



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

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



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

UNIVERSITE CLAUDE BERNARD - LYON I
INSTITUT DES SCIENCES PHARMACEUTIQUES ET BIOLOGIQUES
FACULTE DE PHARMACIE DE LYON
8, avenue Rockefeller – 69373 LYON Cedex 08

Année 2016

THESE n°03-2016

MÉMOIRE
DU DIPLOME D'ÉTUDES SPÉCIALISÉES DE
BIOLOGIE MEDICALE

Travail effectué dans le laboratoire de Génétique Moléculaire de l'Hôpital Edouard Herriot (Lyon)
Sous la direction de M. le Docteur Stéphane Pinson

Soutenu devant le jury interrégional le 12 février 2016

Par M. **Adrien BUISSON**

Né le 12 Janvier 1988 à Lyon 2^{ème} (69)

Conformément aux dispositions du décret n° 90-810 du 10 septembre 1990, tient lieu de

THÈSE

POUR LE DIPLOME D'ÉTAT DE DOCTEUR EN PHARMACIE

**MISE AU POINT D'UNE TECHNIQUE DE SÉQUENÇAGE HAUT-
DÉBIT PAR TECHNOLOGIE ION TORRENT™ POUR L'ÉTUDE DES**

GÈNES *NF1* ET *SPRED1*

JURY

PRESIDENT : Mme le Professeur Christine VINCIGUERRA (PU-PH)

MEMBRES : M. le Professeur Alain CALENDER (PU-PH)

Mme le Docteur Claire BARDEL (MCU-PH)

M. le Docteur Stéphane PINSON (PH)

UNIVERSITE CLAUDE BERNARD LYON 1

- Président de l'Université M. François-Noël GILLY
- Vice Président du Conseil d'Administration M. Hamda BEN HADID
- Vice-Président de la Commission Recherche M. Germain GILLET
- Vice-Président de la Formation et de la Vie Universitaire M. Philippe LALLE

Composantes de l'Université Claude Bernard Lyon 1

SANTE

- UFR de Médecine Lyon Est Directeur : M. Jérôme ETIENNE
- UFR de Médecine Lyon Sud Charles Mérieux Directeur : Mme Carole BURILLON
- Institut des Sciences Pharmaceutiques et Biologiques Directrice : Mme Christine VINCIGUERRA
- UFR d'Odontologie Directeur : M. Denis BOURGEOIS
- Institut des Techniques de Réadaptation Directeur : M. Yves MATILLON
- Département de formation et centre de recherche en Biologie Humaine Directeur : Anne-Marie SCHOTT

SCIENCES ET TECHNOLOGIES

- Faculté des Sciences et Technologies Directeur : M. Fabien DE MARCHI
- UFR de Sciences et Techniques des Activités Physiques et Sportives (STAPS) Directeur : M. Yannick VANPOULLE
- Ecole Polytechnique Universitaire de Lyon (ex ISTIL) Directeur : M. Pascal FOURNIER
- I.U.T. LYON 1 Directeur : M. Christophe VITON
- Institut des Sciences Financières et d'Assurance (ISFA) Directeur : M. Nicolas LEBOISNE
- ESPE Directeur : M. Alain MOUGNIOTTE

ISPB – Faculté de Pharmacie Lyon
LISTE DES DEPARTEMENTS PEDAGOGIQUES

DEPARTEMENT PEDAGOGIQUE DE SCIENCES PHYSICO-CHIMIQUE ET PHARMACIE GALENIQUE

- **CHIMIE ANALYTIQUE, GENERALE, PHYSIQUE ET MINERALE**

Monsieur Raphaël TERREUX (Pr)
Monsieur Pierre TOULHOAT (Pr - PAST)
Madame Julie-Anne CHEMELLE (MCU)
Monsieur Lars-Petter JORDHEIM (MCU-HDR)
Madame Christelle MACHON (AHU)

- **PHARMACIE GALENIQUE -COSMETOLOGIE**

Madame Marie-Alexandrine BOLZINGER (Pr)
Madame Stéphanie BRIANCON (Pr)
Madame Françoise FALSON (Pr)
Monsieur Hatem FESSI (Pr)
Monsieur Fabrice PIROT (PU - PH)
Monsieur Eyad AL MOUAZEN (MCU)
Madame Sandrine BOURGEOIS (MCU)
Madame Ghania HAMD-DEGOBERT (MCU-HDR)
Monsieur Plamen KIRILOV (MCU)
Monsieur Damien SALMON (AHU)

- **BIOPHYSIQUE**

Monsieur Richard COHEN (PU – PH)
Madame Laurence HEINRICH (MCU)
Monsieur David KRYZA (MCU – PH - HDR)
Madame Sophie LANCELOT (MCU - PH)
Monsieur Cyril PAILLER-MATTEI (MCU-HDR)
Madame Elise LEVIGOUREUX (AHU)

DEPARTEMENT PEDAGOGIQUE PHARMACEUTIQUE DE SANTE PUBLIQUE

- **DROIT DE LA SANTE**

Monsieur François LOCHER (PU – PH)
Madame Valérie SIRANYAN (MCU - HDR)

- **ECONOMIE DE LA SANTE**

Madame Nora FERDJAOUI MOUMJID (MCU - HDR)

Madame Carole SIANI (MCU – HDR)
Monsieur Hans-Martin SPÄTH (MCU)

- **INFORMATION ET DOCUMENTATION**

Monsieur Pascal BADOR (MCU - HDR)

- **HYGIENE, NUTRITION, HYDROLOGIE ET ENVIRONNEMENT**

Madame Joëlle GOUDABLE (PU – PH)

- **INGENIERIE APPLIQUEE A LA SANTE ET DISPOSITIFS MEDICAUX**

Monsieur Gilles AULAGNER (PU – PH)
Monsieur Daniel HARTMANN (Pr)

- **QUALITOLOGIE – MANAGEMENT DE LA QUALITE**

Madame Alexandra CLAYER-MONTEBAULT (MCU)
Monsieur Vincent GROS (MCU-PAST)
Madame Audrey JANOLY-DUMENIL (MCU-PH)
Madame Pascale PREYNAT (MCU PAST)

- **MATHEMATIQUES – STATISTIQUES**
Madame Claire BARDEL-DANJEAN (MCU-PH)
Madame Marie-Aimée DRONNE (MCU)
Madame Marie-Paule PAULTRE (MCU - HDR)

DEPARTEMENT PEDAGOGIQUE SCIENCES DU MEDICAMENT

- **CHIMIE ORGANIQUE**
Monsieur Pascal NEBOIS (Pr)
Madame Nadia WALCHSHOFER (Pr)
Monsieur Zouhair BOUAZIZ (MCU - HDR)
Madame Christelle MARMINON (MCU)
Madame Sylvie RADIX (MCU -HDR)
Monsieur Luc ROCHEBLAVE (MCU - HDR)
- **CHIMIE THERAPEUTIQUE**
Monsieur Roland BARRET (Pr)
Monsieur Marc LEBORGNE (Pr)
Monsieur Laurent ETTOUATI (MCU - HDR)
Monsieur Thierry LOMBERGET (MCU - HDR)
Madame Marie-Emmanuelle MILLION (MCU)
- **BOTANIQUE ET PHARMACOGNOSIE**
Madame Marie-Geneviève DIJOUX-FRANCA (Pr)
Madame Marie-Emmanuelle HAY DE BETTIGNIES (MCU)
Madame Isabelle KERZAON (MCU)
Monsieur Serge MICHALET (MCU)
- **PHARMACIE CLINIQUE, PHARMACOCINETIQUE ET EVALUATION DU MEDICAMENT**
Madame Roselyne BOULIEU (PU – PH)
Madame Magali BOLON-LARGER (MCU - PH)
Madame Christelle CHAUDRAY-MOUCHOUX (MCU-PH)
Madame Céline PRUNET-SPANO (MCU)
Madame Catherine RIOUFOL (MCU- PH-HDR)

DEPARTEMENT PEDAGOGIQUE DE PHARMACOLOGIE, PHYSIOLOGIE ET TOXICOLOGIE

- **TOXICOLOGIE**
Monsieur Jérôme GUITTON (PU – PH)
Madame Léa PAYEN (PU-PH)
Monsieur Bruno FOUILLET (MCU)
Monsieur Sylvain GOUTELLE (MCU-PH)
- **PHYSIOLOGIE**
Monsieur Christian BARRES (Pr)
Monsieur Daniel BENZONI (Pr)
Madame Kiao Ling LIU (MCU)
Monsieur Ming LO (MCU - HDR)
- **PHARMACOLOGIE**
Monsieur Michel TOD (PU – PH)
Monsieur Luc ZIMMER (PU – PH)
Monsieur Roger BESANCON (MCU)
Monsieur Laurent BOURGUIGNON (MCU-PH)

Madame Evelyne CHANUT (MCU)
Monsieur Nicola KUCZEWSKI (MCU)
Madame Dominique MARCEL CHATELAIN (MCU-HDR)

- **COMMUNICATION**
Monsieur Ronald GUILLOUX (MCU)
- **ENSEIGNANTS ASSOCIES TEMPORAIRES**
Monsieur Olivier CATALA (Pr-PAST)
Madame Corinne FEUTRIER (MCU-PAST)
Madame Mélanie THUDEROZ (MCU-PAST)

DEPARTEMENT PEDAGOGIQUE DES SCIENCES BIOMEDICALES A

- **IMMUNOLOGIE**
Monsieur Jacques BIENVENU (PU – PH)
Monsieur Guillaume MONNERET (PU-PH)
Madame Cécile BALTER-VEYSSEYRE (MCU - HDR)
Monsieur Sébastien VIEL (AHU)
- **HEMATOLOGIE ET CYTOLOGIE**
Madame Christine VINCIGUERRA (PU - PH)
Madame Brigitte DURAND (MCU - PH)
Monsieur Yohann JOURDY (AHU)
- **MICROBIOLOGIE ET MYCOLOGIE FONDAMENTALE ET APPLIQUEE AUX BIOTECHNOLOGIE INDUSTRIELLES**
Monsieur Patrick BOIRON (Pr)
Monsieur Jean FRENEY (PU – PH)
Monsieur Frédéric LAURENT (PU-PH-HDR)
Madame Florence MORFIN (PU – PH)
Monsieur Didier BLAHA (MCU)
Madame Ghislaine DESCOURS (MCU-PH)
Madame Anne DOLEANS JORDHEIM (MCU-PH)
Madame Emilie FROBERT (MCU - PH)
Madame Véronica RODRIGUEZ-NAVA (MCU-HDR)
- **PARASITOLOGIE, MYCOLOGIE MEDICALE**
Monsieur Philippe LAWTON (Pr)
Madame Nathalie ALLIOLI (MCU)
Madame Samira AZZOUZ-MAACHE (MCU - HDR)

DEPARTEMENT PEDAGOGIQUE DES SCIENCES BIOMEDICALES B

- **BIOCHIMIE – BIOLOGIE MOLECULAIRE - BIOTECHNOLOGIE**
Madame Pascale COHEN (Pr)
Monsieur Alain PUISIEUX (PU - PH)
Madame Emilie BLOND (MCU-PH)
Monsieur Karim CHIKH (MCU - PH)
Madame Carole FERRARO-PEYRET (MCU - PH-HDR)
Monsieur Boyan GRIGOROV (MCU)
Monsieur Hubert LINCET (MCU-HDR)
Monsieur Olivier MEURETTE (MCU)
Madame Caroline MOYRET-LALLE (MCU – HDR)
Madame Angélique MULARONI (MCU)
Madame Stéphanie SENTIS (MCU)
Monsieur Anthony FOURIER (AHU)

