.3:c.-5A>G) du patient 3 situé dans une région ayant une profondeur de lecture < à 30X dans le test extérieur et une profondeur de lecture > à 100X avec le test DI117

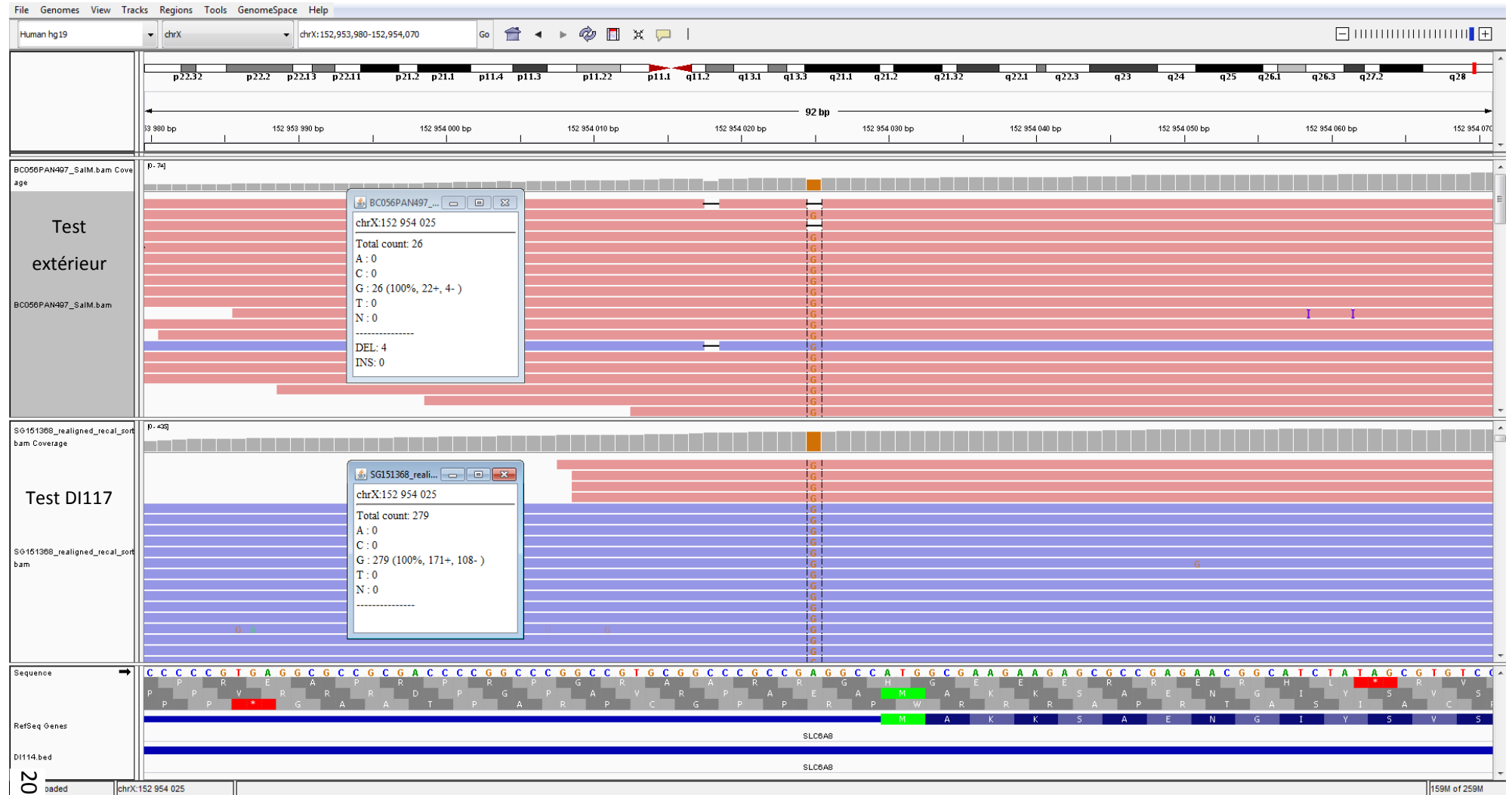


Figure 62 : Visualisation sur IGV du variant retrouvé chez le patient 10 dans le gène *KMT2D* (NM_003482.3 :c.13885A>C) avec le test extérieur, mais non décelé avec le test DI117

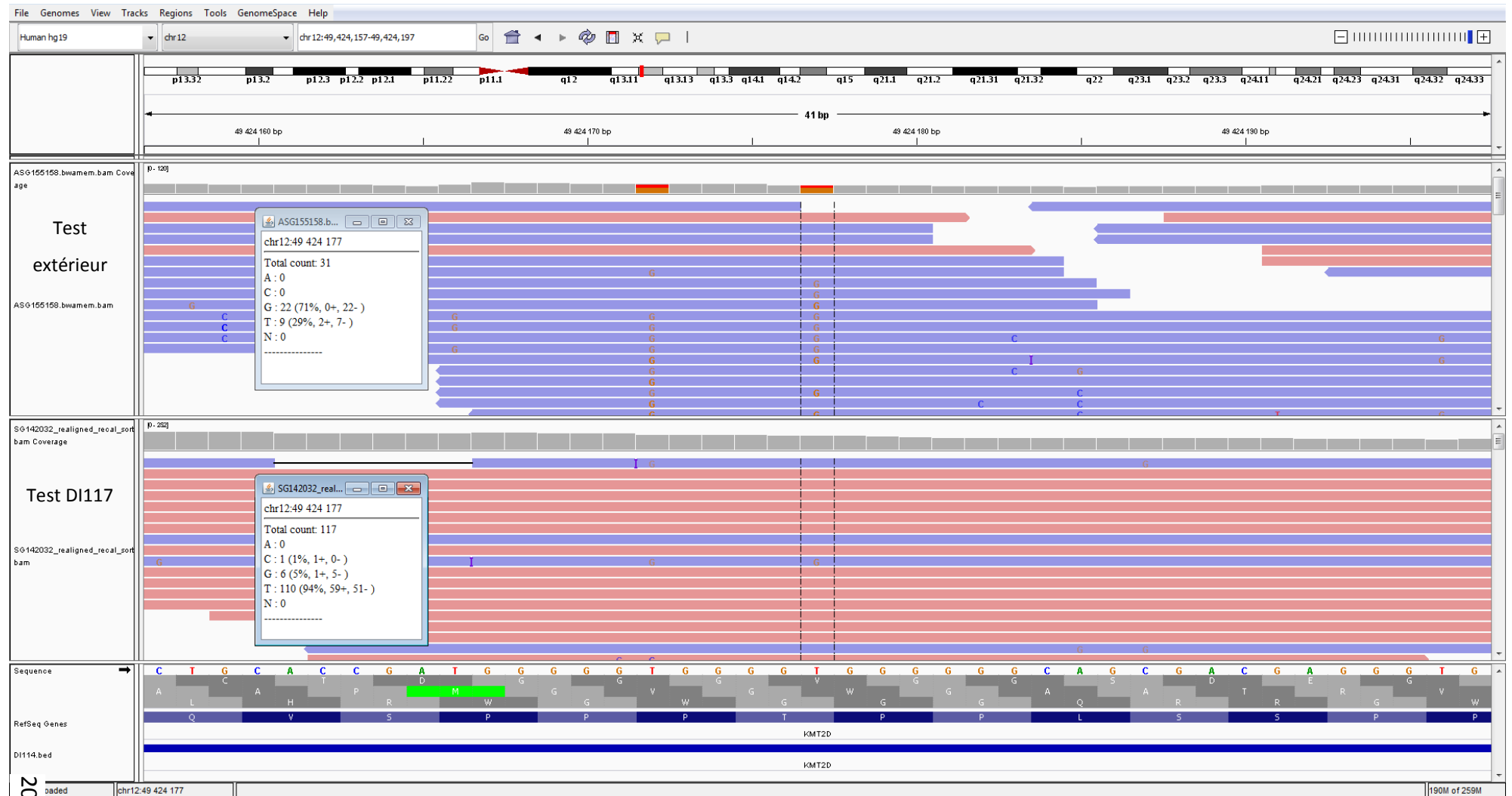


Figure 63 : Visualisation sur IGV du variant retrouvé chez le patient 10 dans le gène *DLG3* (NM_021120.3 :c.202A>C) avec le test extérieur, mais non décelé avec le test DI117

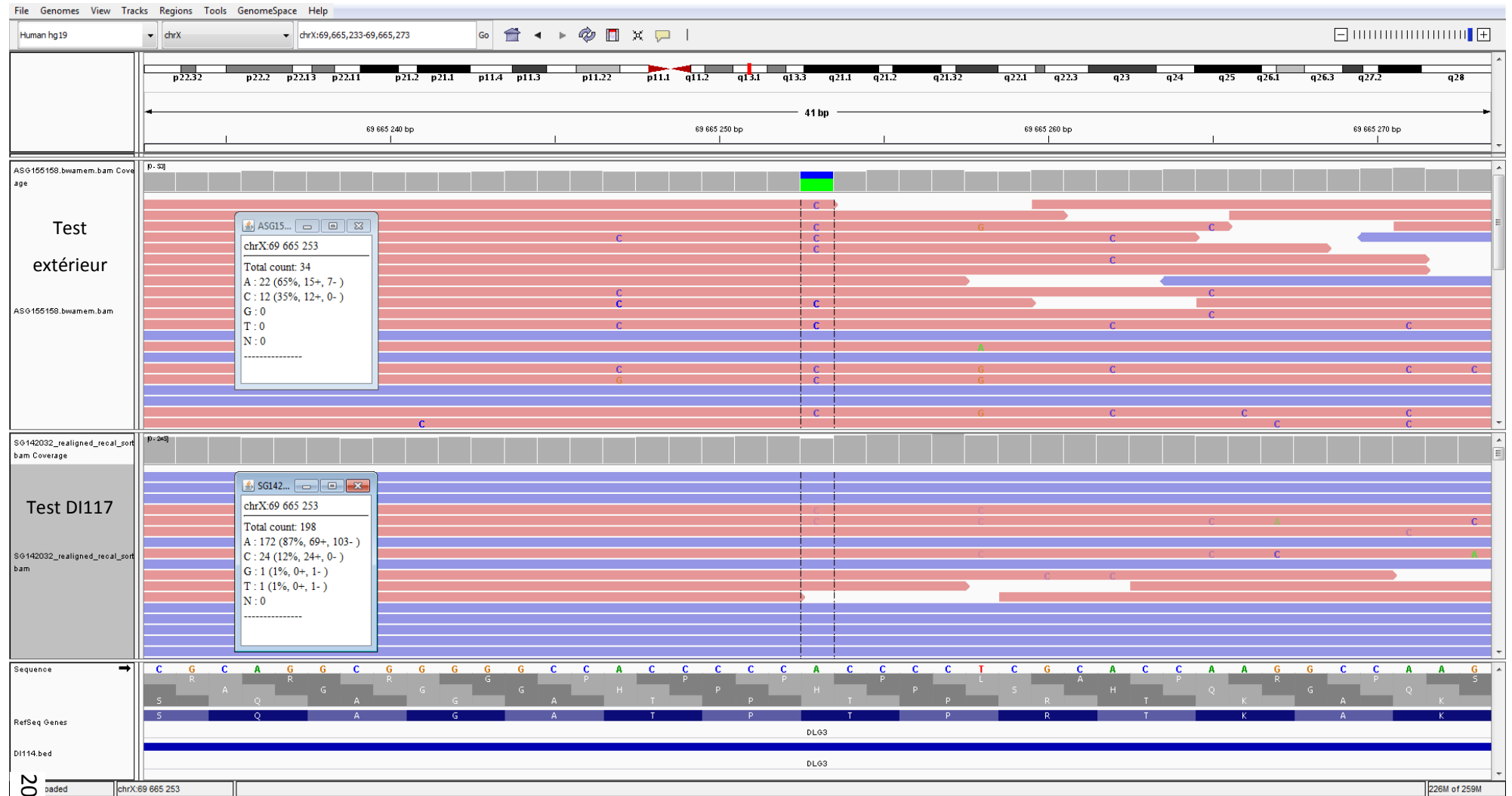
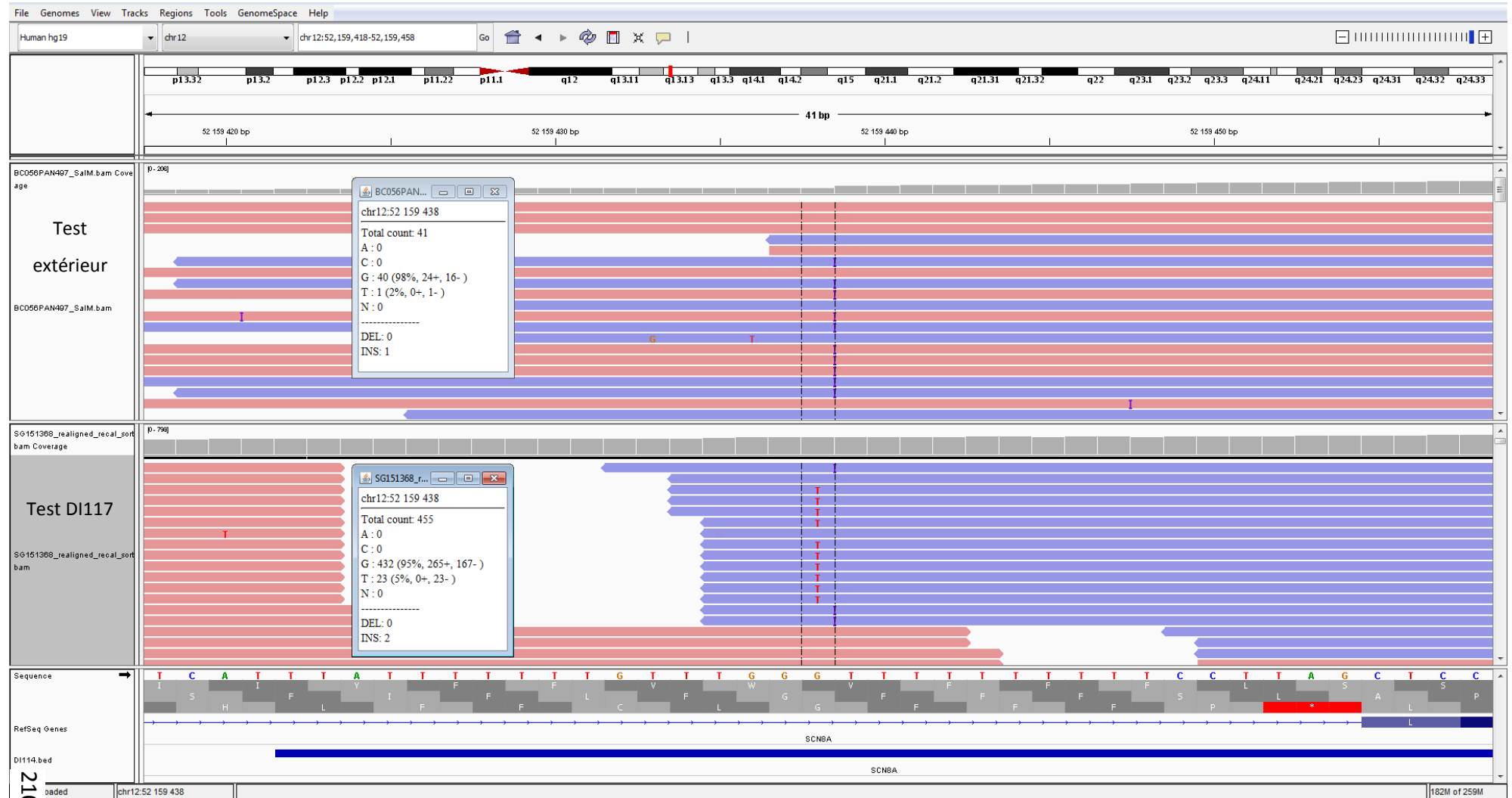


Figure 64 : Représentation sur IGV du variant du gène *SCN8A* (NM_014191.3:c.2545-7dupT) du patient 3, probable variant faux-positif du à une erreur de séquençage, pour le test extérieur et pour le test DI117



Discussion

La déficience intellectuelle fait l'objet de nombreuses consultations au sein du service de génétique clinique du CHU de Toulouse. La grande majorité de ces patients présente une DI sans orientation clinique évidente.

Dans ce contexte et selon un algorithme décisionnel consensuel élaboré par l'ANPGM (**Figure 1**, page 48), il est systématiquement prescrit en première intention une recherche du syndrome de l'X fragile et une ACPA (Analyse Chromosomique sur Puce à ADN). La prévalence de ces prescriptions chaque année au CHU de Toulouse peut être évaluée au moyen des rapports annuels d'activité de notre laboratoire publiés par l'ABM (Agence de la BioMédecine). Ainsi, sur les 4 dernières années (2015 à 2018), c'est en moyenne 227 recherches du syndrome de l'X fragile chez des patients DI qui ont été réalisées annuellement par notre laboratoire, avec un rendement diagnostique de 1,21% (908 tests en 4 ans, dont 11 revenus positifs). Au cours des années 2015 à 2017, le nombre d'ACPA est estimé à 457 par an, avec un rendement de 20,58% (1370 tests en 3 ans, dont 282 avec une anomalie). L'écart entre ces 2 chiffres s'explique par le fait que l'ACPA est prescrite pour de nombreuses autres indications que la DI isolée ou s'accompagnant de troubles du spectre de l'autisme. Par exemple, l'association de la DI à un syndrome polymalformatif ou à un syndrome dysmorphique incitera le prescripteur à rechercher en première intention une anomalie chromosomique de type CNV (variations du nombre de copies) plutôt qu'un syndrome de l'X fragile.

A l'issue de ce premier bilan génétique, la DI de la plupart des patients reste sans étiologie identifiée.

L'arrivée des technologies de séquençage à haut-débit (NGS) a enrichi l'offre de tests diagnostiques en ajoutant en tête de l'arbre décisionnel la possibilité d'explorer en un seul examen des panels de gènes connus pour leur implication dans la DI.

Le laboratoire de génétique du CHU de Toulouse a la maîtrise depuis plusieurs années de la technologie NGS qu'il utilise notamment pour le diagnostic de la mucoviscidose et pour les pathologies du développement oculaire. Vu l'incidence de la DI, estimée à au moins 200 nouveaux cas par an pour notre CHU, il était important pour le laboratoire de mettre à disposition des prescripteurs un test NGS performant ciblant un panel de gènes connus pour leur implication dans la DI.

Le développement de ce nouveau test NGS, appelé test DI117, a nécessité plusieurs décisions stratégiques primordiales, portant notamment sur le choix et le nombre de gènes cibles ainsi que sur le choix de la technologie utilisée. Le Dr Amélie Piton du CHU de Strasbourg, dotée d'une expertise reconnue dans le domaine de la DI, nous a apporté une aide précieuse tout au long de la réalisation de ce projet.

Nous avons finalement opté pour une préparation des bibliothèques avec un enrichissement par capture et un séquençage sur technologie Illumina. Ce choix repose sur un faisceau d'arguments. D'une part, comme nous l'avons montré dans la partie Résultats, paragraphe I.1 (page 104) la méthode d'enrichissement par capture est plus robuste, plus adaptée pour la détection des variations du nombre de copie (CNV) et plus souple dans son développement et ses possibilités de modifications ultérieures. La technologie d'enrichissement par capture revêt de nombreux avantages, mais nous pouvons néanmoins citer quelques inconvénients potentiels. L'enrichissement par capture est précis dans son échantillonnage des régions d'intérêt, mais la localisation des régions d'intérêt par leur amplification sélective est plus spécifique (63). Les amorces sont environ 5 fois plus petites que les sondes de capture. S'il existe des régions homologues dans le génome, les amorces peuvent être conçues pour n'amplifier que la séquence ciblée. La variation d'un seul nucléotide au sein du site de liaison de l'amorce suffit à entraver l'hybridation d'une amorce, et ainsi à empêcher l'amplification de la région homologue. En revanche, du fait de leur plus grande taille, les sondes de capture ne sont pas gênées dans leur hybridation en cas de variation d'un nucléotide, et les régions homologues aux régions ciblées sont également capturées. Il est crucial d'identifier les régions ciblées pour lesquelles une région homologue

existe au sein du génome, et de prendre en considération cette limite technique lors de l'interprétation des résultats après un enrichissement par capture.

Les autres désavantages à l'enrichissement par capture sont le coût des réactifs et le temps technique nécessaire, comme détaillé dans la partie Résultats, paragraphe II.3 (page 176, et **Annexe 11** et **Annexe 12**). Ils sont en effet un peu plus élevés avec l'enrichissement par capture qu'avec l'enrichissement par amplification, bien que cette différence tende à s'amenuiser. Malgré ces quelques désagréments, au vu de ses nombreux atouts, la technique d'enrichissement par capture a été choisie pour notre test.

Le choix de la technologie utilisée pour le séquençage a également été pris après analyse de deux options possibles, les technologies NGS Ion Torrent et Illumina. Le séquençage sur technologie Illumina génère un taux d'erreur moins élevé que la technologie Ion Torrent (64). Les erreurs de séquençage peuvent rendre l'identification d'un variant pathogène compliquée et parfois le masquer par erreur. Les avantages du séquençage sur Ion Torrent restent un temps de *run* plus court et un coût des réactifs légèrement inférieur (65), bien que ces écarts tendent à diminuer avec le développement de ces technologies (cf. **Figure 13**, page 70).

Pourquoi un panel de 117 gènes ?

Le choix du nombre de gènes à inclure dans le test DI117 a été le fruit d'une longue réflexion mêlant plusieurs paramètres. Il a fallu trouver un compromis entre un rendement diagnostique élevé et l'accessibilité du test au plus grand nombre de patients DI.

Un test de première intention requiert non seulement une bonne sensibilité diagnostique mais aussi une bonne applicabilité pour un usage en routine. Autrement dit, un test trop complexe, trop coûteux et ayant des délais de rendu trop long sera inapproprié pour une application en première intention. Il est par conséquent crucial en matière de test NGS ciblé de bien choisir les gènes à séquencer et de se limiter à un nombre qui soit compatible avec

une activité de routine. Dans la pratique, une augmentation du nombre de gènes dans un panel NGS implique, lors des *runs*, une diminution des profondeurs de lecture pour chaque patient ou une baisse du nombre de patients séquencés. Ce qui signifie une qualité de séquençage inférieure ou une augmentation du coût du test par échantillon.

De plus, l'augmentation du nombre de gènes séquencés a pour corollaire l'augmentation du nombre de variations de séquences par patient, et donc du temps d'analyse requis pour l'interprétation de ces variants. Un test de première intention doit pouvoir être réalisé dans un temps acceptable et avoir un délai de rendu de résultat raisonnable. L'errance diagnostique, fréquente chez les patients DI, a des conséquences sanitaire, psychologique, et sociale. Il est par conséquent important qu'un test réalisé en première intention ait pour vocation d'affirmer ou d'écarter des étiologies rapidement. Le retard diagnostique empêche la mise en place d'un suivi médical optimal, voire d'un traitement adapté, et peut avoir un impact sur le devenir du patient. Par ailleurs, l'absence de reconnaissance d'une maladie génétique peut entraver une prise en charge par les organismes sociaux et l'accès à des aides diverses qui permettent d'améliorer la qualité de vie des patients en situation de handicap.

De ce fait, il est crucial de pouvoir réaliser des *runs* fréquemment, sur un nombre de patients en accord avec la population ciblée par ce test, c'est-à-dire sur au minimum une vingtaine de patients DI par mois.

La prise en considération de l'ensemble de ces critères nous a conduits à limiter le nombre de gènes du panel NGS à 117. Ce chiffre est, pour notre laboratoire et les moyens disponibles, le meilleur compromis actuel, médical et médico-économique, conjuguant bonne sensibilité diagnostique et applicabilité satisfaisante pour un test de première intention. La sélection inclut les 44 gènes du test DI44, auxquels ont été ajoutés 70 gènes par ordre de prévalence dans la DI, et 3 gènes impliqués dans le syndrome de Prader-Willi like, constituant ainsi ce test nommé « DI117 » (**Annexe 4**).

La rentabilité diagnostique de ce nouveau test a été estimée à environ 20% sur la base des données de la filière nationale de santé DéfiScience, à 9,33% sur les données issues des publications scientifiques internationales (Pubmed), et à 10,58% en additionnant l'ensemble de ces données (**Annexe 5**). Ces écarts de chiffres peuvent s'expliquer par le choix de la population étudiée et le recrutement d'individus qui diffèrent selon les études. Certaines publications portent sur des DI de sévérité variable alors que d'autres sont focalisées sur des DI sévères ou sur des DI avec un phénotype particulier. Certaines s'intéressent uniquement à des patients ayant eu un bilan de première intention négatif, quand d'autres ont un recrutement plus large. Par ailleurs, la prévalence des DI d'origine génétique par rapport aux DI d'origine environnementale est fortement dépendante de la localisation géographique et du contexte socio-économique. Les données du réseau DéfiScience proviennent de laboratoires de CHU français qui réalisent un test NGS pour le diagnostic de la DI. Il est possible qu'il y ait également un biais de recrutement, puisque les patients ont été explorés pour leur DI suite à une consultation dans un CHU français. Les patients avec une DI légère ne font pas tous l'objet d'une consultation dans un CHU, et n'ont par conséquent pas tous eu un test NGS. De plus, il est possible que le test NGS ait été proposé à certains patients après un premier bilan négatif. Ainsi, la rentabilité diagnostique de notre test DI117 ne pourra être réellement appréciée qu'après sa mise en œuvre dans notre CHU.

Le test DI117 est-il au moins aussi performant que le test DI44 ?

Pour justifier l'utilisation du test NGS DI117 dans notre laboratoire, nous avons évalué les performances (sensibilité et spécificité) du test DI117 par rapport à celles du test DI44 sur les 44 gènes communs aux deux tests.

Un taux élevé de couverture est nécessaire pour une bonne sensibilité du test. La quasi-totalité des variants pathogènes d'un gène sont localisés dans les exons codants et dans les régions d'épissage (62). Notre laboratoire s'est donné comme objectif une couverture de ces régions à 100% afin de détecter la majorité des mutations. En cas de couverture inférieure,

un séquençage complémentaire par la technique de Sanger doit être réalisé pour étudier la région non couverte. Ce séquençage Sanger est chronophage, coûteux, et retarde le rendu du résultat au patient ; traits qui ne peuvent être négligés dans l'évaluation d'un test de première intention.

D'un point de vue théorique, il n'existe pas de différence notable concernant la couverture des 44 gènes en commun entre les 2 tests, bien que le test DI117 paraisse un peu moins performant que le test DI44 (cf. partie Résultats, paragraphe II.1, page 109). Cependant, les couvertures théoriques sont seulement le reflet de la position des amorces ou des sondes de capture conçues pour chacun des tests, la couverture déduite de la pratique expérimentale peut être très différente.

La mise en œuvre expérimentale des 2 tests met en évidence un important défaut de couverture à 30X concernant le test DI44, alors que les couvertures théoriques des 2 tests sont sensiblement identiques.