- **BIOLOGIE CELLULAIRE**

Madame Bénédicte COUPAT-GOUTALAND (MCU)

Monsieur Michel PELANDAKIS (MCU - HDR)

- **INSTITUT DE PHARMACIE INDUSTRIELLE DE LYON**

Madame Marie-Alexandrine BOLZINGER (Pr)

Monsieur Daniel HARTMANN (Pr)

Monsieur Philippe LAWTON (Pr)

Madame Sandrine BOURGEOIS (MCU)

Madame Marie-Emmanuelle MILLION (MCU)

Madame Alexandra MONTEBAULT (MCU)

Madame Angélique MULARONI (MCU)

Madame Valérie VOIRON (MCU - PAST)

- **Assistants hospitalo-universitaires sur plusieurs départements pédagogiques**

Madame Florence RANCHON

- **Attachés Temporaires d'Enseignement et de Recherche (ATER)**

Madame Charlotte BOUARD (86^{ème} section)

Madame Laure-Estelle CASSAGNES(85^{ème} section)

Monsieur Karim MILADI (85^{ème} section)

Madame Laurence PAGES (87^{ème} section)

Pr : Professeur

PU-PH : Professeur des Universités, Praticien Hospitalier

MCU : Maître de Conférences des Universités

MCU-PH : Maître de Conférences des Universités, Praticien Hospitalier

HDR : Habilitation à Diriger des Recherches

AHU : Assistant Hospitalier Universitaire

PAST : Personnel Associé Temps Partiel

REMERCIEMENTS

Aux Membres du Jury

À Madame le Professeur Christine Vinciguerra,

Vous me faites le très grand honneur de présider mon jury de thèse. Vous faites partie des enseignants qui transmettent leur passion naturellement, pour le plus grand plaisir des étudiants qui n'attendent que ça. Merci pour vos efforts déployés à préparer les étudiants de pharmacie au concours de l'internat. Votre optimisme, votre sagesse et votre charisme nous offrent de précieux encouragements lors de cette année difficile. Votre aptitude à réaliser des clips vidéo de très grande qualité montre vos multiples talents. Que ce travail soit l'expression de ma sincère reconnaissance et de mon profond respect.

À Monsieur le Docteur Stéphane Pinson,

Merci de m'avoir proposé ce vaste et passionnant projet NGS qui me permet de conjuguer mes impératifs de master 2 et de faire un grand saut dans ce qui est dès à présent, le futur de la génétique. Merci de votre patience et de votre compréhension à mon égard. Merci aussi pour l'encadrement de ce travail et les corrections de mes « envolées lyriques » à travers vos relectures. J'espère que mon travail sera à la hauteur de vos attentes. Soyez assuré de mon respect et de ma plus grande reconnaissance.

À Madame le Docteur Claire Bardel,

Je suis très honoré de vous avoir dans mon jury, à la fin d'un travail qui n'aurait pas pu commencer sans votre investissement dans le master 2 qui m'a permis d'intégrer les notions essentielles au NGS. J'espère que vous pourrez le développer encore plus, car il y a un vrai besoin de formation dans les nouvelles techniques de séquençage haut-débit. Merci pour votre gentillesse, votre patience et pour votre disponibilité. Je garde un très bon souvenir des cours que vous dirigiez pendant mes années de pharmacie. Trouvez dans ce travail la marque de mon respect et de ma gratitude.

À Monsieur le Professeur Alain Calender,

Vous me faites l'honneur de juger ce travail et je vous en remercie chaleureusement. Je vous suis tout particulièrement reconnaissant de m'avoir accepté pour mon premier stage de spécialisation dans votre service et de m'avoir fait confiance par la suite. Merci d'avoir œuvré et accepté de donner une suite à ce travail. Je suis honoré par votre offre d'assistantat et me réjouis de continuer d'apprendre et travailler dans un domaine qui me passionne. J'espère ne pas décevoir, à travers ce travail, votre confiance placée en moi.

À toute l'équipe du laboratoire de Génétique de l'hôpital Edouard Herriot,

Pour leur gentillesse, leur accueil et leur soutien, merci. Soyez assurés que je continuerai de vous apporter le café autant que possible, et des tartes aux pralines. Que non, je ne changerai (accent allemand) pas 1 mois après ma prise de fonction et que du coup, nous n'aurons pas d'obligation à nous vouvoyer. Je ne me fais pas trop de soucis à ce que la bonne humeur prime toujours.

Aux collègues du CLB, merci de m'accueillir dans votre équipe. J'espère réaliser avec vous de belles choses et que le NGS deviendra une routine d'une facilité déconcertante dans les prochaines semaines. Chantal promis je vais essayer de ne plus perdre la clé du café...

Un remerciement un peu plus marqué à :

Anne-Lyse, merci pour ta patience, ta gentillesse et ta bienveillance à mon égard. Sans toi, ce travail n'aurait pas été le même et ton aide m'a été précieuse. Saches que tu as souvent de très bonnes idées ! Merci (ou pas) pour les pauses café, ta bonne humeur, tes propos francs, tes ragots, et ton travail sur nextGENe (lol).

Cécile, merci pour m'avoir formé lors de mon premier semestre en génétique. Tu m'as fait découvrir le gène *NF1* et toute sa complexité. Merci d'avoir été patiente avec moi, et de m'avoir toujours écouté et aidé du mieux possible. Merci pour ta participation aux vérifications Sanger de ce projet. Après avoir été mon professeur, j'espère te transmettre à mon tour les joies du NGS. Je ne me fais pas de soucis à ce que ta nature curieuse et pragmatique y trouve un intérêt et un certain plaisir. Merci pour me rappeler de temps à autre et à travers tes histoires, mes quelques mois de passage sur les terres grenobloises. L'air des montagnes me manque parfois et je te souhaite de très longues et belles randonnées dans les années à venir.

Manue, merci de m'avoir passé mes puces et fait tout le « sale travail » à ma place. Merci pour prendre en charge l'intégralité du processus, même quand j'en oublie moi-même une partie ! Ton travail m'a été d'une grande utilité et je t'en remercie chaleureusement. Je te souhaite beaucoup de bonheur avec ta petite famille.

Nadia, merci d'avoir remué ciel et terre pour avoir un assistant ! Merci pour m'avoir laissé du temps afin de rédiger ce travail. Merci pour ta bienveillance et pour tes efforts à m'intégrer au mieux à l'équipe CLB. C'est parti pour 2 ans de binôme Chef ! Et oui maintenant je vais signer tes bilans et faire tourner les séries sur NextGENe. Merci de me faire confiance, en espérant ne pas te décevoir.

À d'autres personnes rencontrées lors de mes études,

À Monsieur le Professeur D. Sanlaville,

Merci de m'avoir accepté dans votre laboratoire pendant un semestre d'internat. Je vous remercie pour votre bienveillance et la qualité de vos connaissances ainsi que votre passion à les transmettre. Soyez assuré de mon respect et de ma profonde reconnaissance.

À Monsieur le Docteur P. Joly,

Merci Philippe de m'avoir accepté pour ma 5AHU. Tu es le premier à m'avoir appris les bases de la génétique, et c'est un peu grâce à toi que j'en suis maintenant ici. Merci pour tes efforts déployés, ta disponibilité ainsi que ta patience à mon égard. Désolé d'avoir interverti les tubes de Taq et d'avoir jeté le plein et gardé le vide ! Je te suis sincèrement reconnaissant.

A travers leurs qualités, je tenais à remercier pour diverses raisons,

Le Dr. Karim Chikh, pour son aura de professeur et la qualité de ses enseignements et aussi pour avoir donné confiance à son sosie, mon ami S.Meniri.

Le Pr. Cohen pour sa rigueur scientifique et ses imposantes connaissances qui forcent le respect.

Benjamin Ribba pour m'avoir fait lâcher la filière ingénieur après un cours de trop sur la thermodynamique, un grand merci.

À ma Famille

À mes parents :

Un grand merci pour votre soutien sans faille et votre amour que vous me donnez chaque jour (sans que je vienne toujours le chercher). Ce travail vous est dédié. Merci à boule et boulette pour les efforts consentis et que l'on reconnait une fois adulte. Pour votre éducation sans faille (Ben a bien réussi à s'en sortir haha), votre dévotion à nous emmener partout pour le basket ou toute autre activité, tes efforts maman à suivre nos études et s'occuper des papiers (sauf pour l'ouverture de mon courrier), merci. Votre amour sera toujours une force pour moi, et malgré toutes les petites piques, sachez que je suis très fier de vous et d'être votre fils. Vous avez bien fait de passer devant cette poubelle ! Notre famille est unie et je suis ravi de vous voir maintenant grands-parents. Puisse votre future retraite être joyeuse et pleine de vie.

À mon frère :

Merci frérot pour avoir mené le chemin vers la vie adulte. Merci pour m'avoir sûrement guidé vers le basket en te voyant jouer. Merci de m'avoir poussé à récolter le plus de feux verts dès mon plus jeune âge, peut être que tout vient de là (pas sûr quand même). Et merci pour nous avoir donné avec Cam les plus beaux neveux et nièce que l'on puisse espérer (en attendant les nôtres). Après une petite période difficile, je te souhaite de trouver un nouvel équilibre avec Gaëlle, d'être heureux et de développer ta sagesse. Je me retrouve souvent en toi, ce qui n'est pas toujours pour me plaire, et t'aime fort. La bise aux gitans de Saint-Tropez pour moi.

À ma soeur :

Merci caca (désolé) pour avoir été la meilleure des grandes sœurs. Malgré nos 11 mois d'écart, merci d'avoir joué le rôle de traductrice quand nous étions petits. Merci d'avoir été la sœur jolie et intelligente que tu es. Nous sommes désolés avec Ben de n'avoir jamais été sérieux à tes galas de G.R.S, comme je suis sûr que tu peux nous comprendre aussi ! Merci par contre de n'avoir pas été rancunière et de nous avoir souvent accompagné et toujours encouragé (go sanglier !). Je te souhaite beaucoup de bonheur avec le Beau-f, vos enfants seront de toute beauté et je ne me fais aucun souci pour que tu les élèves de la meilleure des façons. En espérant que vous n'allez pas trop vous enterrez dans le fin fond de l'Ain ou encore pire, car maman ne serait pas contente. Je t'aime fort sœurette.

À ma grand-mère:

Merci Mamie de vous être aussi bien occupé de nous tous avec Papi. Notre jeunesse est fondée en grande partie sur nos vacances à Vénéjan. Merci d'œuvrer comme jamais nous pourrions le faire à entretenir la maison et tout ce qui va avec. Merci de nous recevoir comme des rois quand nous ne voyons pas toujours les efforts que cela te coûte. Je te souhaite de

profiter au maximum de ta retraite, que les genoux arrêteront un peu leur bazar, et que le souvenir de Papi soit moins douloureux. Merci de nous avoir inculqué les valeurs essentielles à notre éducation. Et merci pour ton appartenance à la Provence qui est ancrée au fond de nous. Je suis fière d'avoir une grand-mère comme toi que beaucoup de nos camarades envient.