Cet écart peut s'expliquer par le procédé d'enrichissement choisi pour préparer les librairies qui diffère entre les 2 tests.

Dans le test DI44, la technique d'enrichissement des librairies est basée sur l'amplification sélective des régions cibles au moyen d'une PCR avec des amorces spécifiques. La couverture théorique du test DI44 est déduite des positions où sont supposées s'hybrider chacune des amorces ; elle correspond à la couverture obtenue si toutes les amorces s'hybrident et si toutes les régions cibles sont amplifiées de façon optimale. Si certaines amorces ne s'hybrident pas correctement, la couverture expérimentale devient inférieure à la couverture théorique (63). Ainsi la couverture à 30X constatée du test DI44 est largement inférieure à sa couverture théorique, avec de nombreux *gaps 30X* (cf. partie Résultats, paragraphe II.2.a, page 123).

Dans le test DI117, l'enrichissement des librairies repose sur la sélection de régions cibles à l'aide de sondes de capture spécifiques à ces régions d'ADN. La méthode d'enrichissement par capture est plus robuste que la technique d'enrichissement par amplification, comme

expliqué dans la partie Résultats, paragraphe I.1 (page 104). Ainsi, pour un patient donné et pour une région cible donnée, il est possible que des amorces ne s'hybrident pas correctement alors que les sondes de capture se lient de façon optimale. La couverture expérimentale à 30X est plus susceptible d'équivaloir à la couverture théorique lorsque les librairies sont obtenues par capture plutôt que par amplification (63).

De plus, avec l'enrichissement par capture, la couverture réelle d'un gène peut s'avérer être supérieure à sa couverture théorique. En effet, lors de la préparation de la librairie, l'ADN est découpé en fragments de 200 à 300 pb, alors que les sondes de capture ne font que 120 pb. Une partie seulement des fragments d'ADN capturés est liée à la sonde de capture et appartient donc à la région ciblée (63). La partie des fragments d'ADN capturés non liée à la sonde de capture peut correspondre à une région adjacente à la région cible. Par conséquent, les régions adjacentes aux régions ciblées peuvent être séquencées, et une région d'un gène non couverte en théorie peut l'être en pratique. Ainsi le paragraphe II.2.a de notre partie Résultats (page 123) montre que pour le test DI117, 14 des 44 gènes étudiés ont une couverture à 30X mesurée supérieure à leur couverture théorique.

Il est de ce fait primordial d'évaluer expérimentalement la couverture d'un test qui peut se révéler très éloignée de celle attendue en théorie.

Le test DI117 s'avère avoir une meilleure couverture que le test DI44, laissant supposer une meilleure sensibilité diagnostique. De plus, le nombre de séquençages complémentaires par la technique de Sanger nécessaires est estimé deux à quatre fois moindre avec le test DI117 qu'avec le test DI44 (cf. partie Résultats, paragraphe II.2.a.ii, page 135).

Nous avons choisi d'évaluer expérimentalement les 2 tests sur 2 autres critères qui nous paraissaient pertinents pour un test NGS diagnostique : la profondeur de lecture et la capacité à mettre en évidence les variants candidats.

La profondeur de lecture est liée à la probabilité de détection d'un variant candidat, et par conséquent à la sensibilité diagnostique du test.

Par ailleurs, la profondeur de lecture moyenne sur l'ensemble des régions ciblées (PLM) de chaque patient est un des reflets de la qualité de la librairie du patient. Les 2 tests DI44 et DI117 présentent une PLM élevée pour l'ensemble des 6 patients en commun, et il n'y a pas de différence notable lorsque l'on compare la PLM entre les 2 tests pour chaque patient (cf. partie Résultats, paragraphe II.2.b, page 139).

La performance diagnostique d'un test NGS se mesure par sa capacité à révéler les variants candidats (sensibilité), mais également par son potentiel à simplifier leur détection en évitant de les mêler à d'innombrables variants faux-positifs (spécificité).

Les variants pathogènes, confirmés par un séquençage Sanger, de l'ensemble des 6 patients en commun ont été mis en évidence par les 2 tests, sans différence notable (cf. partie Résultats, paragraphe II.2.c.i, page 143). La sensibilité des 2 tests est équivalente pour ces 6 patients.

Cependant, le test DI117 semble plus performant par rapport au test DI44 pour la recherche de variants candidats au sein des 44 gènes en commun.

Lors de l'analyse des résultats des 6 patients en commun avec le filtre bio-informatique « Filtre 2 – 2 VC », 6 variants non pathogènes ne sont pas détectés par le test DI44 contre 2 pour le test DI117. Cela peut s'expliquer notamment par la couverture à 30X du test DI117, nettement supérieure à celle du test DI44 (cf. partie Résultats, paragraphe II.2.c.iii, page 151).

Le test DI117 a également démontré une meilleure spécificité par rapport au test DI44. Le test DI44 génère en effet un nombre beaucoup plus élevé de variants faux-positifs comparé au test DI117.

Cette différence importante de spécificité peut être expliquée par la technologie utilisée lors du séquençage.

La technologie Ion Torrent est reconnue pour engendrer un taux d'erreurs de séquençage bien supérieur par rapport à la technologie Illumina, notamment dans les régions du génome réputées compliquées à séquencer tels que les homopolymères* (65). Le principe du séquençage sur technologie Ion Torrent repose sur la détection de la variation du pH engendrée par la libération d'ions H⁺ lors de l'incorporation d'un nucléotide à l'ADN en cours de séquençage (cf. partie Introduction, paragraphe II.1.b.i.4, page 67). Plus le nombre de bases identiques successives est important, plus le nombre de nucléotides intégrés durant le même cycle augmente, et plus la variation de pH détectée est élevée. Cependant, au-delà de 3 bases identiques à la suite, le nombre de bases responsables de la variation du pH est difficile à déterminer avec précision par cette technologie.

Le séquençage de bases identiques avec le système Illumina se traduit par l'émission d'un fluorophore de même couleur, et ce à chacun des cycles de séquençage (cf. partie Introduction, paragraphe II.1.b.i.3, page 61).

Le nombre de bases identiques détectées successivement est donc déterminé avec plus de précision sur la technologie Illumina que sur la technologie Ion Torrent. Il en résulte un nombre de variants faux-positifs (délétion ou insertion) beaucoup moins important avec le système Illumina (cf. partie Résultats, paragraphe II.2.c.ii, page 145).

Ce nombre très inférieur de faux-positifs du test DI117 par rapport au test DI44 atteste de la plus grande spécificité du test DI117.

Ces erreurs de séquençage peuvent se situer au niveau de la position d'un variant candidat, et empêcher ainsi sa détection (66). Le nombre de potentiels faux-négatifs secondaires aux erreurs de séquençage du test DI44 est donc plus élevé comparé au test DI117, cela suggère une meilleure sensibilité du test DI117 par rapport au test DI44.

Tous ces résultats ont permis de démontrer la supériorité en matière de sensibilité et de spécificité du test diagnostique DI117 par rapport au test DI44 sur les 44 gènes en commun pour les 6 patients sélectionnés. Ce nombre de patient est limité, mais suffisant pour justifier

l'utilisation du test DI117 à la place du test DI44 dans notre laboratoire. Toutefois la robustesse de la technique n'a pas été évaluée entre les deux tests. Il serait intéressant de comparer le séquençage par les deux tests de l'échantillon du Patient 7 (cf. **Tableau 2**, page 82), dont l'ADN conservé était de mauvaise qualité.

Le test DI117 satisfait-il pleinement aux exigences de qualité requises pour une application à visée diagnostique ?

Nous avons souhaité évaluer les performances propres au test NGS DI117 sur l'ensemble des 117 gènes en l'expérimentant sur 15 patients. Afin d'objectiver l'apport médical de ce nouveau test diagnostique en ce qui concerne sa sensibilité et sa spécificité, nous avons évalué les mêmes caractéristiques que précédemment, à savoir la profondeur de lecture, la couverture des régions ciblées, et la capacité à détecter le variant pathogène de chaque patient.

Nous avons tout d'abord contrôlé la PLM de chaque patient. La PLM est supérieure à 600X, soit 3 fois plus élevé que le seuil qualité de 200X fixé par notre laboratoire (cf. partie Résultats, paragraphe III.1, page 178). Cette PLM élevée, couplée à une profondeur de lecture par gène élevée pour tous les patients, permettent d'envisager une augmentation du nombre de patients séquencés au sein d'un même *run* du test DI117.

Nous avons poursuivi par l'étude de la couverture de notre test DI117, critère étroitement lié à la sensibilité diagnostique du test. La couverture de l'ensemble des régions ciblées est satisfaisante, supérieure à 99% à 30X et au-dessus de 96% à 100X. Cependant, 20 gènes ont une couverture à 30X comprise entre 95 et 99%, et 6 entre 90 et 95%. Le test DI117 semble

avoir une bonne sensibilité diagnostique pour les trois-quarts des 117 gènes analysés (cf. partie Résultats, paragraphe III.2, page 180).

Nous nous sommes ensuite intéressés à la capacité de détection des variants pathogènes de type SNV confirmés par un séquençage Sanger des 15 patients analysés. Le test DI117 a démontré une sensibilité optimale, et une excellente fiabilité pour détecter les mutations de type SNV chez les patients étudiés. Cela conforte l'intérêt médical du test DI117 pour les patients atteints de DI. Toutefois, le nombre de gènes explorés reste limité du fait du faible nombre de patients analysés, dont certains présentent une mutation de type SNV dans le même gène (cf. partie Résultats, paragraphe III.3, page 184).

Les variants de type CNV ne sont pas recherchés actuellement par le pipeline bio-informatique du test DI117. Les mutations de type CNV du patient 15 n'ont donc pas pu être mises en évidence. Néanmoins, nous avons vérifié l'absence de variant hétérozygote dans la zone de la délétion, et donc en simple copie, du gène *MECP2* chez ce patient.

Nous avons souhaité étayer ces résultats en étudiant la sensibilité, la spécificité et la reproductibilité de notre test DI117 par rapport aux tests NGS de DI validés et réalisés dans d'autres laboratoires de CHU français.

Au sein de notre cohorte, 5 patients avaient fait l'objet d'un séquençage NGS sur un test dédié à la DI dans un autre CHU. Pour chacun d'entre eux, nous avons comparé le fichier VCF généré par notre test DI117 à celui fourni par le LBM extérieur, en concentrant notre analyse sur les 67 gènes communs à l'ensemble des tests.

Nous avons également testé le pipeline d'analyse bio-informatique du test DI117 sur les fichiers BAM récupérés auprès des laboratoires extérieurs, ce qui a permis de comparer la performance de notre pipeline sur le maximum de gènes en commun entre notre test DI117 et chacun des autres tests.

Le séquençage et le pipeline d'analyse bio-informatique mis en œuvre dans le cadre du test DI117 sont exhaustifs sur les régions ciblées ; tous les variants réels pathogènes et non pathogènes décelés par les tests des autres CHU ont été mis en évidence par le test DI117.

Les variants non détectés sont des variants faux-positifs liés à des erreurs de séquençage.

Grâce à cette comparaison minutieuse, nous avons pu éprouver la performance de notre test NGS DI117 en matière de sensibilité et de spécificité sur un nombre important de gènes, mais un nombre limité de patients (cf. partie Résultats, paragraphe IV, page 187).

Ces résultats renforcent l'impression de pertinence médicale de son utilisation en tant que test diagnostique de première intention chez les patients DI.

Néanmoins, l'analyse s'est faite sur un nombre trop limité de patients pour pouvoir statuer définitivement sur ses performances. Il est recommandé par l'ANPGM d'analyser au moins 30 variations de séquence différentes confirmées en séquençage Sanger pour conclure sur la spécificité et la sensibilité analytiques d'un test NGS.

La robustesse du test DI117 n'a pas été éprouvée. Il serait intéressant de séquencer un ADN de mauvaise qualité ou issu d'un autre type de prélèvement (par exemple salivaire) pour vérifier ce critère.

Le test DI117 est-il perfectible ?

L'ensemble de ces réflexions montrent que des perfectionnements peuvent être apportés au test DI117.

La couverture du test DI 117 n'est pas optimale sur les 117 gènes. Une conception de sondes supplémentaires, ajoutée à une modification et à une densification ciblée des sondes de capture existantes, permettrait de réduire le nombre de régions insuffisamment séquencées. Cette adaptation du design profiterait à la sensibilité diagnostique du test qui s'en verrait améliorée.

Les variants de type CNV ne sont actuellement pas recherchés par le pipeline bio-informatique du test DI117. Ces variants CNV sont fréquemment à l'origine de DI, il serait par conséquent intéressant de développer un pipeline d'analyse pour les mettre en évidence avec le test DI117.

Les gènes inclus dans le test DI117 ont été sélectionnés en fonction de leur fréquence d'implication dans la DI en France. Il n'est pas exclu que d'autres gènes responsables de DI soient découverts ou que les prévalences d'implication de certains gènes évoluent.

Ce test DI117 est donc susceptible d'évoluer en fonction des données scientifiques existantes sur l'imputabilité des différents gènes dans la DI.

Par ailleurs, aucun variant pathogène n'a été mis en évidence chez le Patient 3, qui n'a pas encore de diagnostic moléculaire confirmé, et pour qui le test NGS DI-142 n'a décelé aucune mutation dans les 142 gènes de DI étudiés. L'absence de découverte de variant causal chez ce patient peut s'expliquer par l'absence du gène responsable au sein des 117 gènes analysés par le test DI117, ou par les limites d'un test NGS de 2^{ème} génération.

Quelles sont les limites d'un test NGS de 2^{ème} génération ?

Plusieurs problématiques bien connues du séquençage NGS de 2^{ème} génération ont été mises en exergue au long de ce projet.

Le choix des différents outils du pipeline d'analyse bio-informatique revêt une importance cruciale pour la génération des données. Il est essentiel de bien connaître et de prendre en considération leurs limites pour l'interprétation des résultats.

Ainsi, certains *Variants Caller* (VC) sont plus sensibles pour déceler les variants que d'autres. C'est le cas par exemple du VC HaplotypeCaller de GATK (*Genome Analysis ToolKit*) dont la sensibilité pour détecter les indels est bien supérieure à celle du VC UnifiedGenotyper de GATK. Il est d'ailleurs fortement conseillé par le Broad Institute, développeur des deux VC, d'utiliser le VC HaplotypeCaller au lieu du VC UnifiedGenotyper, obsolète. Par ailleurs, certains VC peuvent filtrer les variants bénins fréquents dans la population générale (polymorphisme) ou les variants considérés comme des erreurs de séquençage (faux-positifs).

Les régions répétées sont difficiles à aligner, et par conséquent à annoter, par le pipeline d'analyse bio-informatique. Cette problématique découle de la longueur insuffisante des *reads* produits avec le séquençage NGS de 2^{ème} génération. En effet, si le *read* n'est pas assez grand pour contenir à la fois la région répétée et une extrémité qui permettrait d'orienter l'alignement, le logiciel d'alignement peine à déterminer la position du *read* (67). Ainsi, lors d'une variation du nombre de séquences répétées, le logiciel d'alignement peut générer des alignements différents, conduisant à des annotations différentes, pour la même variation. Cela a pour conséquence une diminution de la VAF d'un éventuel variant, et donc possiblement sa non-détection.

Les séquences homologues existant à un autre endroit du génome peuvent également être sources de difficulté d'alignement (68). Ainsi, si une séquence identique existe à plusieurs endroits du génome, le logiciel d'alignement du pipeline ne peut pas déterminer où aligner la séquence lue (le *read*) entre plusieurs positions candidates. Le logiciel assigne alors une mauvaise qualité d'alignement au *read*. L'alignement se fait aléatoirement, éventuellement au mauvais endroit. Il en résulte deux conséquences nuisibles potentielles :

- L'absence de détection de variant de la séquence étudiée. Le *read* possiblement porteur de variation peut être aligné au niveau de la séquence homologue par erreur. Le *read* n'est alors pas assigné à la séquence étudiée, et le variant n'est pas détecté par le test.
- La détection de variant de la séquence homologue, dont le *read* est aligné par erreur à l'endroit de la séquence étudiée. Ce variant est détecté par le test de manière incorrecte, le *read* porteur de variation n'étant pas issu de la séquence étudiée, mais issu de la séquence homologue.

Dans cette configuration, il est impossible de distinguer le *read* issu de la séquence étudiée par rapport au *read* issu de la séquence homologue. Pour lever le doute, il est nécessaire de réaliser un séquençage par la technique de Sanger avec une amplification sélective basée sur des amorces spécifiques de la séquence étudiée.

Le séquençage NGS de 2^{ème} génération est une technique très performante. Cependant la longueur insuffisante des *reads* rend la détection de certains variants difficile. Le développement du séquençage NGS de 3^{ème} génération a pour objectif l'augmentation de la longueur des *reads* générés. Il paraît donc s'imposer comme la solution idéale pour dépasser ces problématiques. Cependant, comme expliqué dans la partie Introduction, paragraphe I.1.a (page 71) le taux d'erreur est actuellement encore trop élevé avec ces technologies pour une utilisation diagnostique pour les pathologies humaines.

Quelles sont les perspectives pour le diagnostic génétique de la DI ?

L'extrême hétérogénéité génétique de la DI, avec plus de 700 gènes identifiés comme potentiellement impliqués, rend son diagnostic étiologique souvent compliqué à établir. Certaines études pointent que près de 60% des DI pourraient s'expliquer par des causes génétiques (30), alors que le test DI117 aurait un rendement diagnostique de 21% maximum. Les stratégies de séquençage complet de l'exome (WES, *whole exome sequencing*) et du génome (WGS, *whole genome sequencing*) pourraient devenir des alternatives plus profitables sur le plan diagnostique que notre test DI117.

Le WGS est capable de mettre en évidence les réarrangements génomiques et pourrait ainsi également remplacer l'ACPA (69).

Il a été cependant établi que réaliser un WES précocement pouvait être l'approche optimale pour réduire l'errance diagnostique dans la DI, en particulier pour les enfants avec un phénotype sévère, car il couvre tous les gènes, et il n'est pas aussi coûteux et complexe que le WGS (70, 71)

Ces techniques gardent aussi des limites. Elles impliquent la production d'une importante quantité de données de séquençage, ce qui nécessite une infrastructure ayant un séquenceur et des ressources informatiques de traitement et de stockage des données adaptés. Cette infrastructure ne peut pas être établie dans tous les laboratoires de génétique, et cette contrainte rend ces techniques plus onéreuses que notre test DI117.

Leur interprétation est rendue compliquée par le nombre plus important de variants générés. Les variants sont difficiles à interpréter lorsqu'ils se situent dans des gènes insuffisamment documentés en physiopathologie humaine. Il est nécessaire de réaliser des tests fonctionnels pour conclure sur leur pathogénicité, par exemple avec des modèles animaux. Malheureusement, la recherche associée à ces tests prend un temps considérable

qui ne saurait être compatible avec la nécessité d'un diagnostic rapide souhaité pour un test de première intention.

Les stratégies de WES et WGS offrent des taux de couverture encore inférieurs aux panels de gènes sur certains gènes (72). Leur sensibilité est légèrement moins bonne que celle des tests NGS sur un panel de gènes, et elles sont donc un peu moins performantes pour détecter des mutations dans les gènes connus.

La limitation du nombre de gènes séquencés permet de limiter les découvertes fortuites, c'est-à-dire une information génétique sans lien avec l'indication ayant conduit à la prescription de l'examen. Ces découvertes fortuites soulèvent un réel débat éthique sur les informations génétiques à transmettre au patient quand celles-ci sont d'expression incertaine ou liées à une maladie pour laquelle aucune action préventive ou curative ne peut être mise en œuvre. Elles peuvent compliquer le conseil génétique et portent à réfléchir sur l'intérêt pour le patient de l'exploration génétique (73).

L'analyse des régions codantes du génome n'arrive pas à expliquer tous les cas de DI (69). Au maximum la moitié des cas de DI a un diagnostic posé avec un WES. Le rôle des régions non-codantes, correspondant à 98% du génome, tend à être de plus en plus mis en évidence. Elles sont importantes en particulier pour la régulation des gènes. L'étude des ARNm* exprimés par le séquençage haut-débit de l'ARN (*RNA sequencing*) pourrait être un outil complémentaire pour identifier les variants dans les régions non codantes et comprendre en même temps leurs conséquences. Cependant, la limite de cette étude dans le cadre de la DI est l'inaccessibilité du tissu cérébral. Les gènes dont l'expression est altérée spécifiquement dans le cerveau sont difficiles à mettre en évidence. Une alternative pourrait être de réaliser ce séquençage de l'ARNm sur des cellules neuronales induites *in vitro* à partir de cellules issues d'une biopsie de peau d'un patient, mais cette analyse est fastidieuse, longue et coûteuse, ce qui la rend inadéquate pour un test de première intention.

Conclusion

Le test DI117 a montré une sensibilité et une spécificité diagnostiques au moins équivalentes au test DI44, pour leurs 44 gènes en commun. Le test DI117 a également présenté une sensibilité et une spécificité satisfaisantes pour la majorité des 73 gènes supplémentaires qu'il étudie. La comparaison des résultats obtenus avec d'autres tests NGS dédiés à la DI pour 5 patients a permis de tester sa reproductibilité. Certes, ces résultats sont issus d'une cohorte limitée de patients, et une étude sur un nombre plus élevé de patients est nécessaire pour conclure avec certitude sur les performances d'un test diagnostique.

Au vu des résultats de ce travail, des améliorations du design du test DI117 ont été réalisées pour améliorer sa couverture. L'évaluation de la nouvelle version du test est d'ailleurs prévue prochainement sur notre cohorte de patients.

Pour conclure, le test DI117 paraît être un test NGS diagnostique performant, et sa prescription médicale en première intention semble être pertinente chez les patients DI.

Références bibliographiques

1. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem*. 2009;55(4):641-58.
2. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet*. 2018;34(9):666-81.
3. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11(1):31-46.
4. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabil*. 2011;32(2):419-36.
5. Inserm. Déficiences intellectuelles. Montrouge: EDP Sciences; 2016. 1157 p.
6. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev*. 2002;8(3):117-34.
7. Westerinen H, Kaski M, Virta L, Almqvist F, Iivanainen M. Prevalence of intellectual disability: a comprehensive study based on national registers. *J Intellect Disabil Res*. 2007;51(Pt 9):715-25.
8. David M, Dieterich K, Billette de Villemeur A, Jouk PS, Counillon J, Larroque B, et al. Prevalence and characteristics of children with mild intellectual disability in a French county. *J Intellect Disabil Res*. 2014;58(7):591-602.
9. Petterson B, Bourke J, Leonard H, Jacoby P, Bower C. Co-occurrence of birth defects and intellectual disability. *Paediatr Perinat Epidemiol*. 2007;21(1):65-75.
10. Kim MA, Lee YS, Yee NH, Choi JS, Choi JY, Seo K. Prevalence of congenital heart defects associated with Down syndrome in Korea. *J Korean Med Sci*. 2014;29(11):1544-9.
11. Bryson SE, Bradley EA, Thompson A, Wainwright A. Prevalence of autism among adolescents with intellectual disabilities. *Can J Psychiatry*. 2008;53(7):449-59.
12. Fombonne E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J Autism Dev Disord*. 2003;33(4):365-82.
13. Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DE. Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review. *Intellect Dev Disabil*. 2011;49(2):59-85.
14. Patja K, Iivanainen M, Vesala H, Oksanen H, Ruoppila I. Life expectancy of people with intellectual disability: a 35-year follow-up study. *J Intellect Disabil Res*. 2000;44 (Pt 5):591-9.
15. Tyrer F, McGrother C. Cause-specific mortality and death certificate reporting in adults with moderate to profound intellectual disability. *J Intellect Disabil Res*. 2009;53(11):898-904.
16. Liu Y, McDermott S, Lawson A, Aelion CM. The relationship between mental retardation and developmental delays in children and the levels of arsenic, mercury and lead in soil samples taken near their mother's residence during pregnancy. *Int J Hyg Environ Health*. 2010;213(2):116-23.
17. Mendola P, Selevan SG, Gutter S, Rice D. Environmental factors associated with a spectrum of neurodevelopmental deficits. *Ment Retard Dev Disabil Res Rev*. 2002;8(3):188-97.
18. McDermott S, Wu J, Cai B, Lawson A, Marjorie Aelion C. Probability of intellectual disability is associated with soil concentrations of arsenic and lead. *Chemosphere*. 2011;84(1):31-8.
19. Benoist G, Jacquemard F, Leruez-Ville M, Ville Y. [Cytomegalovirus (CMV) congenital infection]. *Gynecol Obstet Fertil*. 2008;36(3):248-60.
20. Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet*. 2012;379(9814):445-52.

21. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nat Genet.* 2010;42(12):1109-12.
22. Winter RM, Baraitser M. The London Dysmorphology Database. *J Med Genet.* 1987;24(8):509-10.
23. Kohler S, Schulz MH, Krawitz P, Bauer S, Dolken S, Ott CE, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet.* 2009;85(4):457-64.
24. Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014;42(Database issue):D966-74.
25. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006;444(7118):444-54.
26. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet.* 2010;19(R2):R131-6.
27. Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A.* 2006;140(19):2063-74.
28. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014;511(7509):344-7.
29. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 2016;17(1):9-18.
30. Kvarnung M, Nordgren A. Intellectual Disability & Rare Disorders: A Diagnostic Challenge. *Adv Exp Med Biol.* 2017;1031:39-54.
31. Guerrini R, Shanahan JL, Carrozzo R, Bonanni P, Higgs DR, Gibbons RJ. A nonsense mutation of the ATRX gene causing mild mental retardation and epilepsy. *Ann Neurol.* 2000;47(1):117-21.
32. Field M, Tarpey P, Boyle J, Edkins S, Goodship J, Luo Y, et al. Mutations in the RSK2(RPS6KA3) gene cause Coffin-Lowry syndrome and nonsyndromic X-linked mental retardation. *Clin Genet.* 2006;70(6):509-15.
33. Hoyer J, Ekici AB, Ende S, Popp B, Zweier C, Wiesener A, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet.* 2012;90(3):565-72.
34. Pavlowsky A, Chelly J, Billuart P. Major synaptic signaling pathways involved in intellectual disability. *Mol Psychiatry.* 2012;17(7):663.
35. van Karnebeek CD, Scheper FY, Abeling NG, Alders M, Barth PG, Hoovers JM, et al. Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard.* 2005;110(4):253-67.
36. Barbelanne M, Tsang WY. Molecular and cellular basis of autosomal recessive primary microcephaly. *Biomed Res Int.* 2014;2014:547986.
37. Tonosaki M, Itoh K, Umekage M, Kishimoto T, Yaoi T, Lemmon VP, et al. L1cam is crucial for cell locomotion and terminal translocation of the Soma in radial migration during murine corticogenesis. *PLoS One.* 2014;9(1):e86186.
38. Guiberson NGL, Pineda A, Abramov D, Kharel P, Carnazza KE, Wragg RT, et al. Mechanism-based rescue of Munc18-1 dysfunction in varied encephalopathies by chemical chaperones. *Nat Commun.* 2018;9(1):3986.
39. Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T, et al. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. *Proc Natl Acad Sci U S A.* 2007;104(46):18163-8.
40. Kishino T, Lalande M, Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet.* 1997;15(1):70-3.
41. Nadif Kasri N, Nakano-Kobayashi A, Malinow R, Li B, Van Aelst L. The Rho-linked mental retardation protein oligophrenin-1 controls synapse maturation and plasticity by stabilizing AMPA receptors. *Genes Dev.* 2009;23(11):1289-302.

42. Roberts AE, Allanson JE, Tartaglia M, Gelb BD. Noonan syndrome. *Lancet*. 2013;381(9863):333-42.
43. Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*. 2015;522(7554):89-93.
44. Stankiewicz P, Beaudet AL. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr Opin Genet Dev*. 2007;17(3):182-92.
45. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*. 2010;86(5):749-64.
46. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet*. 2011;43(9):838-46.
47. Schaaf CP, Wiszniewska J, Beaudet AL. Copy number and SNP arrays in clinical diagnostics. *Annu Rev Genomics Hum Genet*. 2011;12:25-51.
48. Shoukier M, Klein N, Auber B, Wickert J, Schroder J, Zoll B, et al. Array CGH in patients with developmental delay or intellectual disability: are there phenotypic clues to pathogenic copy number variants? *Clin Genet*. 2013;83(1):53-65.
49. de Vries BB, Mohkamsing S, van den Ouweland AM, Mol E, Gelsema K, van Rijn M, et al. Screening for the fragile X syndrome among the mentally retarded: a clinical study. The Collaborative Fragile X Study Group. *J Med Genet*. 1999;36(6):467-70.
50. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*. 1977;265(5596):687-95.
51. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986;321(6071):674-9.
52. Ahmadian A, Ehn M, Hober S. Pyrosequencing: history, biochemistry and future. *Clin Chim Acta*. 2006;363(1-2):83-94.
53. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452(7189):872-6.
54. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res*. 2009;19(9):1527-41.
55. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53-9.
56. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011;475(7356):348-52.
57. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323(5910):133-8.
58. Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*. 2018;46(5):2159-68.
59. Haque F, Li J, Wu HC, Liang XJ, Guo P. Solid-State and Biological Nanopore for Real-Time Sensing of Single Chemical and Sequencing of DNA. *Nano Today*. 2013;8(1):56-74.
60. Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum Mutat*. 2015;36(9):903-14.
61. Redin C, Gerard B, Lauer J, Herenger Y, Muller J, Quartier A, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet*. 2014;51(11):724-36.
62. Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet*. 2004;5(5):389-96.

63. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010;7(2):111-8.
64. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. 2012;13:341.
65. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 2012;30(5):434-9.
66. Yeo ZX, Chan M, Yap YS, Ang P, Rozen S, Lee AS. Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *PLoS One*. 2012;7(9):e45798.
67. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2011;13(1):36-46.
68. Santani A, Murrell J, Funke B, Yu Z, Hegde M, Mao R, et al. Development and Validation of Targeted Next-Generation Sequencing Panels for Detection of Germline Variants in Inherited Diseases. *Arch Pathol Lab Med*. 2017;141(6):787-97.
69. Willemsen MH, Kleefstra T. Making headway with genetic diagnostics of intellectual disabilities. *Clin Genet*. 2014;85(2):101-10.
70. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med*. 2014;6(265):265ra168.
71. Sun Y, Ruivenkamp CA, Hoffer MJ, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat*. 2015;36(6):648-55.
72. LaDuca H, Farwell KD, Vuong H, Lu HM, Mu W, Shahmirzadi L, et al. Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS One*. 2017;12(2):e0170843.
73. Mackley MP, Fletcher B, Parker M, Watkins H, Ormondroyd E. Stakeholder views on secondary findings in whole-genome and whole-exome sequencing: a systematic review of quantitative and qualitative studies. *Genet Med*. 2017;19(3):283-93.

Annexes

Annexe 1 : Tableau récapitulatif des gènes inclus dans les différents tests NGS de DI

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
ANKRD11	X	X	X	X
ARID1B	X	X	X	X
ATRX	X	X	X	X
CASK	X	X	X	X
CDKL5	X	X	X	X
CHD2	X	X	X	X
CHD7	X	X	X	X
CREBBP	X	X	X	X
CTNBN1	X	X	X	X
CUL4B	X	X	X	X
DLG3	X	X	X	X
DYNC1H1	X	X	X	X
DYRK1A	X	X	X	X
EHMT1	X	X	X	X
EP300	X	X	X	X
FLNA	X	X	X	X
FOXP1	X	X	X	X
FOXP2	X	X	X	X
GATAD2B	X	X	X	X
GRIA3	X	X	X	X
GRIN1	X	X	X	X
GRIN2A	X	X	X	X
GRIN2B	X	X	X	X
HDAC8	X	X	X	X
HUWE1	X	X	X	X
IL1RAP1	X	X	X	X
IQSEC2	X	X	X	X
KANSL1	X	X	X	X
KAT6B	X	X	X	X
KCNQ2	X	X	X	X
KDM5C	X	X	X	X
KIAA2022	X	X	X	X
KMT2D	X	X	X	X
MECP2	X	X	X	X
MED12	X	X	X	X
MED13L	X	X	X	X
MEF2C	X	X	X	X
KMT2A	X	X	X	X
NAA10	X	X	X	X
NFIX	X	X	X	X
NSD1	X	X	X	X
OPHN1	X	X	X	X
PPP2R5D	X	X	X	X
PQBP1	X	X	X	X
PTCHD1	X	X	X	X
PTEN	X	X	X	X
RAI1	X	X	X	X
RPS6KA3	X	X	X	X
SATB2	X	X	X	X
SCN2A	X	X	X	X
SCN8A	X	X	X	X
SETBP1	X	X	X	X
SETD5	X	X	X	X
SHANK3	X	X	X	X
SLC16A2	X	X	X	X
SLC2A1	X	X	X	X
SLC6A8	X	X	X	X
SLC9A6	X	X	X	X
SMARCA2	X	X	X	X
SMARCA4	X	X	X	X
SMC1A	X	X	X	X
STXBP1	X	X	X	X
SYNGAP1	X	X	X	X
TCF4	X	X	X	X
TRIO	X	X	X	X
UBE3A	X	X	X	X
UPF3B	X	X	X	X
WDR45	X	X	X	
TBR1	X	X	X	
KAT6A	X	X	X	
GABRB3	X	X	X	
SCN1A	X	X	X	
GNAO1	X	X	X	
SLC6A1	X	X	X	
DNM1	X	X	X	
KCNB1	X	X	X	
ADNP	X	X	X	
AUTS2	X	X	X	
CDK13	X	X	X	
CHD8	X	X	X	
CNKSR2	X	X	X	
DDX3X	X	X	X	
KIF1A	X	X	X	
MYT1L	X	X	X	
PACS1	X	X	X	
PTPN11	X	X	X	

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
PURA	X	X	X	
TBL1XR1	X	X	X	
TRIP12	X	X	X	
USP9X	X	X	X	
POGZ	X	X	X	
EEF1A2	X	X		X
HNRNP1U	X	X		
ASXL3	X	X		X
BCL11A	X	X		X
BRAF	X	X		X
MAGEL2	X	X		X
WAC	X	X		X
AHDC1	X	X		
ASH1L	X	X		
CHD4	X	X		
CSNK2A1	X	X		
CTCF	X	X		
DNMT3A	X	X		
EBF3	X	X		
GNAI1	X	X		
ITPR1	X	X		
PHIP	X	X		
PUF60	X	X		
TCF20	X	X		
FGFR2	X			
FGFR3	X			
PAX6	X			
MTOR	X			
USP7	X			
TRIM27	X			
WDFY3	X			
UBE3B		X	X	
HDAC4		X	X	X
ARID1A		X	X	X
FGD1		X	X	X
KDM6A		X	X	X
MID1		X	X	X
NIPBL		X	X	X
PHF6		X	X	X
SMARCB1		X	X	X
SMARCE1		X	X	X
SMC3		X	X	X
EFTUD2		X		X
ABCC9				X
KCNJ8				X
RAD21				X
BRPF1		X	X	
PHF21A		X	X	
PHF8		X	X	
RBM10		X	X	
SRCAP		X	X	
TBX1		X	X	
KDM5A			X	
KDM5B		X		
SETD1A		X		
ARHGEF9		X	X	X
ARX		X	X	X
KCNJ10		X	X	X
MBD5		X	X	X
PIGA		X		X
CPA6		X	X	
ADSL		X	X	
ALG13		X	X	
BTD		X	X	
CHRNA4		X	X	
CHRN2		X	X	
FOLR1		X	X	
GABRA1		X	X	
GABRG2		X	X	
KCNQ3		X	X	
KCNT1		X	X	
LG11		X	X	
PCDH19		X	X	
PNKP		X	X	
PRRT2		X	X	
SCN1B		X	X	
SPTAN1		X	X	
TBC1D24		X	X	
WWOX		X	X	
ALDH7A1		X	X	
CSTB		X		
EPM2A		X		
GOSR2		X		
KCTD7		X		
NHLRC1		X		
PIGN		X		

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
PNPO		X		
POLG		X		
PRICKLE1		X		
SCARB2		X		
DEPDC5		X	X	
HCN1		X	X	
NEDD4L		X	X	
SLC13A5		X	X	
SLC35A2		X	X	
KCNC1		X		
KCNH1		X		
STAMBP		X		
STX1B		X		
CHRNA2		X	X	
PLCB1		X	X	
SLC25A22		X	X	
ST3GAL3		X	X	
ST3GAL5		X	X	
SYN1		X	X	
ATIC		X		
PIGO		X		
GABBR2		X	X	
QARS		X	X	
SLC1A1		X	X	
SYNJ1		X	X	
SZT2		X	X	
GABRB2		X		
RORB		X		
ATP6AP2		X	X	X
ATP7A		X	X	X
BCOR		X	X	X
BRWD3		X	X	X
CLCN4		X	X	X
CNTNAP2		X	X	X
DCX		X	X	X
DEAF1		X	X	X
DHCR7		X	X	X
EIF2S3		X	X	X
FMR1		X	X	X
FOXP2		X	X	X
GPC3		X	X	X
HIVEP2		X	X	X
HPRT1		X	X	X
HRAS		X	X	X
KLHL15		X	X	X
L1CAM		X	X	X
MAN1B1		X	X	X
MYCN		X	X	X
NRXN1		X	X	X
NSDHL		X	X	X
NSUN2		X	X	X
OCRL		X	X	X
PLP1		X	X	X
PRPS1		X	X	X
SKI		X	X	X
TAF1		X	X	X
TRAPPC9		X	X	X
TUBA1A		X	X	X
UBE2A		X	X	X
ZEB2		X	X	X
COL4A3BP		X		X
ZBTB18		X		X
CCBE1				X
GNAS				X
GPR56				X
KIF11				X
KIF2A				X
KIF5C				X
LARP7				X
LMAN2L				X
PIGL				X
RBBP8				X
SLC2A17				X
TBCE				X
THOC2				X
TWIST1				X
TWIST2				X
ACAN		X	X	
ACAT1		X	X	
ACSL4		X	X	
ACTB		X	X	
ACTG1		X	X	
ACY1		X	X	
ADA		X	X	
AFF2		X	X	
AGA		X	X	
AH1		X	X	

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
ALDH18A1		X	X	
ALDH3A2		X	X	
ALDH5A1		X	X	
ALDOB		X	X	
ALG1		X	X	
ALG12		X	X	
ALG3		X	X	
ALG6		X	X	
AMT		X	X	
ANK2		X	X	
ANK3		X	X	
AP1S2		X	X	
AP4B1		X	X	
AP4E1		X	X	
AP4M1		X	X	
AP4S1		X	X	
ARFGEF2		X	X	
ARIH1		X	X	
ARSA		X	X	
ASAH1		X	X	
ASL		X	X	
ASPA		X	X	
ASPM		X	X	
ASS1		X	X	
ATP7B		X	X	
ATR		X	X	
BCKDHA		X	X	
BCKDHB		X	X	
BCKDK		X	X	
BMP3		X	X	
BRD3		X	X	
C12orf57		X	X	
CA8		X	X	
CACNA1C		X	X	
CACNA1D		X	X	
CACNA1F		X	X	
CACNG2		X	X	
CAMTA1		X	X	
CASR		X	X	
CBS		X	X	
CC2D1A		X	X	
CCDC22		X	X	
CDH10		X	X	
CDH9		X	X	
CDKSRAP2		X	X	
CDK6		X	X	
CENPJ		X	X	
CEP152		X	X	
CHRNA7		X	X	
CLCN2		X	X	
CLIC2		X	X	
CLN3		X	X	
CLN5		X	X	
CLN6		X	X	
CLN8		X	X	
CNOT3		X	X	
CNTN4		X	X	
CPS1		X	X	
CRBN		X	X	
CSDE1		X	X	
CTSF		X	X	
CUX2		X	X	
CYFIP1		X	X	
CYP11A1		X	X	
CYP27A1		X	X	
CYP2U1		X	X	
D2HGDH		X	X	
DBT		X	X	
DDHD2		X	X	
DISC1		X	X	
DKC1		X	X	
DLG1		X	X	
DNM3		X	X	
DP71		X	X	
DPAGT1		X	X	
DPM1		X	X	
DSCAM		X	X	
EIF4E		X	X	
ELP2		X	X	
EN2		X	X	
EPB41L1		X	X	
ERLIN2		X	X	
FRMPD4		X	X	
FTSJ1		X	X	
FUCA1		X	X	
GABRD		X	X	
GALT		X	X	

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
GAMT		X	X	
GAP43		X	X	
GATM		X	X	
GBA		X	X	
GCDH		X	X	
GCH1		X	X	
GCSH		X	X	
GDI1		X	X	
GFPT2		X	X	
GK		X	X	
GLDC		X	X	
GLUD1		X	X	
GPRASP1		X	X	
GRID2		X	X	
GRIK2		X	X	
GRIP1		X	X	
GTF21		X	X	
GUSB		X	X	
HCCS		X	X	
HCFC1		X	X	
HEPACAM		X	X	
HEXA		X	X	
HGSNAT		X	X	
HMGA2		X	X	
HOXA1		X	X	
HSD17B10		X	X	
IDS		X	X	
IDUA		X	X	
IGF1		X	X	
IGF1R		X	X	
IKBKG		X	X	
IVD		X	X	
KANSL2		X	X	
KATNAL2		X	X	
KCNA1		X	X	
KCNK9		X	X	
KCNMA1		X	X	
KCTD13		X	X	
KLF8		X	X	
LAMC3		X	X	
LAMP2		X	X	
LAS1L		X	X	
LJM1		X	X	
LRP2		X	X	
MACROD2		X	X	
MAGI2		X	X	
MAN2B1		X	X	
MANBA		X	X	
MAOA		X	X	
MAST1		X	X	
MBTPS2		X	X	
MCPH1		X	X	
MED17		X	X	
MED23		X	X	
MET		X	X	
MFSD8		X	X	
MMACHC		X	X	
MSL3		X	X	
MTHFR		X	X	
MTR		X	X	
MUT		X	X	
NAGLU		X	X	
NDE1		X	X	
NDST1		X	X	
NDUFA1		X	X	
NF1		X	X	
NGFR		X	X	
NHS		X	X	
NLGN3		X	X	
NLGN4X		X	X	
NPC1		X	X	
NPC2		X	X	
NRXN2		X	X	
NTNG1		X	X	
OFD1		X	X	
OTC		X	X	
OXTR		X	X	
PAH		X	X	
PAK3		X	X	
PAX5		X	X	
PCCA		X	X	
PCCB		X	X	
PCDH8		X	X	
PDHA1		X	X	
PDHX		X	X	
PEX1		X	X	
PEX12		X	X	

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
PEX6		X	X	
PEX7		X	X	
PHGDH		X	X	
PIGV		X	X	
PMM2		X	X	
POMGNT1		X	X	
PORCN		X	X	
POU1F1		X	X	
PPT1		X	X	
PRKRA		X	X	
PRODH		X	X	
PRSS12		X	X	
PSAP		X	X	
PTS		X	X	
RAB39B		X	X	
RELN		X	X	
RLIM		X	X	
RLIM		X	X	
RNF135		X	X	
RORA		X	X	
RPL10		X	X	
SEMA5A		X	X	
SETD2		X	X	
SGSH		X	X	
SHANK1		X	X	
SHANK2		X	X	
SLC17A5		X	X	
SLC25A13		X	X	
SLC25A15		X	X	
SLC35A3		X	X	
SLC35C1		X	X	
SLC46A1		X	X	
SLC6A4		X	X	
SM5		X	X	
SOBP		X	X	
SOX11		X	X	
SOX3		X	X	
SOX5		X	X	
SOX7		X	X	
SRD5A3		X	X	
STAG1		X	X	
STIL		X	X	
STX1A		X	X	
SYP		X	X	
TAOK2		X	X	
TECR		X	X	
TIMM8A		X	X	
TMLHE		X	X	
TNK2		X	X	
TPP1		X	X	
TPP1		X	X	
TSC1		X	X	
TSC2		X	X	
TSPAN7		X	X	
TUBB2B		X	X	
TUSC3		X	X	
UBR1		X	X	
UPB1		X	X	
USP27X		X	X	
VCP		X	X	
VLDLR		X	X	
VPS13B		X	X	
WDR62		X	X	
WHSC1		X	X	
WNK3		X	X	
YWHAE		X	X	
ZBTB20		X	X	
ZC3H14		X	X	
ZDHC9		X	X	
ZFH4		X	X	
ZMYND11		X	X	
ZNF292		X	X	
ZNF711		X	X	
AGO1			X	
ANKH			X	
ARHGFG			X	
ATXN3L			X	
CACNA1G			X	
DLG2			X	
DOCK8			X	
FTCD			X	
HAL			X	
IGBP1			X	
KIRREL3			X	
PAFAH1B1			X	
SHROOM4			X	
SRGAP3			X	

	Test DI117	Test TC-520 (CHU Strasbourg)	Test TC-456 (CHU Strasbourg)	Test DI-142 (CHU Marseille)
SYNE1			X	
UROCI			X	
WDR45B			X	
ZC3H15			X	
ZMYM3			X	
ZNF526			X	
ZNF81			X	
ADAT3		X		
ADCK3		X		
ASXL1		X		
BRWD1		X		
BTF3		X		
CHAMP1		X		
CHD5		X		
CIC		X		
CLTC		X		
COL4A3BP		X		
CUL3		X		
DNMT3B		X		
DPP6		X		
DPYD		X		
GNB1		X		
HCN2		X		
HERC1		X		
KMT2C		X		
LRRC43		X		
NALCN		X		
NECAP1		X		
PCGF2		X		
PGAP2		X		
PGAP3		X		
PPM1D		X		
PPP2R1A		X		
PPP2R5D		X		
QRICH1		X		
SET		X		
SLC1A4		X		
SMAD4		X		
SUV420H1		X		
TAB2		X		
TBCK		X		
THOC6		X		
TRPC5		X		
UNC13		X		
UNC80		X		
ZC4H2		X		

1. PCR Multiplex : Date : xx/xx/2018-XXX

Mix 1 PCR	Volume (µl) + Sample ID	Volume Mix (µl)			
Nombre de tubes	1	8			
5 X Ion Amplicon HiFi Mix (bouchon rouge)	5	40			
H2O nuclease Free	2,5	20			
20X ion Amplicon Sample ID	1	8			
Volume final	8,5	68			
Volume par pool	3	30	Répartir dans 2 tubes 1,5ml pour chaque pool		
Mix 2 PCR par pool	Volume (µl) Pool 1	Volume Mix (µl)	Volume (µl) Pool 2	Volume Mix (µl)	
Nombre de tubes	1	10	1	10	
Mix 1 PCR	3	30	3	30	
2X ion Amplicon Primer Pool	5	50	5	50	
Volume par microtube		8		8	
ADN		2		2	
Volume final par microtube		10		10	
Cycles PCR	Temp (°C)	Temps	Cycles		
Activation de l'enzyme	99°C	2 min			
Dénaturation	99°C	15 s	X 15 cycles		
Hybridation/Elongation	60°C	8 min			
	10°C	∞			

Rq : 1108 amorces dans un pool et 1116 amorces dans l'autre pool

Important : travailler dans la glace et en barrettes

Régler la pipette à 8 µl et pipeter « up and down » 5 fois pour mélanger.

Thermocycleur : Veriti LAB5509, 067430

Stockage une nuit à 4°C des librairies amplifiées :

oui non

Stockage à -20°C des librairies amplifiées :

oui non

Dilution des barcodes (pré-pcr) :

Date :

lot :

Préparation d'un mix pour 20 réactions	
Ion P1 Adaptateur	10 µl
Ion Xpress Barcode X	10 µl
H2O Nuclease Free	20 µl
Total	40 µl

Utilisation de barcodes déjà dilués :

oui non

Si oui, date de dilution :

Stockage à -20°C des barcodes dilués en post-PCR :

oui non

2. Digestion des primers et ligation des adaptateurs :

Date : xx/xx/2018-XXX

Important : travailler sous hotte UV de post-PCR

Transférer les échantillons en plaque à l'aide d'une pipette multicanaux : pour chaque échantillon, transférer 10µl de chaque pool dans un même puits. Le volume final est de 20µl.

2.1 Digestion des primers :

- Ajouter **2 µl de Fupa** (bouchon marron) à chacun des échantillons. Volume final 22 µl
- Vortexer puis centrifuger brièvement ou régler la pipette sur 15 µl et pipeter « up and down » 5 fois

Technique validée sur les thermocycleurs Life Technologies : LAB 5509 (VERITI), LAB 5090 (9700), LAB 3605 (2720), 5137 (2700), 5507 (2720), 5508 (2720), 5604 (2720), 067430 (VERITI).

Cycles PCR	Temp (°C)	Temps
1	50°C	10 min
2	55°C	10 min
3	60°C	20 min
4	10°C	jusqu'à 1h

2.2 Ligation des adaptateurs

Si présence d'un précipité dans la swich solution, pipeter up and down à température ambiante

- Ajouter les réactifs suivant dans chaque puits :

Réactif	Volume
Swich Solution (hexafluor)	4 µl
Mix de Barcode dilué (1 barcode différent / patient)	2 µl
Volume total (incluant les 22 µl d'amplicons digérés)	28 µl

- Ajouter **2 µl** de ligase dans chaque puits
- Vortexer puis centrifuger brièvement ou régler la pipette sur 20 µl et pipeter « up and down » 5 fois

Cycles PCR	Temp (°C)	Temps
1	22°C	30 min
2	72°C	10 min
3	10°C	jusqu'à 1h

Technique validée sur les thermocycleurs Life Technologies : LAB 5509 (VERITI), LAB 5090 (9700), LAB 5605 (2720), 5137 (2700), 5507 (2720), 5508 (2720), 5604 (2720), 067430 (VERITI).

Stockage à -20°C des librairies : *oui* *non*

3. Purification des librairies :

Date : xx/xx/2018-XXX

- Important :**
- travailler sous hotte UV de post-PCR
 - **Sortir le réactif Ampure au minimum une demi-heure avant utilisation, bien le vortexer et pipeter doucement**

Préparer 10 ml d'éthanol à 70% extemporanément : 7ml d'éthanol absolu (conditionnement 100 ml) + 3 ml d'eau ultrapure du jour *lot Ethanol :*

3.1 Purification

- Ajouter **45 µl** d'Ampure à chaque librairie. Pipeter « up and down » 5 fois
 - Incuber **5 min** à température ambiante sur la paillasse
 - Incuber **2 min** sur le barreau magnétique (ou jusqu'à ce que la solution soit claire)
 - **Éliminer** le surnageant sans toucher les billes
 - Ajouter **150 µl** d'éthanol à 70%
 - Décaler la plaque sur le barreau magnétique **2 fois** pour rincer les billes
 - **Éliminer** le surnageant sans toucher les billes
 - Recommencer **une fois** le lavage à l'éthanol
 - Laisser l'éthanol s'évaporer, **4 min** à température ambiante (vérifier l'état d'évaporation)
- Attention de ne pas laisser trop sécher !!!!**

3.2 Réamplification

- Enlever la plaque du barreau magnétique
- **Préparation d'un master mix (prévoir n+2 tubes):**

Mix PCR	Volume (µl)	Volume (µl)
Nombre de tubes	1	10
Platinum PCR superMix High Fidelity (bouchon noir)	50	500
Library Amplification Primer Mix (bouchon blanc)	2	20
Volume à distribuer par tube		52

- Distribuer **52µl** de mix par tube
- Vortexer puis centrifuger brièvement ou pipeter « up and down » 5 fois
- Incuber **2 min** à température ambiante
- Incuber **2 min** sur le barreau magnétique
- **Transférer le surnageant** dans de nouveaux puits sans toucher les billes
- Enlever la plaque du barreau magnétique
- Couvrir la plaque et mettre dans le thermocycleur

N° Fiche : FT XX

XX/XX/2018 Opérateur : XXX

Cycles PCR	Temp (°C)	Temps	Cycles
Activation de l'enzyme	98°C	2 min	
Dénaturation	98°C	15 s	X 5
Hybridation/Elongation	64 °C	1 min	cycles
	10°C	∞	

Technique validée sur les thermocycleurs Life Technologies :
LAB 5509 (VERITI), LAB 5090 (9700), LAB 5605 (2720), 5136
(2700), 5137 (2700), 5507 (2720), 5508 (2720), 5604 (2720)
Thermocycleurs utilisés :

Stockage à -20°C des librairies :
oui non

4. Seconde purification des librairies :

Date : xx/xx/2018-XXX

Important : travailler sous hotte UV de post-PCR

Sortir le réactif Ampure au minimum une demi-heure avant utilisation, bien le vortexer et pipetter doucement

Préparer 10 ml d'éthanol à 70% extemporanément, 7ml d'éthanol absolu (conditionnement 1.00ml) + 3 ml d'eau milliQ du jour

4.1 1^{ère} purification :

- Ajouter **25µl** d'Ampure à chaque tube contenant environ 50µl
- Pipeter « up and down » 5 fois
- Incuber **5 minutes** sur la paillasse à température ambiante
- Placer la plaque **5 minutes** sur le barreau magnétique (ou jusqu'à ce que la solution soit complètement claire)
- **Récupérer le surnageant** dans de nouveaux puits sans toucher les billes
- Enlever la plaque du barreau magnétique

4.2 2^{ème} purification

- Ajouter **60µl** d'Ampure au surnageant récupéré
 - Pipeter « up and down » 5 fois
 - Incuber **5 minutes** à température ambiante
 - Incuber **3 minutes** sur le barreau magnétique
 - **Éliminer le surnageant**
 - Ajouter **150µl** d'éthanol à 70% préparé extemporanément
 - Décaler la plaque sur le barreau magnétique **2 fois** pour rincer les billes
 - **Éliminer** le surnageant
 - Ajouter **150µl** d'éthanol à 70% préparé extemporanément
 - Décaler la plaque sur le barreau magnétique **2 fois** pour rincer les billes
 - **Éliminer** le surnageant
 - Laisser l'éthanol s'évaporer, **4 min** à température ambiante (vérifier l'état d'évaporation)
- Attention de ne pas laisser trop sécher !!!!**

4.3 Elution

- Enlever la plaque du barreau magnétique
- Ajouter **50µl** de low TE et homogénéiser les billes
- Pipeter « up and down » 5 fois
- Incuber **2 min** à température ambiante
- Incuber **2 min** sur le barreau magnétique
- **Récupérer le surnageant** dans un tube 1.5 ml Lo-bind correctement identifié (étiquette zebra, modèle [\Srv-pu-gene\genetique\QUALITE\5-Exigences techniques\5.5 Procédures analytiques\3-Enregistrement\ENR 085 réactifs BM\TRAMES ETIQUETTES BM\Etiquettes Librairies Ion Torrent.doc](#))

N°ADN	N°ADN
DI-44-P1	DI-44-P2
Barcode X	Barcode X
Date -opérateur	Date- Opérateur
Cc°	Cc°

Annexe 3 : Feuille de travail récapitulant le protocole utilisé pour le test DI117

N° Fiche : FT XX

XX/XX/2018

Opérateur : XXX

**Gènes DI : Préparation des librairies avec la méthode SureSelectQXT (Agilent)
pour un séquençage sur technologie Illumina**

Panel DI117

1. Préparation des échantillons (pré-PCR)

Liste des échantillons :

Kit reçu le :

Ouvert le :

N° Fiche : FT XX

XX/XX/2018

Opérateur : XXX

1^{er} dosage Qubit dsDNA BR ASSAY (Thermo Fisher Scientific)

Date :

Dosage avec 1µl d'ADN

N° échantillon	N°ADN	Boîte	Concentration Nanodrop (ng/µl)	Dosage Qubit (ng/µl)	Quantité d'ADN nécessaire pour avoir 525 ng (µl)	Volume d'H2O nécessaire pour une dilution à 35 ng/µl (Vf = 15µl)
1						
2						
3						
4						
5						
6						
7						
8						

Dilution des échantillons à 35 ng/µl en eau « nuclease-free » dans un tube LoBind.