Au beau-frère Patrick (non je déconne c'est Patrice):

Désolé pour la blague d'en dessus. Merci à toi mon le beau-frère qui rend heureuse ma sœur depuis déjà plusieurs années. Merci d'avoir apporté un peu de nouveauté et de Lusitanie dans notre famille. Merci pour le séjour sur Saint-Trop, c'était de la balle. Quand on disait qu'il fallait refaire ça chaque année, je pense que la plupart des acteurs du weekend ne jouait pas du pipeau ! Merci pour ta bienveillance et le très joli mariage que tu as réalisé avec ma sœur. En attendant de boire le café chez vous hein ! (merci bien). Plein de bisous et un jour je vous souhaite avec cam de racheter le domaine viticole qui a toujours été le vôtre.

À la belle-sœur Gaëlle :

Merci Gaëlle de nous avoir donné la première petite Buisson de la famille. Nous en sommes très heureux. Merci de t'occuper de mon frère et merci pour lui de ta patience et de ta bienveillance. Sache que nous sommes heureux de te compter parmi nous et que ta place dans la famille est toute faite. Tu auras très vite ta tarte à la praline promis ! Des bisous et félicitations pour votre ténacité dans l'arrêt des clopinettes et du coca.

Aux trois petits bouts de choux :

Merci à Alphonse, Amaury, et Alicia d'être de merveilleux bébés et enfants. Quand vous lirez ceci les pages de couverture auront un peu vieilli, mais sachez que Tonton et Dino vous aime fort et que vous faites son bonheur à chaque fois que l'on se voit. Profitez bien de votre jeunesse dans le Béarn, et svp, ne vous mettez pas au foot ou au rugby (il reste le basket de libre du coup). Tonton sera toujours là pour vous et j'ai hâte de vous revoir. Des bisous les fifous.

À mes Amis

A Bibou,

Jolie petite salade de fruits, rencontrée pendant un tango Corse lors d'un petit bal perdu de pharmacie. Merci d'être une fille merveilleuse et d'avoir marqué ma vie. Si tu étais un sujet d'étude, tu serais incroyablement passionnante mais d'une complexité folle. Nous nous sommes fondés l'un de l'autre. Tu sais que je serais toujours là pour toi. Merci pour nos voyages, nos rires, nos pleurs, et tout ce qui fait que tu me connais mieux que personne. Si ton cœur a toujours été partagé entre la mer de ton « petit pays » (que tu aimes beaucoup) et la terre Lyonnaise de tes parents, ta tête est bien souvent perchée très haut dans les airs, à mon profond désespoir. J'espère que tu trouveras la sérénité sur Madinina et enfin une douceur de vie tant recherchée. Je te souhaite de trouver une « fin de race » de qualité qui te rendra heureuse, et que malgré tout nous garderons ce lien particulier qui nous unit ! En attendant, prend soin de toi, et tienbè raid pas moli ti'mal ! Zoubi

(Merci à Françoise et Jacques de m'avoir si bien accueilli, et de m'avoir fait confiance)

À Roman, R2G, R2, Roms2Gar, Garin, La Garinette et d'autres.

Merci mon pote d'être là depuis la première année. Merci à ton sens de l'amitié, à tes folies, à ton intégration dans tes connaissances multiples (je pense aux mecs du 6), aux weekends à Serez' et à tous nos moments passés. A nos mots inventés. Je pense que l'on se souviendra longtemps de quelques épisodes vécus ensemble et que l'on pourra se remémorer dans quelques années comme des vieux bien dégoutey que cela soit passé si vite.

N'hésites pas à me rappeler ces instants dont tu te souviens souvent mieux que moi. Pêle-mêle, mais je suis sûr que ces quelques références te feront sourire : Merci pour les skis pharma (le lit cassé, Jésus et Lombard, d'avoir pris la porte de la copine de sami, d'avoir écrit au sang que Sami nous avait tuer, ce qui était tout à fait vrai), les bêtises de nos folles soirées à la bonne époque : nos péages, nos emprunts de bouteilles, nos fausses places, le chat « pas-très-en-forme », les AG d'internat, les férias de Bayonne, le cap'tain bloum bloum , les mises sous pression multiples et variées infligées, notre détestation de l'amicale, notre interdiction de ski pharma, ta proposition toujours respectable de boire une vodka contre 100€ en fin de soirée, Nancy Gribelin) et d'autre encore.

Merci aussi pour la formation de notre trio de musique, qui aurait pu faire un carton si sami jouait un peu mieux de la grat. On a bien changé depuis la P1 (surtout ta garde-robe), mais je suis heureux que l'on se retrouve au moins à faire le même métier. Nos chemins vont je l'espère, dans la même direction encore longtemps. Merci de travailler ton shoot et tes mooves dos-au-panier stp afin que je puisse te mettre la misère avec un peu plus d'honneur. Je te souhaite une douce et heureuse relation avec Pauline la lapine. Rien de mieux que de te voir heureux et satisfait sexuellement. Vous formez un joli couple mon gars. Je t'embrasse.

À Sami, Sam2, Jo Coko, El Majico Coko, El chicanos, La cok, J. Cok, Bouboule et d'autres (mais surtout pas Drake, ni un Indien, ou encore Obama).

Je t'ai mis en deuxième mais tu sais bien qu'il n'y a pas de hiérarchie dans mon cœur. Merci mon petit pote d'être aussi là depuis le commencement. Merci à tes parents d'avoir formé un si parfaite personne issue de l'immigration. Les cours du Cha sont bien loin maintenant et je pense qu'on est devenu tous les deux des hommes (la preuve t'as de la barbe maintenant). Merci pour cette magnifique escapade faite au pays des Chicanos. Tu t'es pas moqué de nous sur ce coup-là (bon à part sur Xico un peu quand même), c'était vraiment incroyable. Je ne rigole absolument pas quand je te dis qu'il faut qu'on y retourne (zeubi comme tu aimes bien le dire).

Merci pour le trio de musiciens que nous formions avec R2, dommage que ce dernier soit vraiment trop mauvais chanteur. Merci d'être là quand il faut parler à un mec sensé du groupe. Merci d'accepter les blagues racistes avec le sourire, ça fait plaisir à tout le monde et ça met à l'aise en société. Merci de m'avoir accompagné aux cours d'ingénieur et de nous avoir conseillé de mater Sheitan plutôt que les cours de Benjo la Rib. Plein de bonheur avec Pauline (mince c'est la même phrase que R2 c'est relou). On vous attend comme le premier mariage du groupe les gars. Je te promets de te prendre au squash dès que j'aurai un peu de temps pour me remettre au sport. La bise la bouche pleine de makrout.

À Stéphane, Mon gars sûr, Frérot, Konan, S2K, S2, Le bio d'Abidjan, Couz, Bro et d'autres,

« Bordel, quand on rentre sur Rungins/Comment leur faire confiance ils ont tué ce Fils/Je regarde sur ma gauche, RAS, sur ma droite des chiens du taff tenu en laisse/ Mon Fiston a les idées claires qu'avec des polys verts, y a pas que le fe-ta dans la vie, toto taff à en devenir débile, une pensée pour les taffos disparus comme Sully Sefils». Ba alors poto, t'es dégouté parce que t'es pas thesey ? haha. Merci gros pour tout, tes remix lyriques, ta bonne humeur. On se connaît depuis moins longtemps que les autres zouaves, mais notre relation est forte mon pote. Je n'ai pas toujours été là, donc ton meilleur pote c'était toi ? En tout cas, tu peux m'appeler à 4h du mat je serais là, mais sans pioche franchement.

Allez j'arrête, merci pour entretenir mon côté racailleux avec des gros sons qui tâchent. Je sais que la vie n'a pas toujours été facile pour toi bro, mais t'es un peu en or avec ses convictions (attention avec les ananas quand même). Je te dois une fière chandelle, peut-être plus que toto au final, pour m'avoir laissé faire ce taff en pompant salement tes fiches, surtout celles de bioch (tmtc pourquoi). Je suis fier de t'avoir à mes côtés et tu devenu rapidement un de mes meilleurs potes. Essaie de ne pas partir trop vite au soleil ou à la chasse à l'oseille trop rapidement. Profitons prochainement pour aller au bled, tu m'en parles toujours, j'ai pas peur du couz mecton. Merci pour ta bienveillance et ta bonne humeur ainsi que ta franchise. On se ressemble plus qu'on ne veut bien l'admettre je pense. Je te souhaite bon courage avec ta thèse, au début ça pique mais tkt après c'est cool. Et l'année prochaine franchement on va voir B20 en concert. Je te souhaite aussi plein de bonheur avec Audrey. Paye ton mafé please.

À Arnaud, Ninou, la Nine, Nuggets, Nuggsy nugg, Nonard, Nonado cokel, et autres

Merci Ninou d'être toi-même. Tu es brave, droit, et je sais qu'on peut compter sur toi. Merci pour ces vacances cet été, où j'étais un peu surexcité, mais on s'est bien marré quand même. A refaire vite, quand tu seras revenu de Suisse ? Arrêtes de penser aux millions la nine et reviens par chez nous dans le OL. Merci pour tes connaissances footballistiques et ta sagesse rare. Tes conseils et prévisions font du groupe Nonardo Cunselling and Fils une entreprise à la pointe de l'information. JJ Bourdin n'a qu'à bien se tenir, et les français qui veulent savoir ont tout à y gagner. Beaucoup de bonheur avec la Fof, même si elle a la main un peu lourde sur la moutarde. A très vite pour un footing sur les berges et une pite flette.

À Jean-Baptise, la Jib, JB, M. Président, noob de css, et autres

Merci JB de nous avoir motivé avec les deux autres zozos pour l'internat. Tu nous as montré la voie et ton impressionnant bagou pour la biologie nous a tous renforcé sur nos choix. Merci de t'être battu pour notre métier lors de tes différentes fonctions tout au long de ton internat. Dsl mon pote d'avoir oublié que je te devais une place dans mon jury de thèse... t'aurais pas compris grand-chose de tt manière et t'aurais pas eu le temps de lire ce pavé (enfin sorry quand même). Merci d'être un geek et d'aimer le street art comme moi. Je t'attends sur css quand t'aura passé le grade 2 étoiles. Merci d'être toujours ouvert et disponible comme pote, et d'aimer le son old school de notre jeunesse. Un jour tu me montreras le métro de Rouen, en attendant je te souhaite de retrouver ta Marie aussi vite que possible, et que ton énorme travail dans le privé arrive à payer. A très vite pour un P&T, tu es le seul qui comprenne la beauté de la chose.

À Toto, la Tote, Fils, Fiston, Flonzy, Flonzyflonz, Espuigouin, Boule de poils, M. Pyjama,

Merci Toto d'avoir formé avec S2 notre trio de taff pour l'internat. Merci de nous avoir fait flipper (surtout moi) avec ton travail acharné et si efficace. On ne te félicitera jamais assez d'avoir réussi là ou certaines personnes ne voyaient pas le taff de chien que tu produisais. Je crois que les exos tu les connais encore par cœur. Tu l'as bien mérité cette Bio, qui va te l'enlever maintenant ? Pas merci par contre (carton rose attention) pour ton manque de courage afin de voir les potos et d'avoir gardé ton Pyjama après 22h même lors de l'internat. Maintenant que tout est plus calme, ramène ta moitié et viens nous voir (la normalement je suis grossier). Beaucoup de bonheur avec la Nat, et n'oublie pas de profiter de ton temps libre pour passer le permis de conduire sérieusement. Je t'aime fils et suis fier de toi. Mam's

À Margi, Nabil, François (désolé je vous groupe):

Merci les gars pour être là lors des bons moments, c'est toujours un plaisir de vous voir et de passer du bon temps avec vous. Merci margi d'être le plus gangsta des roumains, d'être un hipster trop swag, de partager tes ptits sons, et d'avoir une jumelle super. Je te souhaite aussi beaucoup de bonheur avec la chérie, et de passer ta thèse de philo-pharmacie rapidement. Attention aux détournements de dons au Musée par contre, la Porsche ça fait louche gros.

Merci Nab2 d'être le Sylvain Mirouf du Maghreb, un vrai clown aux muscles saillants mais avec un gros cœur et une grande tendresse. Merci pour tes gadgets, tes stocks de farces et attrapes ainsi que de conneries.

Merci François d'être toujours de bonne humeur et partant. Courage pour les moments difficiles que tu traverses. Bravo pour le travail que tu fais, tu nous impressionnes tous. Merci aussi pour les contacts toujours sympathiques que tu m'as partagé.

Aux Grenoblois :

À **Charles**, Carles, Carlito. Merci mon poto pour ces 3 semestres et plus, passés à tes cotés sur Gre. Difficile au début, j'ai fini par partir avec un petit pincement au cœur. Merci d'avoir une tête de toxico-chien loup- à la cool, qui a fait qu'on est venu te parler avec Stéphane lors du premier jour à Lyon. Merci de m'avoir fait une place dans la bande de la coloc, et pour les repas et nuits passés la-bas. Merci pour les voyages qu'on a fait et qui étaient tous magnifiques. A quand le prochain mon pote ? Merci d'être un des rares amis que j'ai qui se veut friand de bouquins et d'expos. Même si je pensais que c'était surtout pour te la raconter au début, je dois dire que tu as l'air de vraiment aimer ça, ou en tout cas c'est bien fait. Merci pour les parties de squash que tu me laissais gagner, 9 fois sur 10 c'est quand même de l'altruisme de haut level. Merci à ta famille, les frérots, les cousines, je les ai tous apprécié (même le morback). Je te souhaite plein de bonheur avec Lette-PO. Tu sais qu'elle est pleinement validée par mes soins. Je te souhaite de rester dans le coin après l'internat, qu'on puisse continuer à se capter et tenter de se faire des orguies.

À **Pierre**, Nazillon, SS, Papi, Pierrot le skin. Merci l'ancien pour ta bonne humeur et ta sympathie à mon égard. Je sais que mes jojos le premier jour t'ont permis de reconnaître que j'étais un vrai. Merci pour tes cds gravés que tu me faisais ou que je te volais. Merci pour le semestre en bactério pdt lequel malgré le travail tout moisi qu'on avait, on s'est quand même bien marré. Merci pour ton grain de folie et ton amour pour les jolies filles. J'espère que tu es heureux d'être sur Lyon, c'est la ville t'as vu. Sois pas dégouté, tu vas rester ici toute ta vie ici ☺.

À **Benoît**, la Bin (mon roux préféré, merci pour ton naturel, ta finesse, et ton style usité qui font ton charme), à **Brune** (merci pour ton humour, ton énergie, et ta sympathie, plein de bonheur avec B.), au **Président** (goss bo le prez ! trop un gangster en plus) et sa petite femme Mélanie (merci de former un couple parfait, plein de bonheur et de réussite sur Gre). A **Paulette**, la perle de Saint-Etienne, merci de m'avoir laissé entraîner ton homme dans de sombres soirées et de m'avoir fait confiance. Ce n'est pas le dernier sur les conneries, mais tu as toujours été d'accord avec nos envies, merci. Merci d'être une oreille attentive et une fille d'une douceur extrême. Je vous souhaite beaucoup de bonheur dans votre joli couple.

À **Sultan**, Tektansul, le roi afghan de la techno. Merci mon gars pour les premiers semestres passés ensemble. Merci pour ta motivation à la muscu. Merci pour les aprem console à la place de l'hémato. Merci pour m'avoir montré comment mettre du yaourt dans ses plats, et de m'avoir appris un peu plus de choses sur l'Afghanistan. Il va falloir qu'on se voit un moment et qu'on discute le bout de gras. Il paraît que ta copine est fraîche et bien loin des ptites idiotes du K, félicitations. Plein de bonnes choses dans ton projet en Belgique.

À **Carine, Maëlle**, merci les filles pour la bonne entente, votre bonne humeur et votre dynamisme. Je vous souhaite beaucoup de bonnes choses en biologie et dans vos vies privées.

À Dédé, Gargou (gargamelle), Carole, Anne-so, et d'autres que j'oublie certainement : merci d'avoir fait de ces quelques semestres sur Gre une magnifique expérience et finalement, un très bon moment. Je conseil à tous futurs internes de s'y rendre au moins 1 semestre.

Aux Personnes rencontrées lors de mes études :

Aux anciens, **Binouz, Rami, Jésus, Bento** et leurs femmes respectives (dsl Jez je sais que ça fait mal à entendre). Merci pour avoir fait partir nos études sur les chapeaux de roues. On a commencé avec une barre très haute dès notre premier ski pharma. On a fait que se surpasser par la suite ☺. Merci pour la déconne H24, pour les saloperies de Jésus, les jeux de mots de Bento, la grande gueule de Rami (jte baise sous le panier gros), et la crédulité fantastique de binouz.

Clément, sacré mecton, un vrai apollon, merci pour ce semestre passé ensemble en cytogénétique. C'était un bon semestre, on s'est bien marré entre tes histoires, les miennes, tes concours de muscu, la Laurence et le séjour dans ta maison l'été. Merci d'être un mec sur qui on peut compter. Je sais que je reste un mystère pour toi des fois, mais on n'est pas si différent que ça au final tu sais. Je te souhaite du bonheur dans ton taff et dans ton couple.

Merci aussi à : **Marc** pour tes soirées et ton envie de faire la fête. **Pauline L, Ariane, Océanne, Anne, Matou** (ma préférée !) : pour ces bons moments passés ensemble.

Aux gens du 6 :

Merci les mecs pour les bons moments et bonnes soirées passés ensemble, je pense à **Wilson, Tarik, Brugièrre, Jonjon, Mathilde, Sophie, Pomplard**, et tous les autres.

Merci à elles :

Aude, merci pour ces moments passés ensemble, spécialement le weekend en Espagne. Désolé de ne pas avoir donné le meilleur de moi-même. Tu es une fille super et je te souhaite plein de bonheur sur Lyon.

Charlotte, je suis heureux de te connaître. Tu es une fille douce et bienveillante. Je suis fier de te compter parmi mes amies. Mon seul regret est de ne pas avoir pu contempler un coucher de soleil de Bandol, rosé à la main. J'espère que tu resteras sur Lyon encore un certain temps.

Priscilla, désolé de ne pas avoir été disponible ces derniers temps. J'apprécie ta capacité d'analyse et nos longues conversations. J'espère que tu te plais dans ton nouvel appart.

Annabelle, pour ta folie, ta décontraction et ta franchise. J'espère passer plus de temps en ta compagnie quand tu rentreras de la Réunion.

Aux Caladois :

Merci à Canard **Mickael**. Je suis désolé mon pote de ne pas avoir été assez présent ces dernières années, la pharma m'ayant accaparé beaucoup de temps. Je te souhaite beaucoup de bonheur avec ta famille. Merci pour avoir rendu le lycée très sympa. Merci pour notre escapade un mercredi aprem en moto. Merci pour la visite de Clun's et du foy's. Et merci d'être un mec du beaujolais, qui me rappelle un peu d'où on vient.

Merci à **Aulnay**. Vieux noob, on se connaît depuis la 5^{ème}, soit bientôt 15 ans. Tout ce temps et tu dois être le seul mec que je connaisse qui n'ait pas changé, dans ta diction, ton style, ou ta vieille mouille (bon il y a bien quelques râteaux en plus mais ça arrive). Ma mère était choquée la première fois qu'elle t'a vu, elle a dû prendre 2 paracétamols à cause d'un mal de crâne, mais c'est pour ça que je t'aime. On était bien avec la bande de mecs de Claude B, la vie était belle et pleine de jeux vidéo. Dommage que ta carrière de PGM n'ait pas marché et que tu sois devenu médecin à la place. Merci pour être toi, pour me piquer les yeux à chaque message que je reçois de ta part, pour ta folie et ton humour. Un jour tu seras le Roi de la Calade j'en suis sûr mon pote ! A toi les cagoles de la rue nat' ou de la Sto. Il va falloir qu'on se refasse des soirées fissa fissa avant qu'une morue ne te mette le grappin dessus. La bise payday du zayzay.

A **Pierre**, Pedrolito. Je sais qu'on se voit pratiquement plus mon pote, mais ces quelques mots pour te faire savoir peut être un jour que je pense à toi et que je n'oublie pas que t'étais un pote en or. Je te souhaite plein de bonnes choses dans le sud.

Et à tous-ceux que j'ai oublié,

Adrien

« Seul l'inconnu épouvante les hommes. Mais, pour
quiconque l'affronte, il n'est déjà plus l'inconnu. »

Antoine de Saint-Exupéry

« C'était compliqué, alors j'ai essayé d'expliquer ce qu'était le Graal,
pour que tout le monde comprennent.
C'était difficile, alors j'ai essayé de rigoler,
pour que personne ne s'ennuie.
J'ai raté mais j'veux pas qu'on dise que j'ai rien foutu,
Parce que c'est pas vrai. »

Alexandre Astier, *Kaamelott*, Livre V.

« C'est pas la rue mais l'être humain qui m'attriste.
Comment leur faire confiance, ils ont tué le Christ. »

Elie Yaffa, *Boulbi*.