2^{ème} dosage Qubit dsDNA BR ASSAY (Thermo Fisher Scientific)

Date :

Dosage avec 4µl d'ADN dilué à 35 ng/µl

N° échantillon	N°ADN	Dosage Qubit (ng/µl)	Volume d'ADN nécessaire pour avoir 500ng (µl)	Volume d'H2O nécessaire pour avoir une dilution à 25ng/µl (Vf = 20µl)	Si cc° < 25ng/µl, volume d'ADN à utiliser (µl)
1					
2					
3					
4					
5					
6					
7					
8					

Dilution des échantillons à 25 ng/µl en eau « nuclease-free » dans un tube LoBind.

2. Fragmentation (pré-PCR et post-PCR)

J1 Date ,

a. Fragmentation (pré-PCR)

Préprogrammer le thermocycleur (programme « Fragmentation »)
Lancer le programme et mettre sur pause lorsque la température de 45°C est atteinte

Thermocycleur utilisé :

Cycles PCR	Temp (°C)	Temps
1	45°C	10 min
2	4°C	1 min
3	4°C	∞

Réactions de fragmentation

Important : Réaliser les réactions de fragmentation dans la glace

- Vortexer vigoureusement les tubes SureSelect QXT Enzyme Mix ILM et SureSelect QXT Buffer
- Annoter et mettre une plaque de PCR ou une barrette de tubes dans la glace.
- Bien vortexer et ajouter **17 microL** de **SureSelect QXT Buffer** dans chaque puit.
- Ajouter **2 microL** de chaque échantillon d'**ADNg normalisé à 25 ng/microL** dans leur puit attitré.
ATTENTION : Pour distribuer l'ADN, il est important de placer la pointe de la pipette au fond du puit.
- Vortexer et ajouter **2 microL** de **SureSelect QXT Enzyme Mix ILM** dans chaque puit, puis **pipetter up and down 8 à 10 fois**.
ATTENTION : Pour distribuer l'Enzyme Mix, il est important de placer la pointe de la pipette au fond du puit.
- Sceller les puits
- Centrifuger brièvement (pulse)
- Mélanger minutieusement en vortexant la plaque ou la barrette de tube à vitesse élevée pendant 20 secondes.
- Centrifuger brièvement (pulse)
- Puis placer immédiatement la plaque ou la barrette de tube dans le thermocycleur et appuyer sur le bouton Play (pour reprendre le programme mis sur pause).

- Pendant les 10 minutes d'incubation des échantillons dans le thermocycleur, vortexer vigoureusement les billes AMPure XP à vitesse élevée
- Dès que le thermocycleur a fini sa minute d'incubation à 4°C, placer immédiatement les échantillons dans la glace

- Ajouter **32 microL** de **1X SureSelect QXT Stop Solution** (contenant 25% d'éthanol) dans chaque puit.
- Sceller les puits avec de nouveaux bouchons
- Vortexer à vitesse élevée pendant 5 secondes
- Centrifuger brièvement (pulse)
- Laisser incuber à température ambiante pendant 1 minute (et passer directement à l'étape suivante)

Purification de la librairie liée aux adaptateurs avec des billes AMPure XP

- Ajouter **52 microl** de **billes magnétiques** dans chaque puit
- Sceller les puits avec de nouveaux bouchons
- Vortexer pendant 5 secondes
- Centrifuger brèvement (pulse)
- Vérifier que les billes magnétiques sont restées en suspension homogène dans les puits
- Incuber 5 minutes à température ambiante

- Mettre sur le portoir magnétique
- Attendre l'éclaircissement de la solution (environ 3 à 5 minutes), puis enlever délicatement le surnageant
- Ajouter **200 microl** d'**éthanol à 70%**
- Incuber 1 minute
- Enlever le surnageant
- Ajouter **200 microl** d'**éthanol à 70%**
- Incuber 1 minute
- Enlever le surnageant

- Laisser sécher les échantillons sur un thermocycleur (en laissant le couvercle ouvert) à 37°C pendant 1 à 3 minutes
Attention : ne pas laisser trop sécher les échantillons

- Ajouter **11 microl** d'**eau ultrapure** dans chaque puit
- Sceller les puits avec de nouveaux bouchons
- Bien vortexer
- Centrifuger brèvement (pulse)
- Incuber 2 minutes à température ambiante

- Mettre sur le portoir magnétique
- Attendre 2 minutes ou jusqu'à l'éclaircissement de la solution
- **Récupérer** le surnageant (environ 10 microl) dans des puits d'une nouvelle plaque/barrette
- Mettre dans la glace

b. Amplification de la librairie liée aux adaptateurs (post-PCR)

- Décongeler puis vortexer les réactifs suivants :
 - o Herculase II Fusion DNA Polymerase
 - o Herculase II 5x Reaction Buffer
 - o 100 mM dNTP Mix (contenant 25 mM de chaque dNTP)
 - o SureSelect QXT Primer Mix
 - o DMSO
- Conserver tous les réactifs sauf le DMSO dans la glace.
- Préparer le volume nécessaire de mix pour la réaction de PCR pré-capture, dans de la glace, et bien le vortexer

Réactifs	Volume pour 1 réaction (μ l)	Volume pour n+1 réaction (μ l)
Eau ultrapure	25	
Herculase II 5x Reaction Buffer	10	
100 mM dNTP Mix	0,5	
DMSO	2,5	
SureSelect QXT Primer Mix	1	
Herculase II Fusion DNA Polymerase	1	
Total	40	

- Ajouter 40 µL du mix préparé pour la réaction de PCR pré-capture à chaque échantillon de librairie purifiée.
- Sceller les puits avec de nouveaux bouchons
- Vortexer doucement pendant 5 secondes
- Centrifuger brièvement (pulse)
- Placer la plaque ou la barrette de tube dans le thermocycleur (en fermant le couvercle) et lancer le programme

Thermocycleur utilisé :

Cycles PCR	Temp (°C)	Temps
1	68°C	2 min
2	98°C	2 min
3 = X 8	98°C	30 sec
	57°C	30 sec
	72°C	1 min
4	72°C	5 min
5	4°C	∞

Purification de la librairie amplifiée avec des billes AMPure XP

- Ajouter **50 microL** de **billes magnétiques** dans chaque puit
- Sceller les puits avec de nouveaux bouchons
- Vortexer pendant 5 secondes
- Centrifuger brièvement (pulse)
- Vérifier que les billes magnétiques sont restées en suspension homogène dans les puits
- Incuber 5 minutes à température ambiante
- Mettre sur le portoir magnétique
- Attendre l'éclaircissement de la solution (environ 3 à 5 minutes), puis enlever délicatement le surnageant
- Ajouter **200 microL d'éthanol à 70%**
- Incuber 1 minute
- Enlever le surnageant
- Ajouter **200 microL d'éthanol à 70%**
- Incuber 1 minute
- Enlever le surnageant
- Laisser sécher les échantillons sur un thermocycleur (en laissant le couvercle ouvert) à 37°C pendant 1 à 3 minutes (Attention : ne pas laisser trop sécher les échantillons)
- Ajouter **13 microL d'eau ultrapure** dans chaque puit
- Sceller les puits avec de nouveaux bouchons
- Bien vortexer
- Centrifuger brièvement (pulse)
- Incuber 2 minutes à température ambiante
- Mettre sur le portoir magnétique
- Attendre 2 minutes ou jusqu'à l'éclaircissement de la solution
- **Récupérer** le surnageant (environ 13 microL) dans des puits d'une nouvelle plaque/barrette
- Mettre dans la glace

Stop point : stockage à 4°C une nuit ou -20°C

c. Vérification qualité et quantité de la librairie (post-PCR)

Bioanalyser : <http://srv-kalilab/MP-F0-211>

Le pic doit être positionné entre 245 et 325 pb.

Stop point : stockage à 4°C une nuit ou -20°C

3. Hybridation et capture (post-PCR)

J2 Date :

a. Aliquotage des échantillons d'ADN préparés pour l'hybridation

Préparer 500 à 750 ng de librairies d'ADN dans un volume final de 12 µl en H₂O (*en barrettes*).

Barcode	dosage Qubit BR		750 ng/12µl		Volume final
	ADN	[]° ng/µl	Librairie	H ₂ O	12
1					
2					
3					
4					
5					
6					
7					
8					

b. Hybridation des échantillons d'ADN à la librairie de capture Date :

Cette étape utilise les réactifs du kit de réactifs SureSelectQXT suivants :

- SureSelect Fast Hybridization Buffer à décongeler à T° ambiante
 - SureSelect QXT Fast Blocker Mix à décongeler dans de la glace
 - SureSelect RNase Block à décongeler dans de la glace
 - Capture Library à décongeler dans de la glace
- Vortexer et centrifuger brièvement (pulse) chaque réactif
 - Ajouter 5 µL de SureSelect QXT Fast Blocker Mix dans chaque puit de librairie liée aux adaptateurs, puis pipetter up and down 8 à 10 fois
 - Sceller les puits avec de nouveaux bouchons
 - Bien Vortexer pendant 5 secondes
 - Centrifuger brièvement (pulse)
- Placer la plaque ou la barrette de tubes dans le thermocycleur (en fermant le couvercle) et lancer le programme

Thermocycleur utilisé :

Cycles PCR	Temp (°C)	Temps
1	95°C	5 min
2	65°C	10 min
3	65°C	1 min PAUSE
4 = X 60	65°C	1 min
	37°C	3 sec
5	65°C	∞

- Mettre le thermocycleur sur pause durant la 3^{ème} phase du programme
- Préparer une solution contenant 25% de SureSelect RNase Block

Réactifs	Volume pour 1 réaction (μ l)	Volume pour n+1 réaction (μ l)
Eau ultrapure	1,5	
SureSelect RNase Block	0,5	
Total	2	

- Bien Vortexer et conserver dans de la glace
- Préparer le mix pour l'hybridation avec les sondes de capture et le garder à T° ambiante

Réactifs	Volume pour 1 réaction (μ l)	Volume pour n+1 réaction (μ l)
25% RNase Block solution	2	
Capture Library <3 Mb	2	
SureSelect Fast Hybridization Buffer	6	
Eau ultrapure	3	
Total	13	

- Bien Vortexer pendant 5 secondes
- Centrifuger brièvement (pulse)
- Réaliser **immédiatement** l'étape suivant
- Avec le thermocycleur sur pause et en gardant la plaque/barrette de tubes dans le thermocycleur, ajouter 13 μ L du mix pour l'hybridation avec les sondes de capture
- Bien mélanger en pipettant up and down 8 à 10 fois
- Sceller les puits avec de nouveaux bouchons
- Bien Vortexer pendant 5 secondes
- Centrifuger brièvement (pulse)
- Remettre la plaque/barrette de tubes dans le thermocycleur et appuyer sur le bouton Play (pour reprendre le programme mis sur pause).

d. Préparation billes magnétiques recouvertes de Streptavidin T1

- Vortexer vigoureusement les billes magnétiques Dynabeads MyOne Streptavidin T1
- Annoter une nouvelle plaque de PCR ou une nouvelle barrette de tubes
- Ajouter 50 μ L de billes resuspendues dans chaque puit
- Laver les billes :
 - Ajouter 200 μ L de SureSelect Binding Buffer dans chaque puit
 - Mélanger en pipettant up and down 8 à 10 fois
 - Mettre sur le portoir magnétique
 - Attendre au moins 5 minutes ou jusqu'à l'éclaircissement de la solution
 - Enlever le surnageant
 - Répéter ces étapes pour laver 3 fois les billes
- Resuspendre les billes dans 200 μ L de SureSelect Binding Buffer

f. Capture de l'AND hybridé en utilisant les billes magnétiques coatées de streptavidine

- Retirer les échantillons du thermocycleur et les conserver à T° ambiante
- Transférer à l'aide d'une pipette multicanaux la totalité du volume des échantillons (environ 30 µL) dans les puits contenant 200 µL de billes coatées de streptavidine préalablement lavées
- Sceller les puits avec de nouveaux bouchons
- Mettre la plaque de PCR/barrette de tubes dans un agitateur (à 1800 rpm), pendant 30 minutes à T° ambiante

- Pendant ce temps, réchauffer le Wash Buffer 2 à 65°C :
 - Mettre 200 µL de Wash Buffer 2 par puit d'une nouvelle plaque de 96 puits ou d'une nouvelle barrette de tubes (prévoir 3 puits pour chaque échantillon d'ADN)
 - Sceller les puits avec de nouveaux bouchons
 - Mettre la plaque/barrette de tubes dans le thermocycleur (en fermant le couvercle) réglé à 65°C (les laisser dans le thermocycleur jusqu'à utilisation)

- Une fois les 30 minutes écoulées, collecter le liquide au fond des tubes par la méthode appropriée (manuellement si plaque ou par centrifugation brève si barrette de tubes)
- Mettre la plaque de PCR/barrette de tubes sur le portoir magnétique
- Attendre 1 minutes ou jusqu'à l'éclaircissement de la solution
- Enlever le surnageant

- Resuspendre les billes dans 200 µL de SureSelect Wash Buffer 1 en pipettant up and down 8 à 10 fois
- Sceller les puits avec de nouveaux bouchons
- Bien Vortexer pendant 8 secondes
- Collecter le liquide au fond des tubes par la méthode appropriée (manuellement si plaque ou par centrifugation brève si barrette de tubes)
- Mettre la plaque/barrette de tubes sur le portoir magnétique
- Attendre 1 minutes ou jusqu'à l'éclaircissement de la solution
- Enlever le surnageant

- Enlever la plaque/barrette du portoir magnétique et la laisser à T° ambiante
- Laver les billes avec le Wash Buffer 2
 - Resuspendre les billes dans 200 µL de Wash Buffer 2 préchauffé à 65°C
 - Pipetter up and down au moins 10 fois
 - Sceller les puits avec de nouveaux bouchons
 - Bien Vortexer pendant 5 secondes
 - Collecter le liquide au fond des tubes par la méthode appropriée (manuellement si plaque ou par centrifugation brève si barrette de tubes)
 - Incuber les échantillons 10 minutes à 65°C dans le thermocycleur (en fermant le couvercle)
 - Mettre la plaque/barrette de tubes sur le portoir magnétique à T° ambiante
 - Attendre 1 minutes ou jusqu'à l'éclaircissement de la solution
 - Enlever le surnageant
 - Répéter ces étapes pour laver 3 fois les billes
- Centrifuger brièvement
- Mettre sur le portoir magnétique

- Enlever tout le surnageant
- Ajouter 23 µL d'eau ultrapure à chaque puit
- Conserver la plaque/barrette de tubes dans de la glace

4. Indexation des échantillons

12

a. Amplification et indexation *plaque*

Date :

Rq ; Préparer une réaction d'amplification avec indexation propre à chaque librairie

		1	2
		PSi13	PSi14
A	P7i1		
B	P7i2		
C	P7i3		
D	P7i4		
E	P7i5		
F	P7i6		
G	P7i7		
H	P7i8		

- Préparer le volume nécessaire de mix pour la réaction de PCR, dans de la glace.
- Bien vortexer

Réactifs	Volume pour 1 réaction (µl)	Volume pour n+1 réaction (µl)
Eau ultrapure	13,5	
Herculase II 5× Reaction Buffer	10	
100 mM dNTP Mix	0,5	
Herculase II Fusion DNA Polymerase	1	
Total	25	

- Ajouter 25 µL de mix pour la réaction de PCR à chaque échantillon de librairie capturée
- Ajouter 1 µL de l'amorce doublement indexée appropriée P7 (P7 i1 à P7 i12) à chaque puit
- Ajouter 1 µL de l'amorce doublement indexée appropriée P5 (P5 i13 à P5 i20) à chaque puit
- Bien mélanger en en pipettant up and down
- Mettre la plaque de PCR/barrette de tubes dans le thermocycleur et lancer le programme d'amplification de PCR

Thermocycleur utilisé :

Cycles PCR	Temp (°C)	Temps
1	98°C	2 min
2 = X 12	98°C	30 sec
	58°C	30 sec
	72°C	1 min
3	72°C	5 min
4	4°C	∞

- Centrifuger brièvement
- Mettre sur le portoir magnétique à T° ambiante
- Attendre 2 minutes ou jusqu'à l'éclaircissement de la solution
- **Récupérer** le surnageant (environ 50 µL) dans des puits d'une nouvelle plaque/barrette

c. Purification des librairies capturées avec des billes AMPure XP

- Ajouter **60 microL** de **billes magnétiques** dans chaque puit
 - Sceller les puits avec de nouveaux bouchons
 - Vortexer pendant 5 secondes
 - Centrifuger brièvement (pulse)
 - Vérifier que les billes magnétiques sont restées en suspension homogène dans les puits
 - Incuber 5 minutes à température ambiante

 - Mettre sur le portoir magnétique
 - Attendre l'éclaircissement de la solution (environ 3 à 5 minutes), puis enlever délicatement le surnageant
 - Ajouter **200 microL d'éthanol à 70%**
 - Incuber 1 minute
 - Enlever le surnageant
 - Ajouter **200 microL d'éthanol à 70%**
 - Incuber 1 minute
 - Enlever le surnageant

 - Laisser sécher les échantillons sur un thermocycleur (en laissant le couvercle ouvert) à 37°C pendant 1 à 3 minutes
- Attention : ne pas laisser trop sécher les échantillons

- Ajouter **25 microL d'eau ultrapure** dans chaque puit
- Sceller les puits avec de nouveaux bouchons
- Bien vortexer
- Centrifuger brièvement (pulse)
- Incuber 2 minutes à température ambiante

- Mettre sur le portoir magnétique
- Attendre 2 minutes ou jusqu'à l'éclaircissement de la solution
- **Récupérer** le surnageant (environ 25 microL) dans des tubes LoBind.

Stop point : stockage à -20°C

d. Vérification qualité et quantité des librairie d'ADN indexées

Bioanalyser : <http://srv-kalilab/MP-F0-211>

Le pic doit être positionné entre 325 et 450 pb.

Stop point : stockage à 4°C une nuit ou -20°C

Barcode	ADN	[]° nM	[]° ng/μl	facteur conversion nM->ng/μl	commentaire
1					
2					
3					
4					
5					
6					
7					
8					

e. Quantification Qubit dsDNA BR ASSAY (Thermo Fisher Scientific)

Préparation de la solution de travail Qubit :

nb échantillons + 2 std	1	10
Qubit™ dsDNA BR/HS buffer (μl)	1	10,8
Qubit™ dsDNA BR/HS reagent (μl)	199	2149,2
Volume total (μl)	200	2160

Volume échantillon (librairie pure) = 2 μl

ADN	[]° ng/μl dosage 1	[]° ng/μl dosage 2	[]° ng/μl moyenne	[]° nM moyenne	Vol Librairie (μl)	Vol diluant (μl)	Vol Final (μl)	cc° Finale (nM)
							0,0	4
							0,0	4
							0,0	4
							0,0	4
							0,0	4
							0,0	4
							0,0	4
							0,0	4

Stop point : stockage à 4°C une nuit ou -20°C

f. Préparation séquençage

J3 Date :

a. Pool librairies**Re-Dosage Qubit des librairies + dosage du pool**

volume échantillon (librairie 4 nM) = 2 µl

Date :

Barcode	ADN	[]° ng/µl <i>dosage 1</i>	[]° ng/µl <i>dosage 2</i>	[]° ng/µl <i>dosage 3</i>	[]° ng/µl <i>moyenne</i>	[]° nM <i>moyenne</i>	Vol. librairie <i>(4nM final)</i>	Vol. H2O <i>(20µl final)</i>
1						0,0	#DIV/0!	#DIV/0!
2						0,0	#DIV/0!	#DIV/0!
3						0,0	#DIV/0!	#DIV/0!
4						0,0	#DIV/0!	#DIV/0!
5						0,0	#DIV/0!	#DIV/0!
6						0,0	#DIV/0!	#DIV/0!
7						0,0	#DIV/0!	#DIV/0!
8						0,0	#DIV/0!	#DIV/0!
Pool 4nM	<i>(5µl+5µl+...)</i>					0,0	#DIV/0!	#DIV/0!

Pool de toutes les librairies à 4 nM (utilisation de 5 µl de chaque soit 5*8=40 µl au total)

Stop point : rajouter du Tween 20 à 0,1 % final

Stockage -20°C à court terme

b. Préparation séquençage

Chargement 15 pM final de librairies poolées et 1 % PhiX

- 1- Préparation extemporanée NaOH 0,2 M :

Dilution 1 NaOH 1 M	10 µl NaOH 10M + 90 µl H2O
Dilution 2 NaOH 0,2 M	50 µl NaOH 1M + 200 µl H2O

- 2- Dénaturation à la soude

Pool librairies 4 nM	10 µl
NaOH 0,2 M	10 µl
<i>volume final pool Librairie à 2 nM</i>	20 µl

- Vortexer, centrifuger brièvement et laisser incuber 5 min
- Préparer un volume donné selon le tube HT1

- 3- Préparation du pool de librairie à 15pM final dans le tampon HT1 frais (fourni avec la cartouche)

Concentration finale souhaitée librairie	15 pM
Volume Librairie dénaturée 2 nM (étape 2) à ajouter	7,5 µl
Volume HT1	992,5 µl
<i>Volume final</i>	<i>1000 µl</i>

4- Préparation tube à séquencer avec Spike in PhiX

% final souhaité PhiX	1 %
Concentration librairie (étape 3)	15 pM
Volume 20 pM PhiX	7 µl
Volume librairie 15 pM lib (étape 3)	693 µl
Volume Total	700 µl

Utilisation de 600 µl pour charger le run

5- Préparation de la cartouche et des amorces

Laisser décongeler la cartouche 1h dans un bain d'eau froide du robinet, l'agiter et ensuite préparer les primers customs (amorces) selon le protocole.

! Attention il faut cocher les customs primers sur la sample sheet !

Trouer la cartouche et récupérer les solutions selon le tableau ci-dessous pour les mettre dans un tube 1.5 ml correctement annoté. Respecter les volumes mentionnés du tableau.