SOMMAIRE

| | |
|---|-----------|
| Remerciements..... | 9 |
| Vocabulaire..... | 30 |
| Liste des abréviations..... | 31 |
| Introduction :..... | 34 |
| PARTIE I: REVUE BIBLIOGRAPHIQUE..... | 37 |
| 1. Tumeurs et rasopathies..... | 38 |
| 1.1. La voie de signalisation RAS/MAPK..... | 38 |
| 1.1.1. Le signal primaire activateur..... | 38 |
| 1.1.2. Les protéines RAS et RAF..... | 40 |
| 1.1.3. Les protéines MEK et ERK..... | 41 |
| 1.1.4. Effets cellulaires de l'activation de la voie Ras-MAPK..... | 42 |
| 1.2. Mécanismes primordiaux de la cancérogenèse..... | 43 |
| 1.3. De la mutation à la tumeur..... | 48 |
| 1.4. Les Rasopathies : un groupe hétérogène de pathologies..... | 50 |
| 1.5. les différents syndromes composant les RASopathies..... | 50 |
| 2. La neurofibromatose, une « rasopathie » complexe..... | 56 |
| 2.1. La neurofibromatose de type 1..... | 56 |
| 2.2. Historique..... | 56 |
| 2.3. Epidémiologie – prévalence..... | 57 |
| 2.4. Présentation Clinique..... | 57 |
| 2.4.1. Critères du NIH..... | 57 |
| 2.4.2. Critère du NIH : les manifestations cliniques en détail..... | 59 |
| 2.4.3. Les autres atteintes cliniques :..... | 66 |
| 2.5. Diagnostic..... | 69 |
| 2.5.1. Arbre décisionnel..... | 70 |
| 2.5.2. Les formes mosaïques, un diagnostic particulier..... | 73 |
| 2.5.3. Le diagnostic génétique..... | 74 |
| 2.6. Pronostic et suivi..... | 74 |
| 2.7. Conseil génétique..... | 77 |
| 2.8. Prise en charge..... | 78 |
| 3. Le syndrome de Legius, une maladie de découverte récente..... | 80 |
| 3.1. Introduction/ historique..... | 80 |

| | | |
|---------|---|-----|
| 3.2. | Clinique..... | 80 |
| 3.3. | Incidence / pénétrance | 83 |
| 3.4. | Critères diagnostiques..... | 83 |
| 3.5. | Diagnostic génétique..... | 85 |
| 3.6. | Suivi et traitement..... | 86 |
| 3.7. | Conseil génétique | 86 |
| 4. | Etudes Moléculaires des gènes <i>NF1</i> et <i>SPRED1</i> | 88 |
| 4.1. | Neurofibromatose de type 1 : du gène à la protéine..... | 88 |
| 4.1.1. | Le gène <i>NF1</i> | 88 |
| 4.1.2. | Clonage positionnel..... | 88 |
| 4.1.3. | Structure..... | 88 |
| 4.1.4. | les pseudogènes associés au gène <i>NF1</i> | 91 |
| 4.1.5. | ARN messenger..... | 93 |
| 4.1.6. | Les transcrits d'épissage..... | 93 |
| 4.1.7. | La protéine Neurofibromine..... | 95 |
| 4.1.8. | Altérations et mutations du gène <i>NF1</i> | 101 |
| 4.1.9. | Relation génotype-phénotype..... | 104 |
| 4.1.10. | Gènes modificateurs..... | 105 |
| 4.1.11. | Détection des altérations du gène <i>NF1</i> | 106 |
| 4.1.12. | Stratégie actuelle de détection au laboratoire de Lyon | 107 |
| 4.2. | Le gène <i>SPRED1</i> , une nécessité d'analyse..... | 112 |
| 4.2.1. | Clonage positionnel..... | 112 |
| 4.2.2. | Structure..... | 113 |
| 4.2.3. | Pseudogène | 113 |
| 4.2.4. | ARN messenger..... | 113 |
| 4.2.5. | Protéine <i>SPRED1</i> | 113 |
| 4.2.6. | Expression..... | 114 |
| 4.2.7. | Activité fonctionnelle | 114 |
| 4.2.8. | Altération de <i>SPRED1</i> | 116 |
| 4.2.9. | Relation génotype-phénotype..... | 117 |
| 4.2.10. | Détection des altérations du gène <i>SPRED1</i> | 118 |
| 4.2.11. | Stratégie actuelle de détection au laboratoire de Lyon | 118 |
| 5. | NGS..... | 121 |
| 5.1. | Introduction..... | 121 |

| | | |
|--|--|-----|
| 5.2. | Un schéma général de fonctionnement..... | 122 |
| 5.3. | les plateformes actuelles..... | 125 |
| 5.3.1. | Les plateformes Illumina..... | 125 |
| 5.3.2. | Les plateformes Life Technologies/ThermoFisher (Ion Torrent)..... | 129 |
| 5.3.3. | Performances spécifiques aux deux technologies : | 135 |
| 5.4. | Séquençage de 3 ^{ème} génération..... | 137 |
| 5.4.1. | Pacific Biosciences : | 137 |
| 5.4.2. | Oxford Nanopore..... | 141 |
| 5.5. | L'analyse Bio-Informatique | 144 |
| 5.5.1. | Etape d'alignement | 146 |
| 5.5.2. | Etape de l'appel de variant (variant calling)..... | 148 |
| 5.5.3. | Les erreurs et biais du NGS..... | 151 |
| 5.5.4. | Notions de base et différents formats de l'analyse bio-informatique..... | 153 |
| PARTIE II. EXPERIENCES ET RESULTATS..... | | 161 |
| I. | Matériels et Méthodes..... | 163 |
| 6. | Fabrication de la Librairie | 163 |
| 6.1. | La technologie Ampliseq™..... | 163 |
| 6.2. | Premier design Ampliseq™..... | 164 |
| 6.3. | Caractéristiques du 1 ^{er} design | 166 |
| 6.4. | Deuxième design Ampliseq™ | 168 |
| 6.5. | Caractéristiques du 2 ^{ème} design | 168 |
| 6.6. | Préparation de la librairie..... | 171 |
| 6.6.1. | Protocole général | 171 |
| 6.6.2. | Extraction et dosage de l'ADN..... | 172 |
| 6.7.1. | Amplification des régions cibles | 174 |
| 6.7.2. | digestion des primers | 175 |
| 6.7.3. | Ligation des codes-barres et des adaptateurs | 177 |
| 6.7.4. | Purification simple..... | 177 |
| 6.7.5. | Quantification par QuBit™ et LabChip® | 179 |
| 6.7.6. | Quantification par qPCR. | 179 |
| 6.8. | Amplification clonale et séquençage..... | 180 |
| 6.9. | Choix de la puce | 181 |
| 7. | Traitement bio-informatique | 182 |
| 7.1. | Le rapport de run..... | 182 |

| | | |
|--------|---|-----|
| 7.1.1. | Indicateurs de qualité du séquençage..... | 182 |
| 7.1.2. | Indicateurs qualité de l'alignement..... | 184 |
| 7.1.3. | Indicateurs qualité des séquences | 185 |
| 7.1.4. | Critères d'acceptation | 189 |
| 7.2. | Analyses du laboratoire..... | 190 |
| 7.2.1. | Plug-in : File Exporter | 190 |
| 7.2.2. | Plug in : Coverage Analysis | 191 |
| 7.2.3. | Plug in : Variant Caller | 192 |
| 7.3. | Traitement bioinformatique des fichiers VCF obtenus | 196 |
| 7.3.1. | Parcours du fichier VCF <i>Base à base</i> | 196 |
| 7.3.2. | Parcours du fichier VCF <i>Variant</i> | 196 |
| 7.4. | Finalités et analyse du run par le biologiste..... | 198 |
| 7.5. | Stockage | 199 |
| 8. | Echantillons séquencés | 201 |
| 8.1. | Premier design..... | 201 |
| 8.2. | Second Design | 202 |
| II. | Résultats | 204 |
| 1. | Résultats du premier design Ampliseq™ | 204 |
| 1.1. | Comparaison des trois premières séries | 204 |
| 1.1.1. | Performances générales du séquençage..... | 204 |
| 1.1.2. | Analyse de la couverture | 206 |
| 1.1.3. | Analyse de la profondeur | 208 |
| 1.1.4. | Analyse de la détection des variants | 213 |
| 1.2. | Comparaison série 3 et série 4 | 226 |
| 1.3. | Résultat du protocole modifié..... | 227 |
| 2. | Résultats du Second design Ampliseq™ | 229 |
| 2.1. | Comparaison des séries..... | 229 |
| 2.1.1. | performances générales du sequençage..... | 229 |
| 2.1.2. | Analyse de la couverture et des profondeurs | 230 |
| 2.1.3. | Analyse de la détection des variants | 234 |
| 2.2. | Résultats des duplicats de la série 4..... | 248 |
| III. | Discussion | 250 |
| 1.1.1. | la technologie Ampliseq™. | 250 |
| 1.1.2. | La fabrication de librairie..... | 252 |

| | |
|--|-----|
| 1.1.3. L'analyse des données..... | 254 |
| 1.2. Mise en production de la méthode..... | 257 |
| IV. CONCLUSIONS | 259 |
| Annexes | 261 |
| Bibliographie..... | 288 |

INDEX DES FIGURES

| | |
|---|-----|
| Figure 1. Famille des récepteurs de l'EGF (A), activation et vue d'ensemble des voies de signalisation intracellulaire induites (B). | 39 |
| Figure 2. Les quatre isoformes des protéines Ras..... | 40 |
| Figure 3. "The hallmarks of cancer", caractéristiques fondamentales des cellules cancéreuses. | 45 |
| Figure 4. Les nouveaux « hallmarks », caractéristiques fondamentales des cellules cancéreuses. | 46 |
| Figure 5. Effecteurs de la voie Ras-MAPK et leurs implications tumorales..... | 49 |
| Figure 6. La voie de signalisation Ras-MAPK et les différentes RASopathies associées..... | 51 |
| Figure 7. Tâches café au lait de grande taille ³⁰ | 59 |
| Figure 8 Lentigines axillaires..... | 60 |
| Figure 9 Neurofibromes cutanés multiples ³⁰ | 61 |
| Figure 10 Hamartomes iriens ou nodules de Lisch ³² (flèche orange). | 62 |
| Figure 11. Gliome du nerf optique droit ³⁰ (flèche orange)..... | 63 |
| Figure 12. Pseudoarthrose du tibia congénitale ³⁰ (flèches orange)..... | 65 |
| Figure 13. Arbre décisionnel de diagnostic de la NF1..... | 70 |
| Figure 14. Schéma du gène NF1 et de la protéine neurofibromine | 90 |
| Figure 15. Mécanismes de création de pseudogenes..... | 91 |
| Figure 16. Rôle de la neurofibromine dans la voie de signalisation Ras. | 96 |
| Figure 17. Fréquences des altérations génétiques du gène <i>NF1</i> dans les néoplasies humaines.. | 99 |
| Figure 18. Répartition des mutations "Shorts nucléotides Variations (SNV)" du gène <i>NF1</i> | 102 |
| Figure 19. Schéma de la région du gène <i>NF1</i> et les différents type de délétions récurrentes.... | 103 |
| Figure 20. Stratégie du laboratoire pour l'analyse du gène <i>NF1</i> | 111 |
| Figure 21. Variants de <i>SPRED1</i> répertoriés..... | 116 |
| Figure 22. Variants protéiques de <i>SPRED1</i> | 116 |
| Figure 23. Stratégie du laboratoire pour l'analyse du gène <i>SPRED1</i> | 119 |
| Figure 24. Schéma du gène <i>SPRED1</i> et ses altérations génétiques..... | 120 |
| Figure 25. Protocole commun au séquençage de 2 ^{ème} génération. | 122 |
| Figure 26. Séquence des adaptateurs de plusieurs plateformes. | 123 |
| Figure 27. Séquenceurs de 2 ^{ème} génération d'Illumina en 2016 (hors MiniSeq). | 125 |
| Figure 28. Technologie Illumina: amplification clonale et séquençage | 128 |
| Figure 29. Séquenceurs de 2 ^{ème} génération de Life Technologies en 2016..... | 129 |
| Figure 30. Amplification clonale par PCR en émulsion..... | 131 |
| Figure 31. Enrichissement des ISPs..... | 132 |
| Figure 32. Principe de dosage des ISPs. | 133 |
| Figure 33. Schéma conceptuel du séquençage Ion torrent..... | 133 |
| Figure 34. Etude sur plateforme Ion Torrent: précision moyenne des variants appelés en fonction de la longueur de l'homopolymère)..... | 135 |
| Figure 35. Séquenceur PacBio RS II. | 137 |
| Figure 36. SMRTbell de Pacific Biosciences..... | 138 |
| Figure 37. Technologie de la plateforme Pacific Bioscience de 3 ^{ème} génération. | 139 |
| Figure 38. MinIon de Oxford Nanopore. | 141 |
| Figure 39. Technologie de la plateforme Oxford Nanopore de 3 ^{ème} génération..... | 142 |
| Figure 40. Influence de la force appliquée sur la résolution du séquençage par nanopore..... | 142 |
| Figure 41. Principales étapes de l'analyse bio-informatique | 145 |

| | |
|--|-----|
| Figure 42. Schéma de l'alignement sur un génome de référence..... | 146 |
| Figure 43. Exemples de logiciels académiques pour l'alignement, la visualisation, l'appel de variant et leur interprétation..... | 147 |
| Figure 44. Profondeur minimale théorique..... | 149 |
| Figure 45. Workflow du Variant Caller (plateforme Ion torrent)..... | 151 |
| Figure 46. Erreur de PCR propagée dans les duplicats et résultat après leur élimination..... | 152 |
| Figure 47. Exemple de réalignement d'une indel..... | 152 |
| Figure 48. Description du format FASTQ..... | 153 |
| Figure 49. Correspondances du score Phred..... | 154 |
| Figure 50. Score qualité (Phred) par base (outil FASTQC)..... | 154 |
| Figure 51. Format SAM/BAM..... | 156 |
| Figure 52. Format VCF, et nomenclature des différents types de variants..... | 157 |
| Figure 53. Format BED..... | 158 |
| Figure 54. Notions de couverture et profondeur..... | 160 |
| Figure 55. Premier design choisi avec l'outil <i>Ampliseq™ Designer</i> | 165 |
| Figure 56. Visualisation des défauts de couverture (<i>gaps</i>) du design 1 sur IGV..... | 166 |
| Figure 57. Répartition de la taille des inserts du design 1..... | 167 |
| Figure 58. Répartition de la taille des inserts du design 2..... | 169 |
| Figure 59. Protocole de préparation de librairie par la technologie <i>Ampliseq™</i> | 171 |
| Figure 60. Protocole standard de fabrication de banque (pool final)..... | 176 |
| Figure 61. Protocole modifié de fabrication de banque (pool après amplification des cibles) selon la méthode de quantification..... | 176 |
| Figure 62. L'étape de ligation..... | 177 |
| Figure 63. Exemple de rapport de run d'un séquençage (indicateur qualité de séquençage)..... | 183 |
| Figure 64. Exemple de rapport de run d'un séquençage (indicateur qualité de l'alignement)..... | 184 |
| Figure 65. <i>Per base sequence quality</i> de bonne qualité d'un séquençage 400pb..... | 185 |
| Figure 66. Distribution de la qualité moyenne des lectures..... | 186 |
| Figure 67. Distribution moyenne des bases à travers les lectures..... | 186 |
| Figure 68. Distribution taux de GC par base (gauche) et par lectures (droite)..... | 187 |
| Figure 69. Distribution des longueurs des lectures..... | 187 |
| Figure 70. Représentation d'amplicons non chevauchants sur IGV (en rouge)..... | 191 |
| Figure 71. Schéma simplifié de l'analyse bioinformatique du Torrent Server par notre équipe..... | 195 |
| Figure 72. Schéma simplifié du traitement bio-informatique à la validation et stockage d'un run NGS..... | 200 |
| Figure 73. Les séries de validation du design 1..... | 201 |
| Figure 74. Les séries de validation du design 2..... | 202 |
| Figure 75. Caractéristiques et performances générales des séries : 1, 2 et 3 (1er design)..... | 205 |
| Figure 76. Profondeur moyenne des séries 1, 2 et 3 du 1er design..... | 206 |
| Figure 77. Taux de couverture des bases d'intérêt à une profondeur donnée (Bed CDS +/- 20)..... | 207 |
| Figure 78. Pourcentage de lectures alignées hors région d'intérêt..... | 208 |
| Figure 79. Nombre d'exons par série avec minimum une base à un seuil de profondeur inférieur de 100X ou 200X..... | 209 |
| Figure 80. Compilation du nombre de patients ayant au moins 1 base <200 dans une région du gène NF1..... | 211 |

| | |
|---|-----|
| Figure 81. Compilation du nombre de patients ayant au moins 1 base <100 dans une région du gène NF1..... | 212 |
| Figure 82. Distribution du type de variants retrouvés par les 3 premières séries. | 213 |
| Figure 83. Représentation de la qualité en fonction de la profondeur pour les variants VP. En rouge, la prévision de la qualité, selon l'analyse de la régression | 215 |
| Figure 84. Les variants FP de l'exon 4c du gène <i>NF1</i> | 222 |
| Figure 85. Variants faux négatifs des 3 séries avec profondeurs faibles..... | 223 |
| Figure 86. Représentation de l'exon 13, avec la répétition de C , le SNP adjacent et les amplicons le couvrant. Visualisation à partir d'Alamut™ avec un des deux patients porteurs du FN..... | 224 |
| Figure 87. Profondeur des lectures de différents amplicons de la plex B..... | 226 |
| Figure 88. Résultats de l'analyse de profondeur pour les 8 premiers échantillons de la série 4. | 227 |
| Figure 89. Concentration finale des banques (pM) de la série 3. Les flèches orange indiquent les patients avec une librairie de faible concentration. | 228 |
| Figure 90. Nombre de patients ayant au moins 1 base <100 dans une région du gène NF1 ou SPRED1 (+/- 20 pb). | 232 |
| Figure 91. Nombre de régions moyenne avec une base <100X de profondeur par patient..... | 233 |
| Figure 92. Distribution du type de variants retrouvés par les 4 séries du design 2..... | 234 |
| Figure 93. Distribution du type de variants délétères retrouvés par les 3 premières séries..... | 236 |
| Figure 94. Nombre de FP par exons (vérifiés et extrapolés) sur l'ensemble des 4 séries..... | 238 |
| Figure 95. Faux positif c.670delG du gène NF1 dû à un défaut de couverture (IGV). | 242 |
| Figure 96. Visualisation de 3 échantillons porteurs d'une duplication dans l'exon 13 du gène NF1..... | 244 |
| Figure 97. Profondeur des variants en fonction de l'hétérozygotie (VP et FP appelés comme hétérozygotes). | 246 |
| Figure 98. Distribution de la fréquence allélique des variants VP et FP dans les 4 séries..... | 246 |
| Figure 99. Biais de brin en fonction de l'hétérozygotie chez tous les variants VP et FP (flèches noirs : les FP vérifiés en Sanger)..... | 247 |
| Figure 100. Comparaison du nombre de reads selon l'ajout ou non d'amorces du design 1. | 249 |

INDEX DES TABLEAUX

| | |
|--|-----|
| Tableau 1. Les principales pathologies regroupées sous le terme : Rasopathies..... | 55 |
| Tableau 2. Critères diagnostic du NIH (1998)..... | 58 |
| Tableau 3. Les différents tumeurs les plus fréquentes et leur risque associé. | 69 |
| Tableau 4. Fréquence des différents critères diagnostiques du NIH et âge moyen d'apparition..... | 69 |
| Tableau 5. Diagnostic différentiel de la NF1. | 72 |
| Tableau 6. Complications justifiant le suivi des malades NF1 en fonction de l'âge..... | 74 |
| Tableau 7. Critères d'évaluation initiale et de suivi d'un patient NF1 ²⁹ | 76 |
| Tableau 8. Différentes manifestations du syndrome de Légius : nombre de personnes par études chez qui les signes associés sont retrouvés, et leurs fréquences relatives. | 82 |
| Tableau 9. critères du NIH pour la NF1 et confusion possible avec le syndrome de légius..... | 85 |
| Tableau 10. Exon alternatifs du gène NF1. | 93 |
| Tableau 11. Comparaison de deux technologies de séquençage de 2ème génération et une de 3ème génération..... | 136 |
| Tableau 12. Caractéristiques du design 1..... | 166 |
| Tableau 13. Caractéristiques du design 2. | 168 |
| Tableau 14. Caractéristiques des chevauchements du 2ème design..... | 169 |
| Tableau 15. Chevauchement inférieurs à 20 bases..... | 170 |
| Tableau 16. Mix PCR de l'étape d'amplification des cibles (protocole standard et modifié). | 174 |
| Tableau 17. Programme d'amplification PCR de l'étape d'amplification des cibles. | 175 |
| Tableau 18. Programme de l'étape de digestion des amorces. | 175 |
| Tableau 19. Programme de l'étape de ligation..... | 177 |
| Tableau 20. Programme de l'étape d'amplification de la librairie barre-codée. | 178 |
| Tableau 21. Les puces disponibles pour le PGM™. | 181 |
| Tableau 22. Critères qualité du séquençage et de l'alignement..... | 189 |
| Tableau 23. Comparaison des différents paramètres utilisés pour le Variant Caller. | 193 |
| Tableau 24. Les différents rajouts de couples d'amorces testés dans la série 4 du design 2. | 203 |
| Tableau 25. Différents critères des variants VP. | 214 |
| Tableau 26. Variants vérifiés en Sanger (VP). En jaune les 10 SNPs les plus fréquents..... | 216 |
| Tableau 27. Caractéristiques des variants délétères trouvés dans les 3 séries. | 217 |
| Tableau 28. Variants délétères retrouvés dans les 3 premières séries..... | 218 |
| Tableau 29. Caractéristiques des variants faux positifs trouvés dans les 3 séries et vérifiés en Sanger. | 219 |
| Tableau 30. Variants faux positifs retrouvés dans les 3 premières séries et vérifiés en Sanger. | 220 |
| Tableau 31. Motifs proches des variants FP. | 221 |
| Tableau 32. Performances générales des séries : 1, 2,3 et 4 (2 nd design)..... | 229 |
| Tableau 33. Analyse des profondeurs par série..... | 230 |
| Tableau 34. Différents critères des variants VP du design 2. | 235 |
| Tableau 35. Caractéristiques des variants délétères trouvés dans les 4 séries. | 236 |
| Tableau 36. Caractéristiques des variants faux positifs vérifiés en Sanger dans les 4 séries du design 2. | 237 |
| Tableau 37. Variants délétères du design 2. | 241 |
| Tableau 38. Calcul de sensibilité et VPP pour le design 2. | 245 |