Table 27 MiSeq platform custom sequencing primer preparation

Sequencing Read	Volume of SureSelect ^{OX} T Primer	Volume of Illumina TruSeq Primer	Total Volume	Final Cartridge Position
Read 1	3 µl SureSelect OXT Read Primer 1 (brown cap)	597 µl HP10 (well 12)	0.6 ml	well 18
Index	3 µl SureSelect OXT Index 1 Read Primer (clear cap)	597 µl HP12 (well 13)	0.6 ml	well 19
Read 2	3 µl SureSelect OXT Read Primer 2 (black cap)	597 µl HP11 (well 14)	0.6 ml	well 20

Charger la cartouche avec 600µl du tube obtenu en 4

Annexe 4 : Tableau récapitulatif des 117 gènes inclus dans le test DI117

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
ADNP	20	AD	Activity-dependent neuroprotector homeobox	611386	Syndrome de Helsmoortel-van der Aa	615873	DI, TSA, dysmorphie faciale
AHDC1	1	AD	AT-hook DNA-binding motif-containing protein 1	615790	Syndrome de Xia-Gibbs	615829	retard psychomoteur avec retard important de langage, retard de croissance, hypotonie, dysmorphie faciale
ANKRD11	16	AD	Ankyrin repeat domain-containing protein 11	611192	Syndrome KBG	148050	Macrodontie des incisives centrales supérieures, petite taille, anomalies squelettiques, dysmorphie faciale, retard psychomoteur
ARID1B	6	AD	AT-rich interactive domain 1B	614556	Syndrome de Coffin-Siris	135900	retard psychomoteur, DI, microcéphalie, traits grossiers du visage, cheveux épars, problèmes d'alimentation, infections respiratoires récurrentes, petite taille, retard de croissance, ongles hypoplasiques au V ^{ème} doigt et au V ^{ème} orteil
ASH1L	1	AD	ASH1 (absent, small, or homeotic)-like (Drosophila)	607999	Retard mental AD 52	617796	DI-NS
ASXL3	18	AD	Additional sex combs-like 3 (Drosophila)	615115	Syndrome de Bainbridge-Ropers	615485	retard de développement, DI sévère, trouble sévère du langage, hypotonie, difficultés alimentaires, petite taille, dysmorphie faciale
ATRX	X	XL	Alpha thalassemia/mental retardation syndrome X-linked	300032	Syndrome d'alpha-thalassemie/retard mental / Syndrome retard mental-faciès hypotonique	301040 / 309580	DI sévère, retard de développement plus ou moins sévère, dysmorphie faciale, anomalies d'Hb H dans les érythrocytes plus ou moins légères, anomalies génitales, biais fréquent d'inactivation du X chez les femmes porteuses
AUTS2	7	AD	Autism susceptibility candidate 2	607270	Retard mental AD 26	615834	Retard de développement, microcéphalie, retard de croissance, dysmorphie faciale, avec ou sans TSA
BCL11A	2	AD	B-cell CLL/lymphoma 11A	606557	Syndrome de Dias-Logan	617101	retard de développement, DI, microcéphalie, dysmorphie faciale, persistance asymptomatique de l'hémoglobine foetale (HbF)
BRAF	7	AD	V-Raf murine sarcoma viral oncogene homolog B1	164757	Syndrome Cardiofaciocutané / Syndrome LEOPARD 3 / Syndrome de Noonan 7	115150 / 613707 / 613706	DI, retard staturo-pondéral postnatal, anomalies cardiaques, anomalies squelettiques, dysmorphie faciale

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
CASK	X	XL	Calcium/calmodulin- dependent serine protein kinase (MAGUK family)	300172	Retard mental avec ou sans nystagmus / Syndrome MRCPCH (retard mental et microcéphalie avec hypoplasie pontine et cérébelleuse)	300422 / 300749	DI modérée avec ou sans nystagmus congénital / DI sévère associée à une hypoplasie/atrophie ponto-cérébelleuse et à une microcéphalie
CDK13	7	AD	Cyclin-dependent kinase 13	603309	DI, trouble cardiaque congénital et dysmorphie faciale	617360	DI, trouble cardiaque congénital et dysmorphie faciale
CDKL5	X	XL	Cyclin-dependent kinase-like 5	300203	Encéphalopathie épileptique infantile précoce de type 2	300672	Syndrome de Rett atypique avec crises épileptiques débutant précocément
CHD2	15	AD	Chromodomain helicase DNA binding protein 2	602119	Encéphalopathie épileptique débutant dans l'enfance	615369	DI-NS ou DI, épilepsie
CHD4	12	AD	Chromodomain helicase DNA binding protein 4	603277	Syndrome de Sifrim-Hitz-Weiss	617159	DI, anomalies congénitales variables (cardiaques, squelettiques, urogénitales)
CHD7	8	AD	Chromodomain helicase DNA binding protein 7	608892	Syndrome CHARGE / Hypogonadisme hypogonadotrope avec ou sans anosmie	214800 / 612370	Colobome, atrésie/sténose choanale, anomalies des nerfs crâniens et de l'oreille
CHD8	14	AD	Chromodomain helicase DNA binding protein 8	610528	-	-	TSA
CNKS2	X	XL	Connector enhancer of kinase suppressor of Ras 2	300724	Retard mental lié à l'X syndromique (type Hogue)	301008	DI-NS ou DI, épilepsie
CREBBP	16	AD	CREB binding protein	600140	Rubinstein-Taybi syndrome 1	180849	DI, petite taille, retard de croissance postnatale, microcéphalie, pouces et hallux large, dysmorphie faciale
CSNK2A1	20	AD	Casein kinase 2, alpha-1	115440	Syndrome neurodéveloppemental Okur-Chung	617062	Retard de développement, DI avec trouble sévère du langage, anomalies comportementales, dysmorphie faciale
CTCF	16	AD	CCCTC-binding factor	604167	Retard mental AD 21	615502	DI-NS
CTNNA1	3	AD	Catenin (cadherin-associated protein), beta 1	116806	Retard mental AD 19	615075	DI sévère, langage absent ou très limité, microcéphalie, hypotonie et spasticité
CUL4B	X	XL	Cullin 4B	300304	Retard mental lié à l'X syndromique 15 (type Cabezas)	300354	DI, Petite taille, hypogonadisme, ataxie, retard de langage, tremblements

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
DDX3X	X	XL	DEAD (Asp-Glu-Ala-Asp) box 3, X-linked	300160	Retard mental lié à l'X 102	300958	DI légère à sévère, parfois atteintes neurologiques(hypotonie 76 %, mouvements anormaux 45%, troubles du comportement 53 %), atteintes de la peau 37 %, hyperlaxité ligamentaire 37 %, atteintes de la vision 34 %
DLG3	X	XL	Discs, large homolog 3 (Drosophila)	300189	Retard mental lié à l'X 90	300850	DI-NS
DNM1	9	AD	Dynamin 1	602377	Encéphalopathie épileptique infantile précoce 31	616346	DI-NS ou DI, épilepsie
DNMT3A	2	AD	DNA methyltransferase 3A	602769	Syndrome de Tatton-Brown-Rahman	615879	DI, grande taille, dysmorphie
DYNC1H1	14	AD	Dynein, cytoplasmic 1, heavy chain 1	600112	Maladie de Charcot-Marie-Tooth de type axonale 20 / Retard mental AD 13 / Atrophie musculaire spinale prédominant aux extrémités inférieures 1	614228 / 614563 / 158600	DI-NS ou DI, troubles neurologiques
DYRK1A	21	AD	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	600855	Retard mental lié à l'X AD 7	614104	DI, microcéphalie, difficulté d'alimentation, démarche ataxique, stéréotypies, dysmorphie faciale
EBF3	10	AD	Early B-cell factor 3	607407	Syndrome d'hypotonie, ataxie, et retard développemental	617330	Retard de développement, hypotonie, ataxie
EEF1A2	20	AD	Eukaryotic translation elongation factor 1, alpha 2	602959	Encéphalopathie épileptique infantile précoce 33 / Retard mental AD 38	616409 / 616393	DI-NS ou DI, épilepsie
EHMT1	9	AD	Euchromatic histone-lysine N-methyltransferase 1	607001	Syndrome de Kleefstra	610253	DI sévère, hypotonie, épilepsie, dysmorphie faciale
EP300	22	AD	E1A binding protein p300	602700	Syndrome de Rubinstein-Taybi 2	613684	DI, petite taille, retard de croissance postnatale, microcéphalie, pouces et hallux large, dysmorphie faciale
FGFR2	10	AD	Fibroblast growth factor receptor 2	176943	Syndrome d'Antley-Bixler sans anomalie génitale ou trouble de la stéroïdogénèse / Syndrome d'Apert	207410 / 101200	Craniosténose, anomalies osseuses, DI inconstante
FGFR3	4	AD/AR	Fibroblast growth factor receptor 3	134934	Achondroplasie / Hypochondroplasie / Syndrome de Crouzon avec acanthosis nigricans	100800 / 146000 / 612247	petite taille, anomalies osseuses, DI inconstante

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
FLNA	X	XL	Filamin A, alpha	300017	Syndrome FG 2	300321	Retard de développement, retard de langage, anomalies associées diverses, dysmorphie
FOXP1	14	AD	Forkhead box G1	164874	Syndrome de Rett atypique	613454	Syndrome de Rett atypique : DI sévère, microcéphalie progressive, stéréotypies, hypotonie, TSA, absence de langage
FOXP1	3	AD	Forkhead box P1	605515	Retard mental avec trouble du langage et TSA	613670	DI, hypotonie, troubles du langage, TSA
GABRB3	15	AD	Gamma-aminobutyric acid (GABA) A receptor, beta 3	137192	Encéphalopathie épileptique infantile précoce 43	617113	Retard de développement, épilepsie ou TSA
GATAD2B	1	AD	GATA zinc finger domain containing 2B	614998	Retard mental AD 18	615074	DI sévère, hypotonie, langage limité, dysmorphie
GNAI1	7	inconnu	Guanine nucleotide-binding protein, alpha inhibiting activity polypeptide 1	139310	-	-	-
GNAO1	16	AD	Guanine nucleotide-binding protein, alpha activating activity polypeptide O	139311	Encéphalopathie épileptique infantile précoce 17	615473	Retard de développement, épilepsie, avec ou sans mouvements anormaux
GRIA3	X	XL	Glutamate receptor, ionotropic, AMPA 3	305915	Retard mental lié à l'X 94	300699	DI-NS sévère ou Macrocéphalie, épilepsie, TSA
GRIN1	9	AD	Glutamate receptor, ionotropic, N-methyl D-aspartate 1	138249	Retard mental AD 8	614254	DI modérée à sévère, épilepsie infantile précoce, microcéphalie et retard de langage
GRIN2A	16	AD	Glutamate receptor, ionotropic, N-methyl D-aspartate 2A	138253	Epilepsie, focale, avec trouble du langage avec ou sans retard mental	245570	Retard psychomoteur, épilepsie
GRIN2B	12	AD	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	138252	Encéphalopathie épileptique, infantile précoce, 27 / Retard mental AD 6	616139 / 613970	DI-NS
HDAC8	X	XL	Histone deacetylase 8	300269	Syndrome de Cornelia de Lange 5	300882	DI, retard de croissance, anomalies des extrémités, dysmorphie faciale, parfois malformations associées (cardiaques, rénales, etc.)
HNRNPU	1	AD	Heterogeneous nuclear ribonucleoprotein U	602869	Encéphalopathie épileptique infantile précoce 54	617391	Retard de développement, DI sévère, épilepsie

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
HUWE1	X	XL	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	300697	Retard mental, lié à l'X, syndromique, type Turner	300706	DI-NS ou DI, macrocéphalie, macroorchidisme
IL1RAPL1	X	XL	Interleukin 1 receptor accessory protein-like 1	300206	Retard mental lié à l'X 21/34	300143	DI-NS ou TSA, hyperlaxité articulaire
IQSEC2	X	XL	IQ motif and Sec7 domain 2	300522	Retard mental lié à l'X 1/78	309530	DI-NS ou DI sévère, microcéphalie progressive postnatale, retard psychomoteur sévère, TSA, épilepsie
ITPR1	3	AD	Inositol 1,4,5-trisphosphate receptor, type 1	147265	Syndrome de Gillespie / Ataxie spinocérébelleuse AD 15	206700 / 606658	DI, ataxie cérébelleuse non progressive, avec ou sans aniridie partielle bilatérale
KANSL1	17	AD	KAT8 regulatory NSL complex subunit 1	612452	Syndrome Koolen-De Vries	610443	Retard du développement, hypotonie, comportement amical caractéristique, dysmorphie faciale
KAT6A	8	AD	K (lysine) acetyltransferase 6A	601408	Retard mental AD 32	616268	DI, hypotonie, microcéphalie, dysmorphie faciale
KAT6B	10	AD	K (lysine) acetyltransferase 6B	605880	Syndrome Génitopatellaire / Syndrome SBBYSS	606170 / 603736	Retard psychomoteur sévère, anomalies squelettiques, dysmorphie faciale
KCNB1	20	AD	Potassium channel, voltage-gated, Shab-related subfamily, member 1	600397	Encéphalopathie épileptique infantile précoce 26	616056	Encéphalopathie épileptique
KCNQ2	20	AD	Potassium channel, voltage-gated, KQT-like subfamily, member 2	602235	Encéphalopathie épileptique infantile précoce 7	613720	Encéphalopathie épileptique
KDM5C	X	XL	Lysine (K)-specific demethylase 5C	314690	Retard mental lié à l'X syndromique (type Claes-Jensen)	300534	DI légère à sévère, spasticité, cryptorchidie, hypoplasie maxillaire, épilepsie, petite taille, et troubles du langage et du comportement
KIAA2022	X	XL	Neurite extension and migration factor (NEXMIF)	300524	Retard mental lié à l'X 98	30912	DI-NS
KIF1A	2	AR/AD	Kinesin family member 1A	601255	Retard mental AD 9 / Neuropathie AR héréditaire sensorielle type IIC / Paraplégie spastique AR	614255 / 614213 / 610357	DI-NS ou DI, signes neurologiques
KMT2A	11	AD	Lysine (K)-specific methyltransferase 2A	159555	Syndrome de Wiedemann-Steiner	605130	DI, petite taille, anomalies squelettiques, hypertrichose, dysmorphie faciale
KMT2D	12	AD	Lysine (K)-specific methyltransferase 2D	602113	Syndrome de Kabuki 1	147920	DI légère ou modérée, anomalies squelettiques, retard de croissance postnatal, dysmorphie faciale
MAGEL2	15	AD	MAGE-like 2	605283	Syndrome de Schaaf-Yang	615547	Syndrome de Prader-Willi like (retard psychomoteur, hypotonie néonatale, hyperphagie, obésité, troubles du comportement)

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
MECP2	X	XL	Methyl CpG binding protein 2	300005	Syndrome de Rett / Encephalopathie néonatale sévère	312750 / 300673	Retard psychomoteur sévère, régression psychomotrice, microcéphalie, hypotonie, troubles de la marche, TSA, épilepsie, stéréotypies, retard de langage, troubles respiratoires, constipation, retard de croissance, dysmorphie
MED12	X	XL	Mediator complex subunit 12	300188	Syndrome Lujan-Fryns / Syndrome Opitz-Kaveggia / Syndrome Ohdo	309520 / 305450 / 300895	DI, grande taille, troubles comportementaux, dysmorphie faciale, autres signes associés
MED13L	12	AD	Mediator complex subunit 13-like	608771	Retard mental avec signes dysmorphiques cranio-faciaux avec ou sans anomalie cardiaque / Transposition des gros vaisseaux, dextro- rotation 1	616789 / 608808	DI-NS
MEF2C	5	AD	Myocyte enhancer factor 2C	600662	Retard mental, mouvements stéréotypiques, épilepsie et/ou malformations cérébrales	613443	DI, épilepsie, stéréotypies, retard de langage
MTOR	1	AD	Mechanistic target of rapamycin	601231	Syndrome de Smith- Kingsmore	616638	DI, macrocéphalie, trouble neurologique, convulsions, hernie ombilicale, petit thorax
MYT1L	2	AD	Myelin transcription factor 1-like	613084	Retard mental AD 39	616521	DI-NS
NAA10	X	XL	N(alpha)-acetyltransferase 10, NatA catalytic subunit	300013	Microphthalmie syndromique 1 / Syndrome d'Ogden	309800 / 300855	Retard psychomoteur, vieillesse prématurée, arythmie cardiaque
NFIX	19	AD	Nuclear factor I/X (CCAAT-binding transcription factor)	164005	Syndrome de Marshall-Smith / Syndrome de Sotos de type 2	602535 / 614753	DI, avance de l'âge osseux, avance staturale, dysmorphie
NSD1	X	AD	Nuclear receptor binding-SET domain protein 1	606681	Syndrome de Sotos de type 1	117550	DI, gigantisme, avance de maturation osseuse, dysmorphie faciale
OPHN1	X	XL	Oligophrenin 1	300127	Retard mental lié à l'X avec hypoplasie cérébelleuse et dysmorphie faciale	300486	DI modérée à sévère, dysgénésie cérébelleuse, hypotonie, convulsions
PACS1	11	AD	Phosphofurin acidic cluster sorting protein 1	607492	Syndrome de Schuss- Hoeijmakers	615009	DI, anomalies génitales, dysmorphie faciale

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
PAX6	11	AD	Paired box gene 6	607108	Aniridie / Hypoplasie du nerf optique / Dysgénésie du segment antérieur	106210 / 165550 / 604229	Aniridie / Hypoplasie du nerf optique / Dysgénésie du segment antérieur
PHIP	6	AD	Pleckstrin homology domain-interacting protein	612870	Retard de développement, DI, obésité, et dysmorphie faciale	617991	Retard de développement, DI, anomalies du comportement, obésité, dysmorphie faciale
POGZ	1	AD	Pogo transposable element with ZNF domain	614787	Syndrome de White-Sutton	616364	Retard psychomoteur, dysmorphie faciale
PPP2R5D	6	AD	Protein phosphatase 2, regulatory subunit B (B56), delta	601646	Retard mental AD 35	616355	DI-NS
PQBP1	X	XL	Polyglutamine binding protein 1	300463	Syndrome de Renpenning	309500	DI, microcéphalie, petite taille, maigreur
PTCHD1	X	XL	Patched domain containing 1	300828	Autisme, susceptibilité liée à l'X 4	300830	DI-NS ou DI, TSA
PTEN	10	AD	Phosphatase and tensin homolog	601728	Syndrome de Cowden	158350	Polypes hamartomateux, trichilemmone, macrocéphalie, risque accru de cancer du sein, de la thyroïde, de l'endomètre, du rein et du colorectum, dysmorphie faciale, DI inconstante
PTPN11	12	AD	Protein tyrosine phosphatase, nonreceptor-type 11	176876	Syndrome LEOPARD 1 / Syndrome de Noonan 1	151100 / 163950	DI, retard staturo-pondéral postnatal, anomalies cardiaques, anomalies squelettiques, dysmorphie faciale
PUF60	8	AD	Poly-U-binding splicing factor, 60-KDa	604819	Syndrome Verheij	615583	Retard de développement, retard de croissance, anomalie squelettiques, dysmorphie faciale
PURA	5	AD	Purine-rich element-binding protein A	600473	Retard mental AD 31	616158	retard de développement, hypotonie, épilepsie
RAI1	17	AD	Retinoic acid induced gene 1	607642	Syndrome de Smith-Magenis	182290	Retard psychomoteur, DI, troubles du sommeil, troubles du comportement, anomalies squelettiques, dysmorphie faciale
RPS6KA3	X	XL	Ribosomal protein S6 kinase, 90-KDa, 3	300075	Syndrome de Coffin-Lowry (Retard mental lié à l'X 19)	303600	DI-NS ou Retard psychomoteur, retard de croissance, anomalies digitales, déformations squelettiques progressives, dysmorphie faciale
SATB2	2	AD	Special AT-rich sequence-binding protein 2	608148	Syndrome de Glass	612313	DI, fente palatine, dysmorphie faciale
SCN1A	2	AD	Sodium channel, voltage-gated, type I, alpha subunit	182389	Encéphalopathie épileptique infantile précoce 6 (Syndrome de Dravet)	607208	Epilepsie, détérioration cognitive et motrice possible

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
SCN1A	2	AD	Sodium channel, voltage-gated, type I, alpha subunit	182389	Encéphalopathie épileptique infantile précoce 6 (Syndrome de Dravet)	607208	Epilepsie, détérioration cognitive et motrice possible
SCN2A	2	AD	Sodium channel, voltage-gated, type II, alpha subunit	182390	Encéphalopathie épileptique, infantile précoce, 11 / Convulsions bénignes infantiles familiales 3	613721 / 607745	DI-NS ou DI, épilepsie
SCN8A	12	AD	Sodium channel, voltage-gated, type VIII alpha subunit	600702	Encéphalopathie épileptique, infantile précoce, 13 / Convulsions bénignes infantiles familiales 5	614558 / 617080	DI-NS ou DI, épilepsie, ataxie
SETBP1	18	AD	SET binding protein 1	611060	Syndrome de Schinzel-Giedion / Retard mental AD 29	269150 / 616078	DI sévère, anomalies congénitales polymalformatives, dysmorphie faciale ou DI légère à sévère, trouble du langage, dysmorphie faciale
SETD5	3	AD	SET domain containing 5	615743	Retard mental AD 23	615761	DI, troubles du comportement, dysmorphie
SHANK3	22	AD	SH3 and multiple ankyrin repeat domains 3	606230	Syndrome de Phelan-McDermid	606232	DI modérée à sévère, retard de langage, hypotonie néonatale, TSA, dysmorphie faciale
SLC16A2	X	XL	Solute carrier family 16, member 2 (thyroid hormone transporter)	300095	Syndrome de Allan-Herndon-Dudley	300523	DI, hypotonie, hypoplasie musculaire
SLC2A1	1	AD	Solute carrier family 2 (facilitated glucose transporter), member 1	138140	Syndrome de déficit en GLUT1 / Dystonie 9	606777 / 601042	Retard développemental, encéphalopathie épileptique, microcéphalie acquise, ataxie, spasticité
SLC6A1	3	AD	Solute carrier family 6 (neurotransmitter transporter, GABA), member 1	137165	Epilepsie myoclonono-atonique	616421	DI, épilepsie
SLC6A8	X	XL	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8	300036	Syndrome de déficit en créatine cérébrale 1	300352	Retard de développement global, DI, retard proéminent de langage, TSA, convulsions
SLC9A6	X	XL	Solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6	300231	Syndrome de Christianson	300243	DI, microcéphalie, épilepsie, retard de développement sévère ou régression, trouble sévère du langage, hypotonie, mouvements anormaux
SMARCA2	9	AD	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 2	600014	Syndrome de Nicolaidis-Baraitser	601358	DI, petite taille, hypotrichose, brachydactylie, épilepsie, dysmorphie

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
SMARCA4	19	AD	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 2	603254	Syndrome de Coffin-Siris 4	614609	retard psychomoteur, DI, microcéphalie, traits grossiers du visage, petite taille, retard de croissance, ongles hypoplasiques au V ^{ème} doigt et au V ^{ème} orteil
SMC1A	X	XL	Structural maintenance of chromosomes 1A	300040	Syndrome de Cornelia de Lange 2	300590	DI, retard de croissance, anomalies des extrémités, dysmorphie
STXBP1	9	AD	Syntaxin binding protein 1	602926	Encéphalopathie épileptique infantile précoce 4	612164	DI-NS ou DI, épilepsie
SYNGAP1	6	AD	Synaptic Ras GTPase-activating protein 1	603384	Retard mental AD 5	612621	DI-NS modérée à sévère ou DI, épilepsie, TSA
TBL1XR1	3	AD	Transducin-beta-like 1 X-linked receptor 1	608628	Syndrome de Pierpont / Retard mental AD 41	602342 / 616944	DI, anomalies des extrémités, dysmorphie faciale
TBR1	2	AD	T-box, brain, 1	604616	DI avec TSA et retard de langage	606053	DI-NS ou DI, TSA, retard de croissance
TCF20	22	inconnu	Transcription factor 20	603107	-	-	-
TCF4	18	AD	Transcription factor 4	602272	Syndrome de Pitt-Hopkins	610954	DI-NS ou Retard psychomoteur sévère, DI, hyperventilation, anomalies intestinales, stéréotypies, dysmorphie
TRIM27	6	inconnu	Tripartite motif-containing protein 27	602165	-	-	-
TRIO	5	AD	Trio Rho guanine nucleotide exchange factor	601893	Retard mental AD 44	617061	Retard de développement, DI légère, anomalies des doigts
TRIP12	2	AD	Thyroid hormone receptor interactor 12	609761	Retard mental AD 49	617752	DI-NS ou DI, TSA
UBE3A	15	inconnu	Ubiquitin protein ligase E3A	601623	Syndrome d'Angelman	105830	DI sévère, trouble sévère du langage, microcéphalie, troubles neurologiques avec démarche « en pantin », ataxie et crises épileptiques, trouble du comportement, dysmorphie faciale
UPF3B	X	XL	UPF3 regulator of nonsense transcripts homolog B (yeast)	300298	Retard mental lié à l'X syndromique 14	300676	DI-NS ou DI, TSA
USP7	16	inconnu	Ubiquitin-specific protease 7	602519	-	-	-
USP9X	X	XL	Ubiquitin-specific protease 9, X-linked	300072	Retard mental lié à l'X 99	300919	DI-NS
WAC	10	AD	WW domain-containing adaptor with coiled-coil	615049	Syndrome de Desanto-Shinawi	616708	Retard de développement, troubles du comportement, dysmorphie faciale
WDFY3	4	AD	WD repeat- and FYVE domain-containing 3	617485	Microcéphalie primaire AD 18	617520	DI légère à modérée, microcéphalie

Gène	Chr.	Mode de transm.	Nom	OMIM	Syndrome(s) associé(s)	MIM	Phénotype (s)
WDR45	X	XL	WD repeat domain 45	300526	Neurodégénérescence avec surcharge cérébrale en fer type 5	300894	Dystonie progressive, neurodégénérescence chez l'adulte

Chr. = Chromosome

Mode de transm. = Mode de transmission

AD = autosomique dominant

AR = autosomique récessif

XL = lié à l'X

DI-NS = DI non syndromique

Gène inclus dans le DI44


Annexe 5 : Estimation du rendement diagnostique des tests DI44 et DI117

Estimation du rendement diagnostique des tests DI44 et DI117

Gène	Chr.	Mode de transm.	Rendement diagnostique		
			Global	Pubmed	France
ANKRD11	16	AD	0,32%	0,30%	0,56%
ARID1B	6	AD	0,33%	0,34%	0,52%
ATRX	X	XL	0,13%	0,07%	0,14%
CASK	X	XL	0,09%	0,10%	0,14%
CDKL5	X	XL	0,06%	0,05%	0,07%
CTNNB1	3	AD	0,12%	0,11%	0,25%
CUL4B	X	XL	0,11%	0,06%	0,14%
DLG3	X	XL	0,05%	0,06%	0,00%
DYRK1A	21	AD	0,21%	0,19%	0,45%
EP300	22	AD	0,12%	0,13%	0,18%
FOXP1	14	AD	0,08%	0,09%	0,07%
FOXP1	3	AD	0,12%	0,12%	0,24%
GATAD2B	1	AD	0,09%	0,08%	0,25%
GRIA3	X	XL	0,04%	0,03%	0,03%
GRIN1	9	AD	0,05%	0,01%	0,24%
GRIN2A	16	AD	0,05%	0,05%	0,03%
GRIN2B	12	AD	0,16%	0,11%	0,48%
IL1RAPL1	X	XL	0,04%	0,01%	0,10%
IQSEC2	X	XL	0,13%	0,09%	0,35%
KDM5C	X	XL	0,10%	0,03%	0,38%
KMT2A	11	AD	0,28%	0,24%	0,66%
MECP2	X	XL	0,27%	0,23%	0,52%
MED13L	12	AD	0,23%	0,19%	0,55%
NAA10	X	XL	0,06%	0,06%	0,11%
PQBP1	X	XL	0,04%	0,02%	0,07%
PTCHD1	X	XL	0,02%	0,01%	0,03%
RAI1	17	AD	0,03%	0,03%	0,07%
SATB2	2	AD	0,16%	0,16%	0,25%
SCN2A	2	AD	0,25%	0,22%	0,38%
SCN8A	12	AD	0,12%	0,10%	0,07%
SETBP1	18	AD	0,05%	0,03%	0,14%
SHANK3	22	AD	0,13%	0,05%	0,55%
SLC16A2	X	XL	0,05%	0,04%	0,07%
SLC2A1	1	AD	0,08%	0,05%	0,17%
SLC6A8	X	XL	0,04%	0,04%	0,03%
SLC9A6	X	XL	0,05%	0,01%	0,17%
SMARCA2	9	AD	0,10%	0,12%	0,08%
SMC1A	X	XL	0,08%	0,09%	0,03%
STXBP1	9	AD	0,12%	0,11%	0,21%
SYNGAP1	6	AD	0,16%	0,16%	0,31%
TBR1	2	AD	0,06%	0,04%	0,18%
TCF4	18	AD	0,18%	0,14%	0,55%
UPF3B	X	XL	0,04%	0,03%	0,00%
WDR45	X	XL	0,10%	0,10%	0,22%
Au total pour le test DI44			5,09%	4,30%	10,06%