Vocabulaire

- Puce ou *flowcell / chip*
- Adaptateurs ou *adapters*
- Alignement de séquence multiple ou *multiple sequence alignment (MSA)*
- Amplification clonale ou *clonal amplification*
- Billes ou *beads*
- Code-barre ou *barcode*
- Couverture ou *coverage*
- Epingle à cheveux ou *hairpin*
- Exome ou *whole-exome sequencing (WES)*
- Génomes complets ou *whole-genome sequencing (WGS)*
- Micropuits ou *microwells*
- PCR avec amplification par pont ou *bridge-PCR*
- PCR en émulsion ou *emulsion-PCR*
- Séquençage haut-débit ou *high-throughput sequencing (HTS)*
- Séquençage par synthèse ou *sequencing-by-synthesis*
- Séquençage ciblé ou *targeted sequencing*
- Séquençage de régions cibles ou *targeted-sequencing (TS)*
- Séquençage du génome complet ou *whole-genome sequencing (WGS)*.
- Séquençage Massivement parallèle *massive parallel sequencing (MPS)*
- Terminaison cyclique et réversible (*Cyclic Reversible Termination* ou CRT)

LISTE DES ABREVIATIONS

- AA : Acide aminé
- AC: Adenylyl cyclase
- AMPc : AMP cyclique
- BAD: BCL2-Associated Agonist Of Cell Death
- BAM : Binary sequence Alignment/Map
- BCL: B-cell lymphoma
- C.B.P.E : Centre de Biologie et Pathologie Est
- CGH_array: Array comparative genomic hybridization
- CNV : Copy Number Variation
- CREB: C-AMP Response Element-binding protein
- CSRD : Cystein-serine rich domain
- Del : Délétion
- ddNTP : Didésoxynucléotide triphosphate
- dNTP : Désoxynucléotide triphosphate
- DPN: Diagnostics prénatal
- EGF : Epidermal Growth Factor
- EGFR: Epidermal Growth Factor Receptor
- ERK : Extracellular signal-regulated kinases
- FAK : Focal adhesion kinase
- FGFR : Fibroblast Growth Factor Receptor
- FISH: Fluorescent In Situ Hybridization
- GAP : GTPase-Activating Protein
- GDP : Guanosine diphosphate
- GEF : Guanine nucléotide exchange factor
- GTP: Guanosine triphosphate
- GRD : GAP-related domain
- HAVANA: Human and Vertebrate Analysis and Annotation
- HAS : Haute Autorité de Santé
- HB-EGF: Heparin-Binding Epidermal Growth Factor
- HER : Human Epidermal Growth Receptor
- HGNC: HUGO Gene Nomenclature Committee
- HGP : Human Genome Project
- IGFR: Insulin-like Growth Factor Receptor
- Indel : Insertion – délétion
- INCa : Institut National du Cancer

- Ins : Insertion
- LCRs : Low-copy repeats
- MAP : Mitogen-activated protein kinases
- MAPK : Mitogen Activated Protein Kinase
- MBD : méthyl-CPG-binding domain proteins
- Mcl-1: Myeloid Cell Leukemia 1
- MEK: Mitogen-activated extracellular signal regulated kinases
- MLPA : Multiplex ligation-dependent probe amplification
- MNP : Multiple nucleotide polymorphism
- MPNST : Malignant Peripheral Nerve Sheath Tumors
- MPS : massive parallel sequencing
- NCBI : National Center for Biotechnology Information
- NF1 : Neurofibromatose de type 1
- ORF : Open reading frame
- PCR : Polymerase chain reaction
- PI3K : Phosphatidyl Inositol 3-Kinase
- PDGFR: Platelet Derived Growth Factor Receptor
- PGM : Personal Genome Machine
- qPCR : Quantitative PCR
- RSK: Ribosomal s6 kinase
- SAM : Sequence Alignment/Map
- SH2 : Src homology-2
- SNCFC : Syndrome neuro-cardio-facio-cutané
- SPRED-1: Sprouty-related, EVH1 domain-containing protein 1
- SPR: Sprouty-related C-terminal cysteine-rich
- SNV : Short nucléotides variation
- TEM : Transition épithélio-mésenchymateuse
- TNF: Tumor necrosis factor
- TCL : Tâches café au lait
- TGFa : Transforming Growth Factor a
- TMAP: Torrent Mapping Alignment Program
- To : Téraoctets
- TRAIL: TNF-related apoptosis-inducing ligand
- UTR : Untranslated region
- WES : Whole-exome sequencing
- WGS : Whole-genome sequencing

INTRODUCTION :

La neurofibromatose de type 1 (NF1, OMIM : #162200) ou maladie de Von Recklinghausen est une des maladies génétiques les plus fréquentes qui touche de 1/3000 à 1/4000 individus¹. Décrite en 1793 par Tisellius, puis en 1882 par le docteur von Recklinghausen, dont elle garda le nom, son mode de transmission est de type autosomique dominant. La pénétrance est quasi-complète à 8 ans mais les manifestations cliniques sont hétérogènes, y compris intrafamiliales². Dans cette variabilité d'expression clinique, les formes mineures se limitent à des atteintes cutanées caractéristiques de la maladie (nombreuses taches café au lait précoces et des neurofibromes). Plus rares, les formes sévères peuvent être lourdement défigurantes, rapidement invalidantes ou mener à des évolutions malignes qui assombrissent le pronostic. La neurofibromatose de type 1 est secondaire à une modification moléculaire du gène *NF1* (OMIM : *13113). Ce dernier est un gène suppresseur de tumeur codant pour une protéine cytoplasmique de 2818 acides aminés : la neurofibromine.

Le syndrome de Légius (OMIM : #611431), appelé aussi syndrome NF1-like ou NF6, est un trouble génétique plus rare (1/40000) et à priori bénin. L'atteinte cutanée, proche de la NF1, est caractérisée par des macules «café-au-lait» multiples avec ou sans lentigines au niveau des plis cutanés. Le syndrome de Légius a longtemps été confondu avec la NF1 de part un tableau clinique proche pouvant évoquer une NF1 d'expression modérée ou incomplète. Cette pathologie a été caractérisée récemment au niveau génétique par une atteinte moléculaire d'un gène distinct du gène *NF1*. Le gène responsable est le gène *SPRED1* (OMIM : *609291), codant pour une protéine agissant en aval de la neurofibromine sur le contrôle de la voie de signalisation cellulaire Ras-MAPK. Des troubles de l'attention et de l'apprentissage sont le plus souvent décrits. La similitude de certains signes cliniques entre les deux pathologies peut rendre le diagnostic du syndrome de Légius difficile.

Le centre de compétence pour les neurofibromatoses (NF Rhône-Alpes Auvergne) est situé à Lyon et il existe depuis 1999. Il a été labellisé officiellement en 2009 par le ministère de la Santé comme centre de référence régional. De nombreux spécialistes, répartis dans différents hôpitaux lyonnais, prennent en charge les patients. Le centre de compétence prend en charge le parcours de soin du patient, de son diagnostic clinique et génétique, au suivi clinique et au traitement des complications. Il fait partie du réseau de soin NF-France, réseau permettant une harmonisation de la prise en charge au niveau national.

Le service de Génétique Moléculaire et Clinique de l'hôpital Edouard Herriot (Hospices Civils de Lyon) dirigé par le Pr. Calender, fait partie intégrale de cette organisation. Il effectue l'analyse de plus de 200 patients par an.

L'analyse est principalement provoquée en cas de difficultés diagnostiques et en cas de demande d'un diagnostic pré-natal (DPN). Au laboratoire jusqu'en 2015, l'analyse moléculaire des gènes *NF1* et *SPRED1* était faite de manière indépendante. Pour les patients *NF1*, les réarrangements chromosomiques, comptant pour environ 10% des mutations, sont recherchés en premier lieu par méthode MLPA (multiplex ligation-dependent probe amplification). Chez les patients négatifs, un screening dHPLC est ensuite réalisé, avec confirmation par séquençage Sanger des variants trouvés. Ces deux méthodes combinées permettent d'atteindre une sensibilité de 95%. L'analyse moléculaire du gène *SPRED1* est faite par un séquençage Sanger de tous les exons et régions introniques flanquantes. Elle est suivie, si aucun variant n'est mis en cause, d'une recherche de réarrangements chromosomiques par méthode MLPA. L'analyse des deux gènes est réalisée de manière séquentielle, ce qui allonge considérablement le rendu de résultat.

Depuis quelques années, de nouvelles techniques de séquençage dites de « haut-débit » sont apparues. Elles correspondent à une 2^{ème} génération de séquenceurs (ou *next-generation sequencing* (NGS)) bien plus performante et au débit spectaculairement augmenté. Grâce à cette technologie, les capacités d'analyse se sont considérablement accrues, avec la possibilité de réaliser pour une série de patients, le séquençage simultané de plusieurs gènes dans un temps plus court.

Dans l'objectif d'un transfert technologique, les Hospices Civils de Lyon (HCL) se sont dotés depuis 2012 d'une nouvelle plateforme de génétique localisée au Centre de Biologie et Pathologie Est (Bron). L'entreprise Life Technologies, depuis incorporée à Thermo Fisher Scientific, a remporté l'appel d'offre et les premiers séquenceurs haut-débit de paillasse ont été acquis : un PGM™ (*Personal Genome Machine*) et un Proton™. Parallèlement, le Laboratoire de Génétique Moléculaire et Clinique a bénéficié des projets INCa (Institut National du Cancer) de développement de ces nouveaux outils de séquençage dans les hôpitaux Français. Un projet de transfert technologique de la méthode d'analyse actuelle des gènes *NF1* et *SPRED1* a ainsi été initié. La technologie Ampliseq™ de Life Technologies a été choisie pour la création de bibliothèques (analyte final regroupant tous les amplicons à séquencer). Le séquençage haut-débit des deux gènes est ensuite réalisé par le PGM™.

La mise au point, la validation et le passage en routine de cette méthode sont les objectifs principaux de ce projet. La perspective d'un débit d'analyse bien supérieur et du séquençage

simultané des deux gènes permettra un rendu de résultat rapide. Notre attention s'est aussi portée sur d'autres critères, comme une sensibilité équivalente à la méthode actuelle, et un prix convenable par patient.

PARTIE I : REVUE BIBLIOGRAPHIQUE

1. TUMEURS ET RASOPATHIES

La tumorigenèse est un processus cellulaire et moléculaire complexe. Ses effets sur la cellule comprennent divers mécanismes et dérèglements, dont entre autres : la prolifération, l'échappement à l'apoptose et à la sénescence, la dédifférenciation cellulaire, l'angiogenèse des tissus environnants, l'invasion tissulaire et le développement de métastases. L'accumulation d'altérations génétiques, sur plusieurs étapes successives, conduit à l'acquisition du phénotype tumoral. Cette hypothèse, présente dès 1953³, a été démontrée puis largement reprise comme principe fondateur par Knudson en 1971⁴.

La voie de signalisation Ras-Raf-MEK-ERK (voie de Ras-MAPKinase ou Ras-MAPK) est l'une des mieux connues en biologie tumorale⁵. Sa régulation joue un rôle central dans le processus de tumorigenèse. Elle a une activité pléiotrope sur la cellule, comme l'induction de la croissance cellulaire, ou la chimiorésistance. Elle est activée de façon aberrante dans de nombreux cancers, comme par exemple dans 50% des leucémies aiguës⁵ ou encore les cancers colorectaux.

1.1. LA VOIE DE SIGNALISATION RAS/MAPK

1.1.1. LE SIGNAL PRIMAIRE ACTIVATEUR

En général, la voie RAS/MAPK est activée suite à la liaison d'un ligand (un facteur de croissance ou une cytokine) à un récepteur transmembranaire, qui va transmettre le signal extracellulaire à l'intérieur de la cellule.