Gène	Chr.	Mode de transm.	Rendement diagnostique		
			Global	Pubmed	France
ADNP	20	AD	0,19%	0,17%	0,43%
AHDC1	1	AD	0,10%	0,08%	0,37%
ASH1L	1	AD	0,06%	0,05%	0,07%
ASXL3	18	AD	0,10%	0,11%	0,15%
AUTS2	7	AD	0,07%	0,05%	0,22%
BCL11A	2	AD	0,05%	0,05%	0,15%
BRAF	7	AD	0,06%	0,06%	0,07%
CDK13	7	AD	0,08%	0,09%	0,07%
CHD2	15	AD	0,12%	0,12%	0,24%
CHD4	12	AD	0,06%	0,06%	0,15%
CHD7	8	AD	0,07%	0,05%	0,09%
CHD8	14	AD	0,09%	0,06%	0,11%
CNKSRR2	X	XL	0,05%	0,04%	0,11%
CREBBP	16	AD	0,11%	0,11%	0,19%
CSNK2A1	20	AD	0,06%	0,05%	0,15%
CTCF	16	AD	0,05%	0,05%	0,00%
DDX3X	X	XL	0,24%	0,24%	0,58%
DNM1	9	AD	0,05%	0,04%	0,04%
DNMT3A	2	AD	0,05%	0,03%	0,25%
DYNC1H1	14	AD	0,12%	0,12%	0,20%
EBF3	10	AD	0,06%	0,06%	0,07%
EEF1A2	20	AD	0,06%	0,06%	0,07%
EHMT1	9	AD	0,10%	0,09%	0,19%
FGFR2	10	AD	0,05%	0,06%	0,00%
FGFR3	4	AD/AR	0,05%	0,06%	0,00%
FLNA	X	XL	0,07%	0,08%	0,00%
GABRB3	15	AD	0,05%	0,05%	0,04%
GNAI1	7	inconnu	0,05%	0,05%	0,00%
GNAO1	16	AD	0,06%	0,06%	0,04%
HDAC8	X	XL	0,08%	0,09%	0,08%
HNRNPU	1	AD	0,07%	0,08%	0,12%
HUWE1	X	XL	0,05%	0,06%	0,00%
ITPR1	3	AD	0,08%	0,10%	0,00%
KANSL1	17	AD	0,07%	0,06%	0,20%
KAT6A	8	AD	0,11%	0,09%	0,24%
KAT6B	10	AD	0,12%	0,10%	0,28%
KCNB1	20	AD	0,05%	0,01%	0,26%
KCNQ2	20	AD	0,16%	0,14%	0,14%
KIAA2022	X	XL	0,08%	0,05%	0,30%
KIF1A	2	AR/AD	0,07%	0,09%	0,04%
KMT2D	12	AD	0,08%	0,02%	0,30%
MAGEL2	15	AD	0,02%	0,02%	0,00%
MED12	X	XL	0,07%	0,05%	0,08%
MEF2C	5	AD	0,06%	0,07%	0,04%
MTOR	1	AD	0,05%	0,05%	0,15%
MYT1L	2	AD	0,05%	0,04%	0,16%

Gène	Chr.	Mode de transm.	Rendement diagnostique		
			Global	Pubmed	France
NFIX	19	AD	0,06%	0,05%	0,12%
NSD1	X	AD	0,10%	0,09%	0,15%
OPHN1	X	XL	0,06%	0,05%	0,04%
PACS1	11	AD	0,08%	0,07%	0,15%
PAX6	11	AD	0,06%	0,05%	0,07%
PHIP	6	AD	0,05%	0,05%	0,00%
POGZ	1	AD	0,14%	0,09%	0,47%
PPP2R5D	6	AD	0,08%	0,10%	0,04%
PTEN	10	AD	0,08%	0,09%	0,00%
PTPN11	12	AD	0,07%	0,05%	0,17%
PUF60	8	AD	0,05%	0,05%	0,15%
PURA	5	AD	0,12%	0,10%	0,22%
RPS6KA3	X	XL	0,05%	0,03%	0,11%
SCN1A	2	AD	0,10%	0,05%	0,11%
SETD5	3	AD	0,15%	0,12%	0,24%
SLC6A1	3	AD	0,07%	0,07%	0,20%
SMARCA4	19	AD	0,05%	0,05%	0,04%
TBL1XR1	3	AD	0,09%	0,08%	0,24%
TCF20	22	inconnu	0,10%	0,08%	0,30%
TRIM27	6	inconnu	<0,01%	<0,01%	0,00%
TRIO	5	AD	0,05%	0,05%	0,16%
TRIP12	2	AD	0,06%	0,05%	0,09%
UBE3A	15	inconnu	0,05%	0,01%	0,23%
USP7	16	inconnu	0,02%	0,02%	0,00%
USP9X	X	XL	0,08%	0,07%	0,17%
WAC	10	AD	0,05%	0,05%	0,00%
WDFY3	4	AD	0,05%	0,05%	0,00%
Au total pour le test DI117			10,58%	9,33%	19,96%

 Gène inclus dans le DI44

Rendement diagnostique = $\frac{\text{nombre de patients avec un ou plusieurs variant(s) pathogène(s) décelé(s)}}{\text{nombre de patients étudiés}}$

Pubmed = au sein des publications scientifiques sur Pubmed

France = au sein des patients ayant consulté dans les CHU de Strasbourg, Lille, Rennes, La Pitié-Salpêtrière, Lyon et Dijon

Global = regroupement des données Pubmed et des données des CHU français

Chr. = Chromosome

Mode de transm. = Mode de transmission

Annexe 6 : Tableau précisant le nombre de "Sanger" nécessaires pour compléter la couverture théorique sur CDS ± 6 pb des tests DI44 et DI117

Gène	Couverture théorique des tests sur CDS ± 6 pb						nbre de "Sanger" complémentaire(s)	
	Test DI44			Test DI117 (sur les 44 gènes)			Test DI44	Test DI117
	Régions non couvertes en théorie	n. pb n. c.	n. pb n. c. total/gène	Régions non couvertes en théorie	n. pb n. c.	n. pb n. c. total/gène		
ANKRD11 NM_001256182.1	exon 10 (c.7067_7075)	9	9	exon 10 (c.1346_1379) exon 10 (c.4379_4421)	34 43	77	1 (9 pb dans l'exon 10)	2 (34 pb et 43 pb dans l'exon 10)
ARID1B NM_020732.3			0			0	0	0
ATRX NM_000489.4			0	exon 9 (c.2465_2482) exon 15 (c.4322_4386)	18 65	83	0	2 (18 pb dans l'exon 9 et 65 pb dans l'exon 15)
CASK NM_003688.3			0			0	0	0
CDKL5 NM_003159.2			0			0	0	0
CTNNB1 NM_001904.3			0			0	0	0
CUL4B NM_003588.3			0			0	0	0
DLG3 NM_021120.3			0			0	0	0
DYRK1A NM_001396.3			0	exon 12 (c.1836_1847)	12	12	0	1 (12 pb dans l'exon 12)
EP300 NM_001429.3			0			0	0	0
FOXG1 NM_005249.4	exon 1 (c.279_311)	33	33			0	1 (33 pb dans l'exon 1)	0
FOXP1 NM_032682.5			0			0	0	0
GATAD2B NM_020699.3			0			0	0	0
GRIA3 NM_000828.4			0			0	0	0
GRIN1 NM_007327.3			0			0	0	0
GRIN2A NM_000833.4			0			0	0	0
GRIN2B NM_000834.3			0			0	0	0
IL1RAPL1 NM_014271.3			0			0	0	0
IQSEC2 NM_001111125.2			0	exon 5 (c.1950_1981) exon 15 (c.3716_3731)	32 16	48	0	2 (32 pb dans l'exon 5 et 16 pb dans l'exon 15)
KDM5C NM_004187.3			0			0	0	0
KMT2A NM_001197104.1	exon 1 (c.6_195)	190	190			0	1 (190 pb dans l'exon 1)	0
MECP2 NM_004992.3							0	0
MED13L NM_015335.4			0			0	0	0

n. pb. n. c. = nombre de bases non couvertes

nbre de "Sanger" complémentaire(s) = nombre d'amplicon(s) séquencé(s) par la technique de Sanger pour compléter la couverture à 100%

Gène	Couverture théorique des tests sur CDS ± 6 pb						nbre de "Sanger" complémentaire(s)	
	Test DI44			Test DI117 (sur les 44 gènes)			Test DI44	Test DI117
	Régions non couvertes en théorie	n. pb n. c.	n. pb n. c. total/gène	Régions non couvertes en théorie	n. pb n. c.	n. pb n. c. total/gène		
NAA10	NM_003491.3		0			0	0	0
PQBP1	NM_005710.2		0			0	0	0
PTCHD1	NM_173495.2						0	0
RAI1	NM_030665.3		0	exon 3 (c.4896_4917) exon 3 (c.843_856)	22 14	36	0	2 (22 pb et 14 pb dans l'exon 3)
SATB2	NM_001172509.1		0			0	0	0
SCN2A	NM_021007.2		0	exon 13 (c.1523_1547)	25	25	0	1 (25 pb dans l'exon 13)
SCN8A	NM_014191.3		0			0	0	0
SETBP1	NM_015559.2		0			0	0	0
SHANK3	NM_033517.1	5'UTR (c.-25_-1) exon 1 (c.1_63) intron 1 (c.63+1_63+25)	6 63 6			75	0	1 (75 pb dans la région 5'UTR-exon1-intron1)
SLC16A2	NM_006517.4						0	0
SLC2A1	NM_006516.2		0			0	0	0
SLC6A8	NM_005629.3	exon 1 (c.23_32)	10			10	0	1 (10 pb dans l'exon 1)
SLC9A6	NM_001042537.1		0			0	0	0
SMARCA2	NM_003070.4		0	exon 4 (c.647_741) exon 32 (c.4541_4561)	95 21	116	0	2 (95 pb dans l'exon 4 et 21 pb dans l'exon 32)
SMC1A	NM_006306.3		0			0	0	0
STXBP1	NM_003165.3		0			0	0	0
SYNGAP1	NM_006772.2		0	exon 15 (c.2869_2886)	18	18	0	1 (18 pb dans l'exon 15)
TBR1	NM_006593.3		0			0	0	0
TCF4	NM_001083962.1		0			0	0	0
UPF3B	NM_080632.2		0	exon 7 (c.652_729)	78	78	0	1 (78 pb dans l'exon 7)
WDR45	NM_007075.3		0			0	0	0
Au total pour les 44 gènes			85			273	5	14

n. pb. n. c. = nombre de bases non couvertes

nbre de "Sanger" complémentaire(s) = nombre d'amplicon(s) séquencé(s) par la technique de Sanger pour compléter la couverture à 100%

Annexe 7 : Tableau détaillant la couverture déduite de la pratique expérimentale des tests DI44 et DI117 sur les régions CDS ± 6 pb et CDS ± 25 pb

Gène		Régions non couvertes du Test DI44									Régions non couvertes du Test DI117 (sur les 44 gènes)												
		Min			Max			Min			Max												
		Régions non couvertes	nbre de patients concernés /6	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène	Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène	Régions non couvertes	nbre de patients concernés /15	CDS ± 25 pb n. pb n. c. total/gène	CDS +/- 6 n. pb n. c. total/gène	Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS +/- 6 n. pb n. c. total/gène								
ANKRD11 NM_001256182.1	Intron4 exon 10 exon10 exon10	Intron4 (c.88-25_88-24) exon 10 (c.6822) exon10 (c.7067_7075)	6 (récurrent) 6 gap de design	2 1 9		1 9		Intron4 (c.88-25_88-24) exon 10 (c.6817_7066) exon10 (c.7067_7075)	2 250 9		250 9		exon10 (c.1346_1379) exon10 (c.4379_4421)	gap de design gap de design	34 43		34 43		exon10 (c.1346_1379) exon10 (c.4379_4421)	34 43		34 43	
ARID1B NM_020732.3	5'UTR Exon1 Exon1 Exon1 Exon1 Exon1 Exon1 Intron1	5'UTR (c.-25_-1) Exon1 (c.1285_1542) Exon1 (c.1_191) Exon1 (c.401) Exon1 (c.473_690) Exon1 (c.902_1136) intron1 (c.1542+1_1542+25)	6 (récurrent) 6 (suite) (suite) 6 6 (récurrent)	25 258 191 1 218 235 25		6 258 191 1 218 235 6		5'UTR (c.-25_-1) Exon1 (c.1285_1542) Exon1 (c.1_690) Exon1 (c.730_1136) intron1 (c.1542+1_1542+25)	25 258 690 407 25		6 258 690 407 6		pas de gap	1 (run1) (suite)	0				Exon1 (c.580_581) Exon1 (c.592_597)	2 6		2 6	
ATRX NM_000489.4	Intron 8 Exon 9 Exon 9 exon9 Intron11 Intron12 exon15 Intron17 Intron20 Intron24	pas de gap pas de gap pas de gap pas de gap Intron11 (c.3944-25_3944-24) Intron17 (c.4809+14_4809+25) Intron20 (c.5272+24_5272+25)	4 4 4 5 5 5	0 0 0 0 0 0		0 0 0 0 0		Intron 8 (c.663-25_663-1) exon 9 (c.1606_1879) exon 9 (c.663_774) Intron11 (c.3944-25_3944-24) Intron17 (c.4809+14_4809+25) Intron20 (c.5272+24_5272+25)	25 274 112 2 12 2		6 274 112 2 12 2		exon9 (c.2465_2482) Intron12 (c.4121-25_4121-23) exon15 (c.4322_4386) Intron24 (c.5787-25_5787-21)	gap de design gap de design gap de design gap de design	18 3 65 5		18 65 83		exon9 (c.2465_2482) Intron12 (c.4121-25_4121-23) exon15 (c.4322_4386) Intron24 (c.5787-25_5787-21)	18 3 65 5		18 65 83	
CASK NM_003688.3	Intron9 Exon10 Intron10 Intron25 Exon26 Intron26	pas de gap pas de gap pas de gap Intron25 (c.2506-25_2506-1) Exon26 (c.2506_2589) Intron26 (c.2589+1_2589+25)	3 4 3 6 (récurrent) 6 (récurrent) 6 (récurrent)	0 0 0 25 84 25		6 84 6		Intron9 (c.916-25_916-1) Exon10 (c.916_1015) Intron10 (c.1015-1_1015+25) Intron25 (c.2506-25_2506-1) Exon26 (c.2506_2589) Intron26 (c.2589+1_2589+25)	25 100 25 25 84 25		6 100 6 6 84 6				0		0			0		0	
CDKL5 NM_003159.2				0		0			0		0				0		0			0		0	
CTNNB1 NM_001904.3	Intron10	pas de gap	2	0		0		Intron10 (c.1684-25_1684-24)	2		0				0		0			0		0	
CUL4B NM_003588.3	Intron14 Exon15 Intron15	pas de gap pas de gap pas de gap	1 1 1	0 0 0		0		Intron14 (c.1796-25_1796-1) Exon15 (c.1796_1906) Intron15 (c.1906+1_1906+25)	25 111 25		6 111 6				0		0			0		0	
DLG3 NM_021120.3				0		0			0		0				0		0			0		0	
DYRK1A NM_001396.3	Intron9 exon12	pas de gap	1	0		0		Intron9 (c.1546+19_1546+25)	7		0		exon12 (c.1836_1847)	gap de design	12		12		exon12 (c.1836_1847)	12		12	
EP300 NM_001429.3	Intron3 Exon4 Intron6 Intron11 Intron27 Exon28 Intron28 Intron30 Exon31	Intron3(c.907-25_907-1) Exon4 (c.907_908) pas de gap pas de gap pas de gap pas de gap pas de gap Intron30 (c.5062-25_5062-1) Exon31 (c.5062_5294)	6 6 3 3 3 3 2 6 (récurrent) 6 (récurrent)	25 2 0 0 0 0 0 25 233		6 2 0 0 0 0 0 6 233		Intron3(c.907-25_907-1) Exon4 (c.907_909) Intron6 (c.1528+15_1528+25) Intron11 (c.2132-25_2132-23) Intron27 (c.4453-25_4453-1) Exon28 (c.4453_4617) Intron28 (c.4617+1_4617+25) Intron30 (c.5062-25_5062-1) Exon31 (c.5062_5294)	25 3 11 3 25 165 25 25 233		6 3 3 6 165 6 6 233				0		0			0		0	

		Régions non couvertes du Test DI44										Régions non couvertes du Test DI117 (sur les 44 gènes)									
Gène		Min					Max					Min					Max				
		Régions non couvertes	nbre de patients concernés /6	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène		Régions non couvertes	n. pb n. c. total/gène	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène		Régions non couvertes	nbre de patients concernés /15	CDS ± 25 pb n. pb n. c. total/gène	CDS +/-6 n. pb n. c. total/gène		Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS +/-6 n. pb n. c. total/gène		
KDM5C	NM_004187.3	Intron18	Intron18 (c.2623-25_2623-19)	6 (récurrent)	7						Intron18 (c.2623-25_2623-19)	7									
		Intron23	Intron23 (c.4039-25_4039-18)	6 (récurrent)	8						Intron23 (c.4039-25_4039-18)	8									
		Exon25	Exon25 (c.4199_4317)	6 (récurrent)	119	119					Exon25 (c.4199_4317)	119	119								
		Intron25	Intron25 (c.4317+1_4317+25)	6 (récurrent)	25	6					Intron25 (c.4317+1_4317+25)	25	6								
		Intron25	Intron25 (c.4318-25_4318-1)	6 (récurrent)	25	6					Intron25 (c.4318-25_4318-1)	25	6								
		Exon26	Exon26 (c.4318_4392)	6 (récurrent)	75	75	206				Exon26 (c.4318_4392)	75	75	206							
KMT2A	NM_001197104.3	5'UTR	5'UTR (c.-25_-1)	6 (récurrent)	25	6				5'UTR (c.-25_-1)	25	6									
		Exon1	Exon1 (c.1_6)	6 (récurrent)	6	6				Exon1 (c.1_6)	6	6									
		Exon1	Exon1 (c.196_268)	6 (récurrent)	73	73				Exon1 (c.196_268)	73	73									
		Exon1	exon1 (c.6_195)	gap de design	190	190				exon1 (c.6_195)	190	190									
		Intron7	Intron7 (c.4012+20_4012+25)	gap de design	6					Intron7 (c.4012+20_4012+25)	6										
		Intron14	pas de gap	3 (identique)	0					Intron14 (c.4820-17_4820-1)	17	6									
		Exon15	pas de gap	3 (identique)	0					Exon15 (c.4820_5004)	185	185									
		Intron15	pas de gap	3 (identique)	0					Intron15 (c.5004+1_5004+21)	21	6									
		Intron23	pas de gap	4	0					Intron23 (c.6080-25_6080-23)	3										
		Intron25	pas de gap	3	0					Intron25 (c.6319+12_6319+25)	14										
		Exon27	Exon27 (c.10482_10754)	6	273	273				Exon27 (c.10480_10754)	275	275									
		Intron27	Intron27 (c.10754+1_10754+25)	6	25	6				Intron27 (c.10754+1_10754+25)	25	6									
	Intron31	Intron31 (c.11146+23_11146+25)	gap de design	3					Intron31 (c.11146+23_11146+25)	3											
MECP2	NM_0049992.3	Intron2	pas de gap	1	0					Intron2 (c.27-25_27-20)	6										
		Intron3	pas de gap	3	0	0				Intron3 (c.378-25_378-23)	3	9	0								
MED13L	NM_015335.4	5'UTR	5'UTR (c.-25_-1)	6 (récurrent)	25	6				5'UTR (c.-25_-1)	25	6									
		Exon1	Exon1 (c.1_72)	6 (récurrent)	72	72				Exon1 (c.1_72)	72	72									
		Intron1	Intron1 (c.72+1_72+25)	6 (récurrent)	25	6				Intron1 (c.72+1_72+25)	25	6									
		Intron10			122	84					122	84									
NAA10	NM_003491.3	5'UTR	5'UTR (c.-25_-1)	6 (récurrent)	25	6				5'UTR (c.-25_-1)	25	6									
		Intron1	Exon1 (c.1_21)	6 (récurrent)	21	21				Exon1 (c.1_21)	21	21									
PQBP1	NM_005710.2	Intron1	Intron1 (c.21+1_21+25)	6 (récurrent)	25	6				Intron1 (c.21+1_21+25)	25	6									
					71	33					71	33									
PTCHD1	NM_173495.2	5'UTR	5'UTR (c.-25_-1)	6 (récurrent)	25	6				5'UTR (c.-25_-1)	25	6									
		Exon1	Exon1 (c.1_206)	6 (récurrent)	206	206	231	212		Exon1 (c.1_206)	206	206	231	212							
RAI1	NM_030665.3	Exon3																			
		Exon5	Exon5 (c.5692_5709)	6 (récurrent)	18	18				Exon5 (c.5692_5709)	18	18									
		Intron5	Intron5 (c.5709+1_5709+25)	6 (récurrent)	25	6				Intron5 (c.5709+1_5709+25)	25	6									
SATB2	NM_001172509.1	UTR 5'	UTR 5' (c.-25_-11)	gap de design	15					UTR 5' (c.-25_-11)	15										
		Exon5	Exon5 (c.-11_169)	6 (récurrent)	181	181				Exon5 (c.-11_169)	181	181									
		Intron5	Intron5 (c.169+1_169+25)	6 (récurrent)	25	6				Intron5 (c.169+1_169+25)	25	6									
SCN2A	NM_021007.2	Exon13																			
		Intron14	Intron14 (c.2016+22_2016+25)	6 (récurrent)	4					Intron14 (c.2016+22_2016+25)	4										
		Intron15	Intron15 (c.2150-25_2150-20)	6 (récurrent)	6					Intron15 (c.2150-25_2150-20)	6										
		Intron17	Intron17 (c.2563-25_2563-18)	6 (récurrent)	8					Intron17 (c.2563-25_2563-18)	8										
		Intron18	Intron18 (c.2920-25_2920-22)	gap de design	4					Intron18 (c.2920-25_2920-21)	4										
		3'UTR	pas de gap	5 (identique)	0					3'UTR (c.*21_*25)	5										
SCN8A	NM_014191.3	Exon12	Exon12 (c.1760_1998)	6 (récurrent)	239	239				Exon12 (c.1760_1998)	239	239									
		Intron12	Intron12 (c.1998+1_1998+25)	6 (récurrent)	25	6				Intron12 (c.1998+1_1998+25)	25	6									
		Intron19	pas de gap	2	0					Intron19 (c.3646-25)	1										
		Intron20	Intron20 (c.3820-25_3820-20)	gap de design	6					Intron20 (c.3820-25_3820-20)	6										
					270	245					271	245									

n. pb. n. c. = nombre de bases non couvertes

		Régions non couvertes du Test DI44								Régions non couvertes du Test DI117 (sur les 44 gènes)							
Gène		Min				Max				Min				Max			
		Régions non couvertes	nbre de patients concernés /5	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène	Régions non couvertes	n. pb n. c. total/gène	n. pb n. c. total/gène	n. pb n. c. total/gène	Régions non couvertes	nbre de patients concernés /15	CDS ± 25 pb n. pb n. c. total/gène	CDS +/- 6 n. pb n. c. total/gène	Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS +/- 6 n. pb n. c. total/gène	
SLC6A8	S'UTR	S'UTR (c.-25_-1)	6 (récurrent)	25	6	25	6	25	S'UTR (c.-25_-1)	25	6	25	6				
	Exon1	Exon1 (c.1_23)	6 (récurrent)	23	23	23	23	23	Exon1 (c.1_23)	23	23	23	23				
	Exon1	Exon1 (c.53_262)	6 (récurrent)	210	210	210	210	210	Exon1 (c.53_262)	210	210	210	210				
	Exon1	Exon1 (c.33_255)	6 (récurrent)	223	223	223	223	223	Exon1 (c.33_255)	223	223	223	223				
	exon1	exon1 (c.23_32)	gap de design	10	10	10	10	10	exon1 (c.23_32)	10	10	10	10				
	intron1													intron1 (c.262+10_262+25)	16		
	intron1	intron1 (c.262+1_262+25)	6 (récurrent)	25	6	25	6	25	intron1 (c.262+1_262+25)	25	6	25	6				
	intron1	pas de gap	5	0					intron1 (c.263-25_263-1)	25	6	25	6				
	Exon3	Exon3 (c.314)	6	1	1	1	1	1	Exon3 (c.263-25_263-1)	132	132	132	132				
	Exon3	Exon3 (c.355)	(suite)	1	1	1	1	1	Exon3 (c.263_394)								
	Intron3	Intron3 (c.394+9)	6	1	1	1	1	1	intron3 (c.394+1_394+25)	25	5	25	5				
	Exon11	Exon11 (c.1421_1495)	6 (récurrent)	75	75	75	75	75	Exon11 (c.1421_1495)	75	75	75	75				
	Intron11	intron11 (c.1495+1_1495+25)	6 (récurrent)	25	6	25	6	25	intron11 (c.1495+1_1495+25)	25	6	25	6				
	Intron11	intron11 (c.1496-25_1496-1)	6 (récurrent)	25	6	25	6	25	intron11 (c.1496-25_1496-1)	25	6	25	6				
	Exon12	Exon12 (c.1496_1596)	6 (récurrent)	101	101	101	101	101	Exon12 (c.1496_1596)	101	101	101	101				
Intron12	intron12 (c.1596+1_1596+25)	6 (récurrent)	25	6	25	6	25	intron12 (c.1596+1_1596+25)	25	6	25	6					
Intron12	intron12 (c.1597-25_1597-1)	6 (récurrent)	25	6	25	6	25	intron12 (c.1597-25_1597-1)	25	6	25	6					
Exon13	Exon13 (c.1597_1767)	6 (récurrent)	171	171	171	171	171	Exon13 (c.1597_1767)	171	171	171	171					
Intron13	intron13 (c.1767+1_1767+25)	6 (récurrent)	25	6	25	6	25	intron13 (c.1767+1_1767+25)	25	6	25	6					
Intron13													intron13 (c.1767+19_1767+25)	7	0		
SLC9A6	UTR 5'	pas de gap	4	0	991	857	1170	999	S'UTR (c.-25_-1)	25	6	25	6				
	UTR 5'																
	Exon1	pas de gap	5	0					Exon1 (c.1_95)	95	95	95	95				
	Intron3	intron3 (c.604-25_604-1)	6 (récurrent)	25	6	25	6	25	intron3 (c.604-25_604-1)	25	6	25	6				
	Exon4	Exon4 (c.604_680)	6 (récurrent)	77	77	77	77	77	Exon4 (c.604_680)	77	77	77	77				
	Intron4	intron4 (c.680+1_680+25)	6 (récurrent)	25	6	25	6	25	intron4 (c.680+1_680+25)	25	6	25	6				
Intron14	intron14 (c.1728-25_1728-16)	gap de design	10	137	89	257	190	intron14 (c.1728-25_1728-16)	10	10	10	10					
SMARCA2	exon4	intron4 (c.791-25_791-17)	gap de design	9					intron4 (c.791-25_791-17)	9				exon4 (c.647_741)	95	95	
	Intron4	intron4 (c.790+19_790+25)	6 (récurrent)	7	6	7	6	7	intron4 (c.790+19_790+25)	7	6	7	6				
	Intron4	intron4 (c.791-16_791-1)	6 (récurrent)	16	202	109	202	109	intron4 (c.791-16_791-1)	16	202	109	202				
	Exon5	Exon5 (c.791_992)	6 (récurrent)	202	109	202	109	202	Exon5 (c.791_992)	202	109	202	109				
	Exon8	Exon8 (c.1413_1521)	6 (récurrent)	109	6	109	6	109	Exon8 (c.1413_1521)	109	6	109	6				
	Intron8	intron8 (c.1521+1_1521+25)	6 (récurrent)	25	6	25	6	25	intron8 (c.1521+1_1521+25)	25	6	25	6				
	Intron13	intron13 (c.2036+25)	6 (récurrent)	1					Intron13 (c.2036+25)	1		1					
	Intron27	pas de gap	5 (identique)	0					intron27 (c.3982-25_3982-21)	5		5					
	exon32																
	Intron32	intron32 (c.4594+13_4594+25)	gap de design	13	382	323	387	323	intron32 (c.4594+13_4594+25)	13	387	323	387				
SMC1A	Intron5	intron5 (c.855-24_855-23)	6 (récurrent)	2					intron5 (c.855-24_855-23)	2							
	Intron5	intron5 (c.855-25_855-24)	gap de design	2					intron5 (c.855-25_855-24)	2							
	Intron7	pas de gap	3	0					intron7 (c.1254+22_1254+25)	3							
	Intron15	pas de gap	5 (identique)	0					intron15 (c.2421-25_2421-1)	25	6	25	6				
	Exon16	pas de gap	5 (identique)	0					Exon16 (c.2421_2562)	142	142	142	142				
	Intron16	pas de gap	5 (identique)	0					intron16 (c.2562+1_2562+25)	25	6	25	6				
	Intron18	intron18 (c.2863-25_2863-1)	6 (récurrent)	25	6	25	6	25	intron18 (c.2863-25_2863-1)	25	6	25	6				
	Exon19	Exon19 (c.2863_2973)	6 (récurrent)	111	111	111	111	111	Exon19 (c.2863_2973)	111	111	111	111				
	Intron19	intron19 (c.2973+1_2973+25)	6 (récurrent)	25	6	25	6	25	intron19 (c.2973+1_2973+25)	25	6	25	6				
	Intron19																
STXBP1	Intron3	intron3 (c.169+23_169+25)	gap de design	2	2	0	2	2	intron3 (c.169+23_169+25)	2	2	0	2				

n. pb. n. c. = nombre de bases non couvertes

		Régions non couvertes du Test DI44										Régions non couvertes du Test DI117 (sur les 44 gènes)									
Gène		Min					Max					Min					Max				
		Régions non couvertes	nbre de patients concernés /6	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène	Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS ± 6 pb n. pb n. c. total/gène	Régions non couvertes	nbre de patients concernés /15	CDS ± 25 pb n. pb n. c. total/gène	CDS +/-6 n. pb n. c. total/gène	Régions non couvertes	CDS ± 25 pb n. pb n. c. total/gène	CDS +/-6 n. pb n. c. total/gène						
SYNGAP1	NM_006772.2	S'UTR	5'UTR (c.-35_-1)	6 (récurrent)	25	6	5'UTR (c.-25_-1)	25	6	5'UTR (c.-25_-14)	15	12	5'UTR (c.-25_-1)	25	6						
	Exon1	Exon1 (c.1_67)	6 (récurrent)	67	67	Exon1 (c.1_67)	67	67	pas de gap	11	0	Exon1 (c.1_67)	67	67							
	Intron1	Intron1 (c.67+1_67+25)	6 (récurrent)	25	6	Intron1 (c.67+1_67+25)	25	6	Intron1 (c.67+22_67+25)	15	4	Intron1 (c.67+1_67+25)	25	6							
	Intron2	Intron2 (c.190-25_190-1)	6 (récurrent)	25	6	Intron2 (c.190-25_190-1)	25	6													
	Exon3	Exon3 (c.190_295)	6 (récurrent)	106	106	Exon3 (c.190_295)	106	106													
	Intron3	Intron3 (c.295+1_295+25)	6 (récurrent)	25	6	Intron3 (c.295+1_295+25)	25	6													
	Exon7	pas de gap	4	0		Exon7 (c.751_762)	12	12													
	Intron7	pas de gap	4	0		Intron7 (c.762+1_762+25)	25	6													
	Exon8	Exon8 (c.977_1212)	6 (récurrent)	236	236	Exon8 (c.977_1212)	236	236													
	Exon8	Exon8 (c.1097_1235)	6 (récurrent)	139	139	Exon8 (c.1097_1235)	139	139													
	Intron10	pas de gap	4	0		Intron10 (c.1677-25_1677-1)	25	6	pas de gap	7	0	Intron10 (c.1677-21_1677-1)	21	6							
	Intron10								Intron10 (c.1677-25_1677-22)	gap de design	4	4	Intron10 (c.1677-25_1677-22)	4	4						
	Exon11	pas de gap	4	0		Exon11 (c.1677_1832)	156	156	pas de gap	7	0	Exon11 (c.1677_1731)	55	55							
	Exon11								Intron11 (c.1913+22_1913+25)	gap de design	4	4	Intron11 (c.1913+22_1913+25)	4	4						
	Intron11								pas de gap	1	0	Exon14 (c.3407_3408)	2	2							
	Intron14								pas de gap	3	0	Intron14 (c.3408+1_3408+25)	25	6							
	Exon15	Exon15 (c.3142)	6	1	1	Exon15 (c.3134_3214)	81	81	exon15 (c.2869_2886)	gap de design	18	18	exon15 (c.2869_2886)	18	18						
Exon15	Exon15 (c.3146_3209)	(suite)	64	64																	
Exon15	Exon15 (c.3769_3794)	6	26	26	Exon17 (c.3759_3794)	36	36														
Exon15	Exon15 (c.3794+1_3794+25)	6	25	25	Intron17 (c.3794+1_3794+25)	25	6														
			764	669		1008	875			42	18		246	166							
TBR1	NM_006593.3	Exon6	pas de gap	1	0	Exon6 (c.1435)	1	1	UTR 3' (c.*18_*25)	gap de design	8	8	UTR 3' (c.*18_*25)	8	8						
		Exon6	Exon6 (c.1613_1844)	6 (récurrent)	232	232	Exon6 (c.1613_1844)	232	232												
		UTR 3'																			
				232	232			233	233												
TCF4	NM_001083962.1			0	0			0	0					0	0						
UPF3B	NM_080632.2	exon7																			
		Intron9	Intron9 (c.1008-25_1008-21)	6 (récurrent)	5	0	Intron9 (c.1008-25_1008-21)	5	0	exon7 (c.652_729)	gap de design	78	78	exon7 (c.652_729)	78	78					
				5	0			5	0			78	78		78						
WDR45	NM_007075.3	Intron12	Intron12 (c.977-25_977-1)	6 (récurrent)	25	6	Intron12 (c.977-25_977-1)	25	6												
		Exon13	Exon13 (c.977_1086)	6 (récurrent)	110	110	Exon13 (c.977_1086)	110	110												
		3'UTR	3'UTR (c.*1_*25)	6 (récurrent)	25	6	3'UTR (c.*1_*25)	25	6												
					160	122		160	122			0	0		0						
Au total pour les 44 gènes					10295	8984		13864	12014			674	520		1474	1177					

n. pb. n. c. = nombre de bases non couvertes

Annexe 8 : Tableau des taux de couverture déduits de la pratique expérimentale des tests DI44 et DI117 sur les régions CDS ± 6 pb et CDS ± 25 pb

Gène	Taux de couverture par gène (régions codantes ± 25 pb vs ± 6 pb)							
	Test DI44				Test DI117			
	Min		Max		Min		Max	
	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6
ANKRD11 NM_001256182.1	99,86%	99,88%	96,94%	96,97%	99,10%	99,10%	99,10%	99,10%
ARID1B NM_020732.3	87,70%	88,19%	81,87%	82,36%	100,00%	100,00%	99,90%	99,90%
ATRX NM_000489.4	100,00%	100,00%	95,37%	95,75%	99,01%	99,10%	99,01%	99,10%
CASK NM_003688.3	96,74%	97,67%	93,10%	94,95%	100,00%	100,00%	100,00%	100,00%
CDKL5 NM_003159.2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
CTNNB1 NM_001904.3	100,00%	100,00%	99,93%	100,00%	100,00%	100,00%	100,00%	100,00%
CUL4B NM_003588.3	100,00%	100,00%	95,75%	96,76%	100,00%	100,00%	100,00%	100,00%
DLG3 NM_021120.3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
DYRK1A NM_001395.3	100,00%	100,00%	99,75%	100,00%	99,58%	99,58%	99,58%	99,58%
EP300 NM_001429.3	96,76%	97,19%	94,14%	95,17%	100,00%	100,00%	100,00%	100,00%
FOXG1 NM_005249.4	55,72%	57,17%	55,72%	57,17%	99,28%	100,00%	85,46%	86,18%
FOXP1 NM_032682.5	100,00%	100,00%	99,75%	100,00%	100,00%	100,00%	100,00%	100,00%
GATAD2B NM_020699.3	100,00%	100,00%	99,96%	100,00%	100,00%	100,00%	100,00%	100,00%
GRIA3 NM_000828.4	100,00%	100,00%	99,88%	100,00%	100,00%	100,00%	100,00%	100,00%
GRIN1 NM_007327.3	96,78%	97,77%	92,22%	94,21%	100,00%	100,00%	100,00%	100,00%
GRIN2A NM_000833.4	94,72%	95,59%	94,58%	95,59%	99,92%	100,00%	99,92%	100,00%
GRIN2B NM_000834.3	94,68%	95,92%	94,46%	95,71%	99,60%	100,00%	99,60%	100,00%
IL1RAPL1 NM_014271.3	100,00%	100,00%	99,69%	100,00%	100,00%	100,00%	100,00%	100,00%
IQSEC2 NM_001111125.2	87,12%	87,90%	81,77%	82,96%	98,54%	98,81%	97,05%	97,57%
KDM5C NM_004187.3	94,74%	96,56%	94,74%	96,56%	100,00%	100,00%	100,00%	100,00%
KMT2A NM_001197104.1	95,62%	95,96%	93,86%	94,51%	100,00%	100,00%	99,18%	99,18%
MECP2 NM_004992.3	100,00%	100,00%	99,44%	100,00%	100,00%	100,00%	100,00%	100,00%

Gène	Taux de couverture par gène (régions codantes ± 25 pb vs ± 6 pb)							
	Test DI44				Test DI117			
	Min		Max		Min		Max	
	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6	CDS ± 25	CDS ± 6
MED13L NM_015335.4	98,51%	98,97%	98,51%	98,97%	99,82%	100,00%	99,82%	100,00%
NAA10 NM_003491.3	93,59%	97,02%	93,59%	97,02%	100,00%	100,00%	100,00%	100,00%
PQBP1 NM_005710.2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
PTCHD1 NM_173495.2	91,65%	92,34%	91,65%	92,34%	100,00%	100,00%	100,00%	100,00%
RAI1 NM_030665.3	99,27%	99,59%	99,27%	99,59%	99,39%	99,39%	99,39%	99,39%
SATB2 NM_001172509.1	91,82%	93,08%	91,82%	93,08%	100,00%	100,00%	100,00%	100,00%
SCN2A NM_021007.2	99,70%	100,00%	99,63%	100,00%	99,66%	99,66%	99,66%	99,66%
SCN8A NM_014191.3	96,27%	96,62%	96,26%	96,62%	99,94%	100,00%	99,94%	100,00%
SETBP1 NM_015599.2	91,73%	92,10%	91,49%	91,67%	100,00%	100,00%	99,74%	99,74%
SHANK3 NM_033517.1	73,16%	75,59%	68,60%	71,93%	99,32%	99,79%	98,40%	99,43%
SLC16A2 NM_006517.4	99,79%	100,00%	99,79%	100,00%	100,00%	100,00%	99,74%	100,00%
SLC2A1 NM_006516.2	96,56%	98,48%	91,86%	94,74%	100,00%	100,00%	100,00%	100,00%
SLC6A8 NM_005629.3	61,26%	66,50%	54,26%	60,95%	99,37%	100,00%	99,10%	100,00%
SLC9A6 NM_001042537.1	95,29%	96,94%	91,16%	93,46%	100,00%	100,00%	99,72%	100,00%
SMARCA2 NM_003070.4	94,05%	94,97%	93,97%	94,97%	98,19%	98,19%	96,68%	96,98%
SMC1A NM_006306.3	96,67%	97,52%	92,73%	94,41%	100,00%	100,00%	100,00%	100,00%
STXBP1 NM_003165.3	99,93%	100,00%	99,93%	100,00%	100,00%	100,00%	100,00%	100,00%
SYNGAP1 NM_006772.2	84,66%	86,57%	79,77%	82,44%	99,16%	99,64%	95,06%	96,67%
TBR1 NM_006593.3	90,12%	90,12%	90,08%	90,08%	99,66%	100,00%	99,66%	100,00%
TCF4 NM_001083962.1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
UPF3B NM_080632.2	99,75%	100,00%	99,75%	100,00%	96,10%	96,10%	96,10%	96,10%
WDR45 NM_007075.3	89,91%	92,31%	89,91%	92,31%	100,00%	100,00%	100,00%	100,00%
Au total pour les 44 gènes	94,68%	95,36%	92,84%	93,80%	99,65%	99,73%	99,24%	99,39%

Annexe 9 : Tableau des profondeurs de lecture moyennes par gène, patient par patient, sur les 44 gènes (CDS ± 25), avec calcul de la profondeur de lecture moyenne par gène pour l'ensemble des 6 patients, et avec indication de la profondeur de lecture moyenne pour l'ensemble des 44 gènes par patient (PLM) : Test DI44 comparé au Test DI117

Gène	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6		Moyenne (écart-type) pour l'ensemble des patients	
	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117	Test DI44	Test DI117
ANKRD11	778	879	453	769	641	819	822	951	925	873	826	1115	741 (169)	901 (122)
ARID1B	894	967	510	832	717	868	871	1009	1015	943	884	1212	815 (177)	972 (134)
ATRX	886	924	489	816	364	315	466	476	495	464	854	1158	592 (221)	692 (325)
CASK	909	888	518	795	380	354	483	459	525	428	929	1078	624 (235)	667 (294)
CDKL5	1136	982	659	872	454	418	592	523	629	489	1087	1213	759 (282)	749 (320)
CTNNA1	956	1159	540	984	808	919	1132	1157	1234	1114	1036	1406	951 (249)	1123 (169)
CUL4B	899	893	501	751	363	340	488	462	508	452	890	1050	608 (228)	658 (283)
DLG3	1098	957	616	787	466	395	583	487	652	444	1112	1164	755 (279)	705 (314)
DYRK1A	905	978	524	859	734	834	1025	1042	993	966	917	1242	850 (189)	987 (147)
EP300	981	1077	598	938	821	962	1023	1117	1055	1036	951	1317	905 (171)	1074 (137)
FOXP1	334	811	180	668	262	669	343	850	345	783	299	936	294 (64)	786 (105)
FOXP1	746	1017	406	915	604	831	918	1073	1006	1007	823	1262	751 (219)	1018 (147)
GATAD2B	1085	1178	624	1005	882	1039	1174	1232	1282	1140	1113	1480	1027 (237)	1179 (170)
GRIA3	900	1057	489	908	370	409	525	535	538	503	939	1262	627 (235)	779 (346)
GRIN1	487	786	293	706	397	767	408	850	529	778	430	1062	424 (81)	825 (125)
GRIN2A	1146	1276	640	1049	990	1084	1235	1290	1334	1223	1174	1535	1086 (246)	1243 (174)
GRIN2B	910	1232	482	1027	747	1041	1033	1252	1128	1186	964	1485	877 (232)	1204 (168)
IL1RAPL1	960	1026	518	885	394	407	527	551	539	511	936	1245	646 (240)	771 (332)
IQSEC2	865	747	464	613	325	351	405	386	474	351	804	903	556 (223)	558 (233)
KDM5C	1230	1102	658	932	473	478	637	569	707	515	1208	1365	819 (320)	827 (364)
KMT2A	912	1095	529	947	768	1013	1028	1143	1098	1056	959	1380	882 (206)	1106 (150)
MECP2	998	882	595	766	423	407	514	463	546	434	975	1152	675 (248)	684 (301)
MED13L	961	1088	529	915	824	916	1116	1119	1166	1048	1011	1300	935 (232)	1064 (144)
NAA10	1433	1095	832	924	563	496	630	570	805	556	1342	1369	934 (367)	835 (353)
PQB1	1009	1205	595	991	453	522	568	601	656	555	1109	1509	732 (264)	897 (406)
PTCHD1	899	1269	511	1053	367	551	473	638	518	610	851	1546	603 (218)	944 (410)
RAI1	809	960	487	794	684	969	803	1007	964	899	960	1175	768 (165)	967 (126)
SATB2	1079	1001	589	833	899	845	1228	1028	1252	991	1086	1246	1022 (247)	991 (150)
SCN2A	1114	1026	627	889	884	782	1123	1053	1161	1020	1003	1232	985 (203)	1000 (153)
SCN8A	1157	1172	659	999	908	976	1130	1212	1261	1121	1096	1422	1035 (218)	1150 (163)
SETBP1	1167	1106	627	925	937	1009	1228	1114	1338	1031	1148	1306	1074 (255)	1082 (130)
SHANK3	227	746	157	633	236	740	206	774	289	711	232	947	224 (43)	758 (104)
SLC16A2	1251	1071	736	914	514	446	633	549	656	528	1184	1328	829 (310)	806 (354)
SLC2A1	1102	993	630	792	871	860	1128	984	1187	882	1070	1207	998 (210)	953 (146)
SLC6A8	406	713	256	619	263	345	257	379	350	348	352	861	314 (64)	544 (219)
SLC9A6	906	989	496	878	346	399	453	528	486	505	864	1206	592 (233)	751 (321)
SMARCA2	711	902	393	786	622	752	725	931	839	885	702	1113	665 (151)	895 (127)
SMC1A	1146	1203	644	1031	446	493	613	620	648	604	1157	1510	776 (300)	910 (403)
STXB1	1044	1057	613	962	856	982	1166	1162	1262	1069	1117	1397	1010 (237)	1105 (160)
SYNGAP1	858	840	499	745	706	749	880	880	1038	817	884	1066	811 (185)	850 (118)
TBR1	699	909	403	753	547	762	660	952	735	891	644	1124	615 (122)	898 (137)
TCF4	859	979	470	824	711	807	976	1009	1013	963	930	1202	826 (205)	964 (144)
UPF3B	753	731	430	658	292	286	406	391	434	358	753	931	512 (194)	559 (253)
WDR45	1094	1173	559	997	426	508	603	587	641	560	1098	1464	736 (288)	881 (392)
En moyenne pour les 44 gènes (écart-type)	919 (238)	1006 (148)	520 (133)	862 (121)	623 (222)	731 (252)	801 (303)	874 (290)	874 (317)	817 (270)	918 (239)	1240 (179)	776 (167)	922 (180)

Annexe 10 : Nombre de variants retenus selon les différents filtres bio-informatiques, dont nombre de variants *frameshift* (FS), sur les 44 gènes et sur les 6 patients, pour les deux tests DI44 et DI117

Patients	Test DI44 nbre de variants retenus (nbre de variant(s) confirmé(s) par Sanger)				Test DI117 (sur les 44 gènes) nbre de variants retenus (nbre de variant(s) confirmé(s) par Sanger)			
	Filtre 1		Filtre 2		Filtre 1		Filtre 2	
	1 ou 2 VC	2 VC	1 ou 2 VC	2 VC	1 ou 2 VC	2 VC	1 ou 2 VC	2 VC
Patient 1	120 (1/1)	51 (1/1)	10 (1/1)	6 (1/1)	30 (1/1)	30 (1/1)	6 (1/1)	6 (1/1)
dont nombre de FS	80	20	3	1	0	0	0	0
Patient 2	131 (1/1)	67 (1/1)	21 (1/1)	18 (1/1)	46 (1/1)	41 (1/1)	12 (1/1)	11 (1/1)
dont nombre de FS	76	20	2	0	2	0	0	0
Patient 3	128 (0/0)	61 (0/0)	15 (0/0)	11 (0/0)	43 (0/0)	39 (0/0)	19 (0/0)	15 (0/0)
dont nombre de FS	78	21	3	0	2	0	2	0
Patient 4	123 (1/1)	63 (1/1)	17 (1/1)	12 (1/1)	37 (1/1)	36 (1/1)	9 (1/1)	8 (1/1)
dont nombre de FS	79	24	4	1	1	1	1	1
Patient 5	137 (1/1)	60 (1/1)	15 (1/1)	9 (1/1)	37 (1/1)	33 (1/1)	8 (1/1)	8 (1/1)
dont nombre de FS	91	26	5	0	2	0	0	0
Patient 6	140 (1/1)	74 (1/1)	22 (1/1)	18 (1/1)	52 (1/1)	47 (1/1)	18 (1/1)	16 (1/1)
dont nombre de FS	81	25	4	1	3	1	1	1
Moyenne sur l'ensemble des patients	130	63	17	12	41	38	12	11
concernant les FS	81	23	4	1	2	0	1	0

Avec **Filtre 1** : - Localisation dans CDS \pm 25
 - Profondeur de lecture > 30X
 - VAF > 35%
 - Trouvé par 1 ou 2 *Variant Caller* (VC)

Filtre 2 : - Filtre 1
 - avec une fréquence du variant < 4 / run

Annexe 11 : Etude du coût des réactifs et du temps technique nécessaires pour le test DI44

Test DI44 : librairie par technique Ampliseq (Agilent) et séquençage sur le PGM™ Ion Torrent™ (Thermo Fisher Scientific) /Bioanalyzer PUCE 318	Tarif 1 patient	Tarif 15 patients
ION AMPLISEQ LIBRARY KIT 2.0-384LV (384 REACTIONS) REF 4480442	73,83	1328,91
ION AMPLISEQ SAMPLE ID PANEL (96 REACT) REF 4479790	3,75	63,75
ION AMPLISEQ CUSTOM PRIMER POOL (1 PRIMER PAIR) REF 10668077	2,65	26,53
ION XPRESS BARCODE ADAPTATORS KIT(1 KIT DE 160 REACTIONS)	9,56	172,13
KIT HAUTE SENSIBILITE (10 PUCES) REF 5067-4626 (AGILENT)	36,46	36,46
AGENCOURT AMPURE XP (60 ml) REF A63881 (BECKMAN COULTER)	1,80	30,63
TUBES 1,5 ml	0,23	0,39
BARRETTES DE MICROTUBES 0,2ml	1,08	4,32
POINTES FILTRES (9x10)	2,18	26,07
GANTS		
Plaques Lo bind 96 puits (10 plaques)	5,16	10,31
TUBES LOBIND 1,5ml (5X50) REF 033871 (D DUTSCHER)	0,09	0,79
ETHANOL ABSOLU (4X25ml) REF 02860 (SIGMA)	0,36	6,53
Temps technicien : 5h-8h		
<u>Thermocycleur 1,5h</u>		
<u>DynaMag™ -96 Side 12331D</u>		
<u>Bioanalyseur 2100 (coût par run)</u>		
Coût Librairies Ampliseq (Agilent) (par échantillon)	137,15	113,79
ION PGM Hi-Q VIEW OT2 KIT (8 réactions de 200 ou 400 bases) REF A29900	140,40	280,80
ION - DYNABEADS MYONE STREPTAVIDIN C1 (2 mL) REF 65001	2,24	4,48
ION SPHERE QUALITY CONTROL KIT (20 DOSAGES) REF 4468656		
ION - QUBIT ASSAY TUBES (500/pkg) REF Q32856		
TUBES 0,2 ml PP (4X500) REF 72737 (SARSTEDT)	0,06	0,12
TUBES LOBIND 1,5ml (5X50) REF 033871 (D DUTSCHER)	0,31	0,62
POINTES FILTRES (9x10)	1,65	5,40
Temps technicien (2h)		
<u>Thermocycleur (0,1h)</u>		
<u>Ion OneTouch 2 System (1 run)</u>		
<u>Qubit®2.0 Quantitation Starter Kit (1 run)</u>		
Coût OT2+ES (par échantillon)	144,66	19,43
UNITES FILTRATION NALGENE 0,22µm 1000ml (X12) REF 029665 (D DUTSCH	6,95	6,95
ION PGM Hi-Q VIEW SEQUENCING (8 runs de 200 bases ou 4 runs de 400 b	247,50	495,00
ION 318 BC CHIP KIT (4 PUCES) REF 4488146	488,00	976,00
ION PGM WASH BOTTLES (X 40 utilisations) ref A25591	1,63	1,63
TUBES LOBIND 1,5ml (5X50) REF 033871 (D DUTSCHER)	0,09	0,18
POINTES FILTRES (9x10)	0,49	0,98
POINTES LTS RAININ GREENPACK FILTRE GP-L10F 20µl (10X96) REF 732-026	0,63	1,26
POINTES LTS RAININ GREENPACK FILTRE GP-L200F 200µl (10X96) REF RAINI	0,08	0,16
Temps technicien (4h)		
<u>Thermocycleur (0,1h)</u>		
<u>ION PGM System (1 run)</u>		
<u>Mini-centrifugeuse, Mini Galaxi 521-2844 (1 run)</u>		
Coût Séquençage PGM (par échantillon)	745,36	98,81
PC INFORMATIQUE (1,5h/patient ou témoin)		
LOGICIEL NEXTGENE (1h/patient)		
Temps Ingénieur (1,5h/patient)		
Temps Biologiste (2h/patient)		
Coût analyse Test DI44 par patient	1027,17	232,03

Annexe 12 : Etude du coût des réactifs et du temps technique nécessaires pour le test DI117

Test DI117 : librairie par capture SureSelect (Agilent) et séquençage sur le MiSeq (Illumina)	Tarif 1 patient	Tarif 15 patients
SureSelect XT HS / Low Input enzymatic fragmentation kit, 96 réactions	3,25	52,00
SureSelect XT Low Input 1-96 + 0.5-2.9 Mb, (96 réactions) (Tier2)	147,99	2367,90
QUBIT DSDNA HS ASSAY KIT (100 TESTS) REF Q32851	2,07	12,42
KIT HAUTE SENSIBILITE (10 PUCES) REF 5067-4626 (AGILENT)	36,46	145,824
AGENCOURT AMPURE XP (60 ml) REF A63881 (BECKMAN COULTER)	3,60	28,83
TUBES 1,5 ml	0,23	0,39
BARRETTES DE MICROTUBES 0,2ml	1,08	2,70
POINTES FILTRES (9x10)	2,18	19,21
GANTS		
PLAQUE PCR 96 PUIITS TWIN TEC LOBIND DEMI-JUPE INCOLORE	5,16	10,31
TUBES LOBIND 1,5ml (5X50) REF 033871 (D DUTSCHER)	0,09	0,44
ETHANOL ABSOLU (4X25ml) REF 02860 (SIGMA)	0,36	2,72
Temps technicien : 8h-10h		
<u>Thermocycleur 1,5h</u>		
<u>DynaMagTM -96 Side 12331D</u>		
<u>Bioanalyseur 2100 (coût par run)</u>		
Coût Librairies SureSelect (Agilent) (par échantillon)	202,47	176,18
MISEQ REAGENT KIT V2 (300 CYCLES) REF MS-102-2002	931,5	931,50
PHIX CONTROL KIT V3 REF FC-110-3001	1,43	1,43
TUBES LOBIND 1,5ml (5X50) REF 033871 (D DUTSCHER)	0,18	0,18
POINTES FILTRES (9x10)	0,61	0,61
Temps technicien (1h)		
<u>Thermocycleur (0,1h)</u>		
Coût RUN MiSeq (par échantillon)	933,72	62,25
PC INFORMATIQUE (1,5h/patient ou témoin)		
Temps Ingénieur (1,5h/patient)		
Temps Biologiste (2h/patient)		
Coût analyse Test DI117 par patient	1136,19	238,43

Annexe 13 : Tableau des profondeurs de lecture moyennes par gène, patient par patient, sur les 117 gènes (CDS ± 25) ; avec calcul de la profondeur de lecture moyenne par gène pour l'ensemble des 15 patients, et avec indication de la profondeur de lecture moyenne pour l'ensemble des 117 gènes par patient (PLM) pour le Test DI117

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
ADNP	1347	1177	1091	1362	1340	1683	1411	1453	1298	2531	1244	1186	1430	1047	1116	1381 (358)
AHDC1	358	297	357	370	317	443	341	360	294	861	382	377	459	350	533	407 (140)
ANKRD11	879	769	819	951	873	1115	912	941	810	1916	842	910	1116	792	1266	994 (290)
ARID1B	967	832	868	1009	943	1212	969	1009	904	2142	1045	1021	1183	915	1254	1085 (317)
ASH1L	1101	960	881	1160	1067	1395	1150	1188	1045	2139	1005	981	1187	864	950	1138 (309)
ASXL3	975	817	785	983	939	1182	974	995	932	2066	972	935	1071	800	910	1022 (306)
ATRX	924	816	315	476	464	1158	953	987	464	1640	398	751	875	652	692	771 (345)
AUTS2	700	596	548	707	652	849	708	719	653	1381	675	661	744	594	747	729 (194)
BCL11A	717	605	581	735	678	881	736	743	655	1470	706	673	784	605	728	753 (212)
BRAF	718	651	523	744	714	892	749	737	716	1283	645	601	728	527	538	718 (185)
CASK	888	795	354	459	428	1078	923	924	420	1680	423	794	909	696	761	769 (342)
CDK13	651	579	556	686	632	802	649	672	628	1403	690	649	737	563	628	702 (204)
CDKL5	982	872	418	523	489	1213	1013	1015	475	2161	516	1005	1161	893	1088	922 (439)
CHD2	762	680	620	811	734	962	788	799	740	1568	746	714	825	615	659	802 (230)
CHD4	952	828	770	972	909	1188	984	1001	876	1810	838	829	992	741	805	966 (260)
CHD7	886	748	727	893	851	1084	887	903	830	1831	873	850	979	759	812	928 (266)
CHD8	920	774	722	927	887	1135	922	947	852	1777	855	812	981	730	788	935 (256)
CNKS2	725	626	233	366	345	877	702	730	341	1392	339	649	736	557	558	612 (288)
CREBBP	659	582	545	692	626	815	682	678	596	1303	611	604	745	536	661	689 (185)
CSNK2A1	896	821	686	934	889	1107	949	948	871	1617	776	763	951	674	720	907 (229)
CTCF	860	753	753	901	827	1110	898	925	803	1737	827	776	953	708	755	906 (252)
CTNNA1	1159	984	919	1157	1114	1406	1155	1179	1071	2273	1112	1065	1218	932	1023	1184 (325)
CUL4B	893	751	340	462	452	1050	894	904	427	1646	404	748	884	648	694	747 (333)
DDX3X	861	754	323	440	404	1051	847	909	404	1542	359	713	869	615	633	715 (324)
DLG3	957	787	395	487	444	1164	940	967	432	1998	487	935	1096	861	1018	865 (410)
DNM1	691	591	564	728	650	853	712	718	633	1339	621	620	794	581	679	718 (189)
DNMT3A	577	488	483	583	529	723	577	593	532	1171	566	536	647	493	602	607 (168)
DYNC1H1	913	787	737	932	863	1119	900	947	826	1726	814	786	942	708	737	916 (248)
DYRK1A	978	859	834	1042	966	1242	1013	1054	971	2090	995	951	1145	842	1051	1069 (303)
EBF3	604	518	469	638	587	745	633	638	598	1147	556	558	633	496	544	624 (160)
EEF1A2	413	375	382	456	400	528	443	440	370	886	400	406	518	379	506	460 (129)
EHMT1	671	592	564	706	640	833	675	689	616	1339	616	614	767	572	631	702 (191)
EP300	1077	938	962	1117	1036	1317	1096	1126	995	2299	1064	1052	1295	919	1290	1172 (336)
FGFR2	950	811	763	960	896	1142	934	968	894	1855	908	860	980	776	826	968 (263)
FGFR3	263	220	278	274	239	325	247	260	218	669	291	295	358	274	398	307 (111)
FLNA	632	540	269	322	308	798	623	638	286	1321	294	629	733	553	716	577 (275)
FOXP1	811	668	669	850	783	936	811	845	742	1765	826	806	1037	786	1347	912 (288)
FOXP1	1017	915	831	1073	1007	1262	1042	1064	963	2098	1011	959	1183	871	1020	1088 (300)
GABRB3	886	754	714	923	884	1089	875	945	865	1758	855	824	946	722	796	922 (250)
GATAD2B	1178	1005	1039	1232	1140	1480	1232	1252	1109	2507	1166	1144	1373	1002	1270	1275 (365)

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
GNAI1	608	538	440	633	603	725	619	627	599	1052	532	474	580	431	439	593 (153)
GNAO1	817	732	635	843	806	1045	819	839	793	1534	734	730	881	639	698	836 (218)
GRIA3	1057	908	409	535	503	1262	1022	1052	486	2123	494	971	1116	878	975	919 (432)
GRIN1	786	706	767	850	778	1062	845	870	730	1801	802	835	1086	777	1290	932 (287)
GRIN2A	1276	1049	1084	1290	1223	1535	1259	1300	1179	2702	1324	1281	1469	1154	1445	1371 (392)
GRIN2B	1232	1027	1041	1252	1186	1485	1223	1258	1127	2639	1284	1237	1442	1117	1413	1331 (386)
HDAC8	866	739	336	433	428	1062	873	898	409	1701	412	811	902	693	746	754 (347)
HNRNPU	697	632	508	700	707	863	717	740	700	1218	582	566	677	496	539	689 (177)
HUWE1	932	785	353	465	435	1118	898	944	437	1776	429	813	943	742	748	788 (362)
IL1RAPL1	1026	885	407	551	511	1245	1041	1030	491	2111	513	966	1156	859	1036	922 (427)
IQSEC2	747	613	351	386	351	903	730	757	338	1768	409	836	954	732	1137	734 (379)
ITPR1	882	759	673	911	864	1077	893	921	836	1659	818	771	904	697	690	890 (238)
KANSL1	1135	1008	888	974	1076	1168	1160	1197	1068	2284	1050	846	992	904	1167	1128 (338)
KAT6A	928	801	686	956	878	1123	963	965	869	1790	856	806	981	705	806	941 (260)
KAT6B	867	736	725	883	826	1061	845	890	809	1798	850	810	958	740	820	908 (261)
KCNB1	1052	894	861	1100	988	1339	1092	1100	971	2151	969	979	1191	853	1015	1104 (316)
KCNQ2	369	303	336	375	346	449	362	375	310	790	360	373	435	331	415	395 (117)
KDM5C	1102	932	478	569	515	1365	1080	1130	487	2391	575	1121	1324	1025	1351	1030 (498)
KIAA2022	1349	1141	496	686	639	1620	1352	1378	613	2661	636	1214	1356	1067	1179	1159 (544)
KIF1A	521	435	439	533	475	624	501	521	451	1055	483	483	588	443	549	540 (153)
KMT2A	1095	947	1013	1143	1056	1380	1115	1155	1037	2451	1169	1107	1329	986	1254	1216 (362)
KMT2D	745	613	639	757	703	925	744	742	658	1564	709	709	851	640	799	786 (230)
MAGEL2	858	645	836	850	808	1025	805	854	732	2290	1054	1021	1166	895	1507	1023 (407)
MECP2	882	766	407	463	434	1152	926	937	424	2001	458	941	1150	800	1264	867 (428)
MED12	1029	853	393	494	472	1244	1004	1034	479	1907	452	904	1059	826	883	869 (395)
MED13L	1088	915	916	1119	1048	1300	1095	1127	1008	2321	1120	1087	1253	966	1138	1167 (337)
MEF2C	830	738	620	843	825	1031	860	889	813	1568	777	753	851	628	621	843 (229)
MTOR	948	824	776	984	911	1173	952	979	864	1842	875	840	1022	762	804	971 (264)
MYT1L	828	691	646	847	790	1008	841	848	784	1511	741	709	833	627	708	827 (212)
NAA10	1095	924	496	570	556	1369	1124	1169	518	2306	520	1139	1344	1040	1416	1039 (485)
NFIX	735	621	587	767	665	919	748	760	656	1379	618	645	799	598	744	749 (196)
NSD1	947	821	794	989	912	1179	968	999	881	1916	897	870	1054	756	835	988 (279)
OPHN1	854	705	315	425	404	1041	869	862	404	1569	388	732	813	657	686	715 (323)
PACS1	790	684	629	824	740	991	772	815	721	1500	695	694	856	618	716	803 (215)
PAX6	859	730	659	872	822	1045	834	876	813	1600	773	750	831	665	651	852 (231)
PHIP	712	615	460	747	713	872	721	747	696	1254	619	569	674	503	507	694 (190)
POGZ	994	854	808	1039	942	1205	1001	1026	914	2004	930	895	1085	789	860	1023 (293)
PPP2R5D	1095	921	884	1106	1034	1354	1082	1106	1010	2126	977	962	1195	874	907	1109 (309)
PQBP1	1205	991	522	601	555	1509	1180	1227	514	2665	608	1250	1460	1125	1554	1131 (565)
PTCHD1	1269	1053	551	638	610	1546	1249	1260	603	2711	620	1243	1420	1147	1416	1156 (553)
PTEN	657	612	442	696	681	828	699	723	650	1207	591	556	672	483	484	665 (182)
PTPN11	723	664	579	758	715	891	738	776	720	1358	654	602	744	550	559	735 (195)
PUF60	540	488	460	557	508	679	548	544	485	1068	482	508	622	451	581	568 (151)

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
PURA	420	335	337	435	384	528	413	421	356	881	427	418	513	363	472	447 (133)
RAI1	960	794	969	1007	899	1175	949	1018	848	2370	1052	1098	1339	962	1741	1145 (410)
RPS6KA3	606	530	203	306	293	729	628	640	303	1002	241	461	565	409	384	487 (216)
SATB2	1001	833	845	1028	991	1246	1022	1024	940	2243	1037	1051	1194	916	1096	1098 (336)
SCN1A	859	738	544	867	824	1039	858	864	812	1638	828	744	876	649	725	858 (244)
SCN2A	1026	889	782	1053	1020	1232	1042	1061	985	2089	1017	944	1116	819	977	1070 (303)
SCN8A	1172	999	976	1212	1121	1422	1186	1218	1107	2403	1151	1110	1341	988	1189	1240 (344)
SETBP1	1106	925	1009	1114	1031	1306	1061	1096	1021	2605	1223	1196	1358	1058	1387	1233 (402)
SETD5	962	819	759	999	924	1161	960	978	904	1905	927	882	1028	796	858	991 (272)
SHANK3	746	633	740	774	711	947	761	801	668	1830	821	850	1041	792	1328	896 (310)
SLC16A2	1071	914	446	549	528	1328	1072	1089	486	2223	534	1016	1245	925	1287	981 (462)
SLC2A1	993	792	860	984	882	1207	949	969	827	2153	1003	976	1181	901	1212	1059 (329)
SLC6A1	826	718	683	839	805	1006	828	832	779	1666	774	783	913	680	806	862 (237)
SLC6A8	713	619	345	379	348	861	733	744	321	1611	357	766	919	707	1105	702 (348)
SLC9A6	989	878	399	528	505	1206	1027	1074	485	1913	460	898	1079	813	941	880 (389)
SMARCA2	902	786	752	931	885	1113	911	923	850	1853	907	881	1023	783	930	962 (263)
SMARCA4	570	506	484	589	543	722	600	588	525	1165	518	541	671	481	579	605 (168)
SMC1A	1203	1031	493	620	604	1510	1225	1237	563	2403	586	1156	1347	1019	1226	1081 (492)
STXBP1	1057	962	982	1162	1069	1397	1142	1149	1035	2318	1052	1076	1321	967	1207	1193 (335)
SYNGAP1	840	745	749	880	817	1066	900	859	772	1801	834	840	1077	776	1085	936 (265)
TBL1XR1	652	573	451	699	672	817	677	693	657	1185	578	542	632	465	484	652 (178)
TBR1	909	753	762	952	891	1124	940	960	858	2066	992	969	1153	858	1393	1039 (327)
TCF20	1099	941	914	1167	1054	1392	1141	1150	1002	2195	1032	1031	1258	904	1059	1156 (315)
TCF4	979	824	807	1009	963	1202	997	1017	913	2115	991	966	1142	860	1000	1052 (312)
TRIM27	814	700	728	849	801	1054	798	865	770	1881	872	870	989	768	1008	918 (285)
TRIO	801	687	624	829	771	982	810	831	765	1524	737	696	853	641	664	814 (218)
TRIP12	837	722	626	862	813	1037	858	874	800	1565	779	728	832	621	657	841 (228)
UBE3A	922	826	646	980	923	1156	974	972	913	1659	829	769	897	678	738	925 (242)
UPF3B	731	658	286	391	358	931	754	761	375	1224	307	592	728	516	577	613 (257)
USP7	914	845	784	987	953	1174	967	1023	901	1757	867	835	998	727	830	971 (244)
USP9X	806	697	288	413	389	978	812	832	378	1436	345	652	771	554	564	661 (300)
WAC	710	639	549	735	672	862	744	746	702	1332	650	622	747	535	583	722 (190)
WDFY3	810	680	584	817	789	956	797	810	770	1528	734	693	797	604	613	799 (225)
WDR45	1173	997	508	587	560	1464	1162	1243	518	2628	573	1215	1409	1092	1444	1105 (551)
En moyenne (écart-type)	885 (208)	759 (176)	643 (217)	807 (255)	755 (242)	1088 (255)	893 (211)	916 (217)	732 (236)	1790 (469)	754 (255)	825 (217)	979 (252)	735 (195)	877 (305)	896 (272)

Annexe 14 : Tableau des taux de couverture >30X sur les 117 gènes (CDS + 25) du test D1117, détaillés par gène, patient par patient, avec calcul de la moyenne par gène pour l'ensemble des 15 patients

Test D1117

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
ADNP	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
AHDC1	98,50%	98,00%	99,00%	98,70%	98,00%	99,00%	97,60%	98,40%	98,40%	99,90%	99,60%	99,20%	99,60%	98,40%	99,50%	98,79% (0,68%)
ANKRD11	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
ARID1B	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,99% (0,03%)
ASH1L	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
ASXL3	100,00%	100,00%	100,00%	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,99% (0,03%)
ATRX	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
AUTS2	99,80%	99,60%	99,60%	99,60%	99,60%	100,00%	99,60%	99,60%	99,80%	100,00%	100,00%	99,80%	100,00%	99,60%	100,00%	99,77% (0,18%)
BCL11A	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10%	96,10% (0,00%)
BRAF	94,10%	94,10%	94,10%	94,10%	94,10%	94,10%	94,10%	94,10%	94,10%	97,70%	94,10%	94,10%	94,10%	94,10%	97,50%	94,57% (1,23%)
CASK	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CDK13	89,40%	87,30%	91,20%	88,50%	88,00%	88,80%	86,90%	87,00%	86,80%	99,80%	91,80%	93,60%	96,00%	91,70%	98,70%	91,03% (4,28%)
CDKL5	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CHD2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CHD4	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CHD7	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CHD8	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CNKSR2	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,98% (0,08%)
CREBBP	97,70%	97,30%	97,80%	97,80%	97,40%	98,10%	97,30%	97,80%	97,60%	98,30%	98,00%	98,10%	97,80%	97,60%	98,30%	97,79% (0,32%)
CSNK2A1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CTCF	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CTNNB1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
CUL4B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
DDX3X	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
DLG3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
DNM1	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70%	92,70% (0,00%)
DNMT3A	96,80%	96,80%	96,80%	96,80%	96,80%	96,80%	96,80%	96,80%	96,80%	99,90%	96,80%	96,80%	96,80%	96,80%	96,80%	97,01% (0,80%)
DYNC1H1	99,80%	99,00%	99,70%	99,70%	99,40%	99,80%	99,00%	99,30%	98,90%	100,00%	99,90%	100,00%	99,80%	99,90%	100,00%	99,61% (0,39%)
DYRK1A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
EBF3	100,00%	99,00%	100,00%	100,00%	99,00%	100,00%	98,50%	99,00%	97,60%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,54% (0,75%)
EEF1A2	89,80%	89,80%	89,80%	89,80%	89,80%	89,80%	89,80%	89,80%	89,80%	95,80%	89,80%	89,80%	89,80%	89,80%	94,10%	90,49% (1,84%)
EHMT1	98,30%	98,10%	98,60%	98,20%	97,80%	98,60%	97,70%	97,80%	97,80%	98,60%	98,60%	98,60%	98,60%	98,50%	98,60%	98,29% (0,36%)
EP300	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
FGFR2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
FGFR3	95,10%	95,10%	95,10%	94,90%	93,50%	95,10%	94,30%	95,10%	91,50%	97,50%	95,50%	95,10%	96,30%	97,40%	97,10%	95,24% (1,52%)
FLNA	100,00%	99,60%	99,80%	99,60%	99,30%	100,00%	100,00%	100,00%	99,40%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	99,83% (0,25%)
FOXP1	91,60%	86,20%	91,60%	91,30%	87,20%	90,70%	90,50%	86,80%	88,30%	100,00%	94,30%	95,10%	93,10%	91,00%	100,00%	91,85% (4,20%)
FOXP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GABRB3	94,80%	94,70%	96,20%	94,80%	94,10%	95,40%	94,80%	95,30%	94,20%	97,50%	96,30%	95,40%	95,40%	94,80%	96,20%	95,33% (0,91%)
GATAD2B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GNAI1	100,00%	98,20%	99,70%	98,20%	97,50%	100,00%	99,00%	97,70%	94,60%	100,00%	100,00%	100,00%	100,00%	100,00%	99,70%	98,97% (1,52%)
GNAO1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
GRIA3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GRIN1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GRIN2A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
GRIN2B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
HDAC8	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
HNRNPU	98,60%	97,50%	96,90%	98,20%	98,80%	99,50%	98,50%	99,00%	98,20%	100,00%	98,80%	98,70%	99,00%	98,80%	100,00%	98,70% (0,82%)
HUWE1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
IL1RAPL1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
IQSEC2	99,40%	99,50%	98,40%	98,60%	99,30%	99,50%	99,40%	99,40%	98,50%	99,70%	99,20%	99,70%	99,70%	99,50%	99,70%	99,30% (0,44%)
ITPR1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
KANSL1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
KAT6A	100,00%	100,00%	99,60%	99,50%	99,80%	100,00%	99,60%	99,10%	99,60%	100,00%	100,00%	100,00%	100,00%	100,00%	99,70%	99,79% (0,27%)
KAT6B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
KCNB1	97,90%	96,80%	98,00%	96,90%	97,40%	97,00%	96,90%	97,00%	96,80%	99,40%	97,70%	97,50%	98,20%	97,20%	98,00%	97,51% (0,71%)
KCNQ2	98,60%	96,00%	98,20%	97,60%	97,90%	98,30%	97,70%	98,60%	96,30%	100,00%	99,20%	98,90%	99,20%	96,70%	100,00%	98,21% (1,22%)
KDM5C	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
KIAA2022	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
KIF1A	97,70%	95,60%	97,70%	96,00%	97,70%	97,30%	97,10%	97,40%	97,10%	99,70%	96,60%	98,80%	98,40%	97,80%	99,50%	97,63% (1,14%)
KMT2A	99,70%	99,50%	100,00%	100,00%	99,20%	99,60%	99,60%	99,50%	99,40%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,77% (0,28%)
KMT2D	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MAGEL2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MECP2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MED12	100,00%	100,00%	99,10%	99,80%	99,50%	100,00%	100,00%	100,00%	99,40%	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	99,85% (0,28%)
MED13L	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MEF2C	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MTOR	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
MYT1L	100,00%	98,90%	100,00%	100,00%	100,00%	100,00%	99,40%	100,00%	99,80%	100,00%	100,00%	100,00%	99,20%	99,40%	99,10%	99,72% (0,40%)
NAA10	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
NFIX	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70%	95,70% (0,00%)
NSD1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
OPHN1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PACS1	97,10%	95,30%	96,50%	96,80%	96,30%	96,70%	96,40%	95,50%	95,00%	99,80%	96,70%	96,70%	98,70%	97,90%	99,70%	97,01% (1,45%)
PAX6	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,90%	100,00%	100,00%	99,90%	99,99% (0,04%)
PHIP	97,50%	97,50%	97,40%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	98,30%	98,00%	98,00%	97,70%	97,40%	97,70%	97,63% (0,26%)
POGZ	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PPP2R5D	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PQB1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PTCHD1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
PTEN	100,00%	100,00%	95,40%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	96,90%	98,00%	98,00%	99,50%	93,70%	98,77% (2,00%)
PTPN11	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50%	97,50% (0,00%)
PUF60	100,00%	100,00%	100,00%	100,00%	99,50%	100,00%	100,00%	100,00%	99,80%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,95% (0,14%)
PURA	88,90%	84,70%	89,20%	87,10%	87,30%	85,80%	88,60%	86,40%	86,80%	100,00%	94,90%	90,40%	97,90%	91,00%	100,00%	90,60% (5,15%)
RAI1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
RPS6KA3	97,10%	96,40%	93,80%	96,40%	96,30%	96,40%	96,40%	96,40%	93,80%	99,00%	93,80%	93,80%	97,70%	95,60%	97,10%	96,00% (1,57%)
SATB2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SCN1A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SCN2A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)

Gène	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Moyenne (écart-type) pour l'ensemble des patients
SCN8A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SETBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	100,00%	99,98% (0,08%)
SETD5	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SHANK3	98,60%	98,70%	99,10%	99,20%	98,90%	99,10%	98,90%	99,10%	98,80%	99,50%	99,30%	99,00%	99,50%	99,30%	99,50%	99,10% (0,29%)
SLC16A2	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,98% (0,08%)
SLC2A1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SLC6A1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SLC6A8	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,98% (0,08%)
SLC9A6	100,00%	100,00%	100,00%	99,80%	99,80%	100,00%	100,00%	100,00%	99,70%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,95% (0,10%)
SMARCA2	98,80%	98,50%	100,00%	99,90%	98,80%	99,60%	99,40%	99,60%	98,60%	100,00%	99,90%	99,90%	100,00%	100,00%	100,00%	99,53% (0,57%)
SMARCA4	99,60%	99,40%	99,60%	99,70%	99,70%	99,70%	99,60%	99,70%	99,50%	99,80%	99,70%	99,70%	99,80%	99,60%	99,80%	99,66% (0,11%)
SMC1A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
STXBP1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
SYNGAP1	97,70%	99,30%	97,70%	98,60%	99,10%	99,60%	99,60%	97,70%	99,50%	99,80%	98,50%	99,60%	99,60%	99,80%	99,80%	99,06% (0,81%)
TBL1XR1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
TBR1	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
TCF20	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
TCF4	100,00%	99,90%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,99% (0,03%)
TRIM27	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
TRIO	96,60%	95,60%	96,60%	96,10%	95,60%	96,90%	96,20%	96,10%	95,90%	98,30%	96,80%	97,20%	98,60%	98,20%	98,80%	96,90% (1,09%)
TRIP12	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
UBE3A	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
UPF3B	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
USP7	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
USP9X	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
WAC	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60%	96,60% (0,00%)
WDFY3	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
WDR45	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00% (0,00%)
En moyenne (écart-type)	99,30% (2,22%)	99,10% (2,73%)	99,30% (2,21%)	99,30% (2,37%)	99,20% (2,55%)	99,30% (2,41%)	99,20% (2,40%)	99,20% (2,60%)	99,10% (2,69%)	99,70% (1,09%)	99,40% (1,92%)	99,40% (1,98%)	99,50% (1,69%)	99,40% (2,07%)	99,60% (1,35%)	99,33% (0,17%)

Gène inclus dans le test DI44

MISE AU POINT D'UNE MÉTHODE DE SÉQUENÇAGE DE NOUVELLE GÉNÉRATION POUR LE DIAGNOSTIC GÉNÉTIQUE DE LA DÉFICIENCE INTELLECTUELLE

RÉSUMÉ EN FRANÇAIS :

La déficience intellectuelle (DI) constitue un problème majeur de santé publique. Son extrême hétérogénéité génétique, avec plus de 700 gènes identifiés, complique son diagnostic étiologique. L'arrivée des technologies de séquençage haut-débit (NGS) a enrichi l'offre de tests diagnostiques en permettant d'explorer en un examen un ensemble de gènes.

Dans ce contexte, nous avons développé un test NGS de première intention, appelé test DI117, ciblant 117 gènes impliqués dans la DI. Le test DI117 a montré une sensibilité et une spécificité au moins équivalentes au test DI44 sur leurs 44 gènes communs chez 6 patients. Les résultats obtenus chez 15 patients ont également suggéré une performance remarquable pour la majorité des 73 gènes supplémentaires. La comparaison des données obtenues avec 3 autres tests NGS validés pour 5 patients a permis d'évaluer sa reproductibilité. Ce test paraît être un test diagnostique efficace, et sa prescription médicale semble pertinente chez les patients DI.

TITRE EN ANGLAIS : Design of a next generation sequencing method for the genetic diagnosis of Intellectual Disability

RÉSUMÉ EN ANGLAIS :

Intellectual Disability (ID) is a serious public health problem. With over 700 genes identified, its extreme genetic diversity makes its etiological diagnosis a challenge. Massive parallel sequencing (NGS) has widened the diagnostic testing options by enabling the investigation of many genes in one examination.

In this context, we have designed a NGS first-line test, named DI117 test, covering 117 genes linked with ID. The DI117 test has proven at least equal sensitivity and specificity compared to the DI44 test for their 44 genes in common among 6 patients. Results gathered among 15 patients also suggest a striking efficacy for most of the 73 additional genes in its panel. Comparing the gathered data with 3 other approved NGS tests among 5 patients has enabled us to verify its reproducibility. This test seems to be an effective diagnostic test, and so its medical prescription appears relevant for ID patients.

MOTS-CLÉS : Déficience Intellectuelle, DI, diagnostic génétique, séquençage Illumina, NGS, séquençage haut-débit, enrichissement par capture, DI117, test diagnostique.

DISCIPLINE ADMINISTRATIVE : Médecine Spécialité Biologie Médicale

Directeur de thèse : Dr. Eric BIETH

Date et lieu de soutenance : Le 28 mars 2019 à la Faculté de Médecine Toulouse-Purpan

INTITULÉ ET ADRESSE DE L'UFR OU DU LABORATOIRE :

Université Toulouse III - Paul Sabatier, Faculté de Médecine Rangueil
133 route de Narbonne – 31062 Toulouse Cedex 04 - France