La fixation du ligand induit la dimérisation du récepteur membranaire, qui possède alors la particularité d'avoir une activité enzymatique de phosphorylation de résidus tyrosine, appelée : « activité tyrosine-kinase » (figure 1). Elle permet l'activation du récepteur par autophosphorylation, et ainsi l'activation d'une cascade de phosphorylations de nombreuses protéines intracytoplasmiques qui sont généralement d'autres kinases.

Parmi les récepteurs de facteurs de croissance capables d'activer la voie Ras-MAPK, les principaux sont l'EGFR et les autres membres de la famille HER (et d'autres encore comme FGFR, IGFR, PDGFR).

Ces résidus phosphorylés servent de site d'amarrage pour un certain nombre de protéines intracellulaires dénommées adaptatrices, et reconnues par un domaine SH2.

Ces protéines à domaine SH2 jouent donc un rôle central dans la transmission des signaux intracellulaires, et recrutent pour cela des facteurs d'échanges de nucléotides (GEF). Ainsi, le complexe Grb2/hSos (protéine adaptatrice avec son GEF) active la protéine Ras dans la voie des Ras-MAPK et la protéine PI3K, induisant le recrutement de la kinase AKT dans la voie PI3K/AKT⁶.

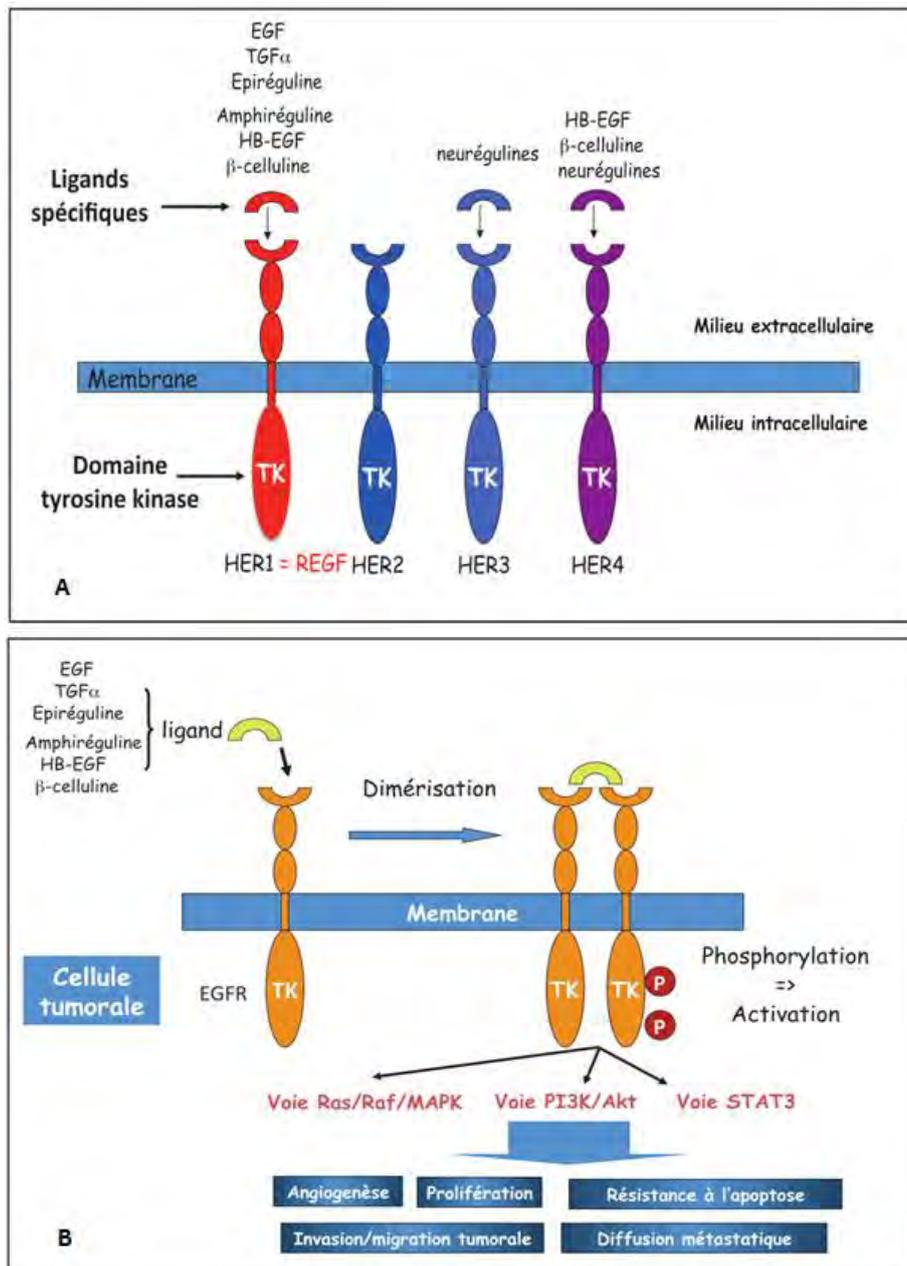


Figure 1. Famille des récepteurs de l'EGF (A), activation et vue d'ensemble des voies de signalisation intracellulaire induites (B)⁷.

1.1.2. LES PROTEINES RAS ET RAF

La famille des proto-oncogènes *RAS* comprend trois gènes bien caractérisés : *HRAS* (OMIM : *190020), *NRAS* (OMIM : *164790) et *KRAS* (OMIM : *190070, avec 2 isoformes: 4A et 4B) qui ont une très forte homologie (figure 2).

Les 4 protéines issues de ces gènes ont un poids moléculaire de 21 000 daltons, d'où leurs noms « p21 » souvent associés. Elles sont localisées au niveau de la face interne de la membrane cytoplasmique, ancrées dans la bicouche phospholipidique par leur extrémité C terminale (figure 2).

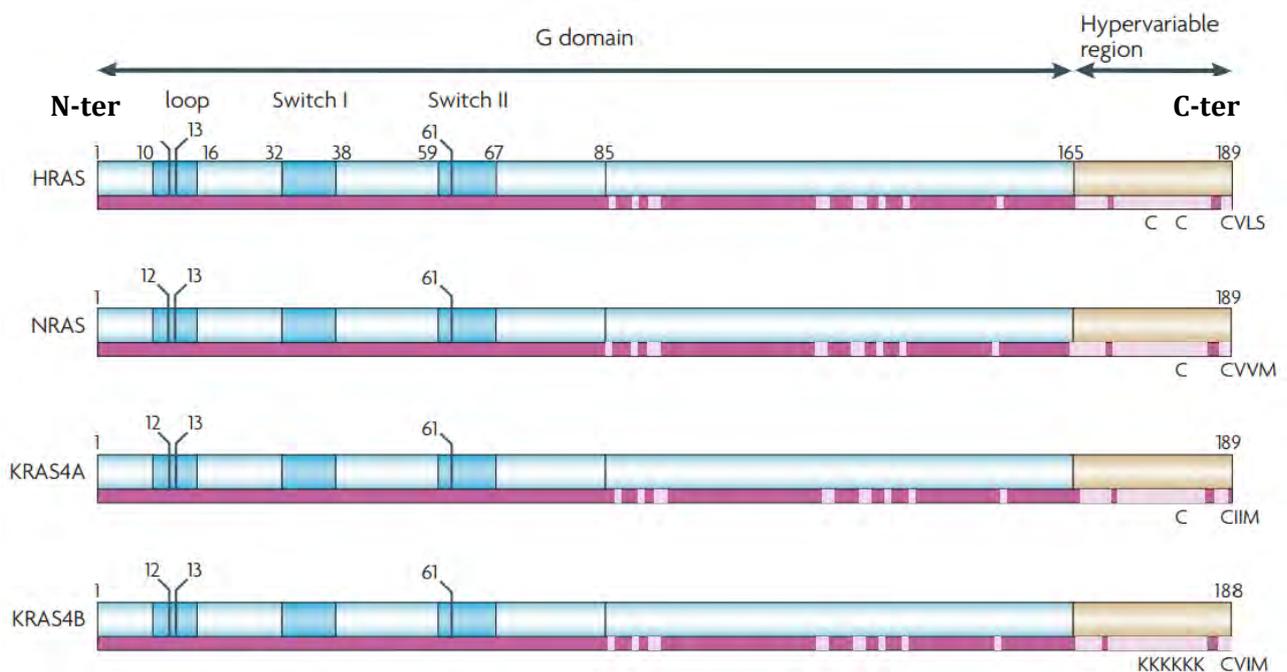


Figure 2. Les quatre isoformes des protéines Ras (d'après Schubbert et al⁸,2007).

Légende : *HRAS*, *NRAS*, *KRAS4A* et *KRAS4B* ont une très forte homologie de leur domaine G (des acides aminés 1 à 165). Les 85 premiers acides aminés sont identiques et se lient spécifiquement au GDP et GTP. La boucle P (P-loop) se lie au γ -phosphate du GDP, switch 1 et 2 permet un contrôle de la liaison des régulateurs et effecteurs de Ras. Les 80 acides aminés suivants ont une homologie de séquence de 80 à 90%. Enfin, le domaine hypervariable C-terminal spécifie la localisation membranaire des protéines grâce aux modifications post-traductionnelles subies (farnésylation sur les motifs CVLS, CVVM, CIIM et CVIM, et palmytylation sur les cystéines clés). Les mutations somatiques les plus fréquemment retrouvées dans les cancers se retrouvent aux positions 12, 13 et 61.

Les protéines Ras font partie de la famille des GTPases monomériques, et jouent un rôle important dans la transmission de signaux extracellulaires provenant des récepteurs membranaires vers le noyau. Les protéines Ras jouent un rôle « d'interrupteur » au sein des voies de signalisation. Elles se retrouvent dans deux états : un état actif où elles sont liées au GTP

(Guanosine Tri-Phosphate), ce qui permet transitoirement l'interaction de Ras avec d'autres molécules intracellulaires effectrices et l'activation de différentes voies de signalisation (RAS/MAPK, PI3K/AKT). L'activation de Ras est possiblement inhibée par les protéines de la famille Sprouty, dont SPRED1 fait partie (chapitre 4.2).

A contrario, elles sont à l'état inactif quand elles sont liées au GDP. Cette inactivation est provoquée par l'hydrolyse spontanée du GTP en GDP grâce à l'activité GTPasique intrinsèque de Ras⁹.

Cette activité GTPase peut aussi être modulée par des protéines de régulation, comme les protéines GAP (protéines activatrices de GTPase), avec la Neurofibromine ou la protéine p120GAP. Elles se lient aux protéines G et activent leur activité GTPase intrinsèque.

Les protéines Ras activent et recrutent les protéines de la famille Raf, constituée des sérine-thréonine-kinases ARAF, BRAF et CRAF (RAF-1), dont l'activité est modulée par des phosphorylations/déphosphorylations sur différents domaines et par des interactions avec de multiples cofacteurs.

1.1.3. LES PROTEINES MEK ET ERK

Les cibles des kinases Raf sont multiples, dont les protéines MEK 1 et 2, qui sont activées directement par phosphorylation (B-raf ayant le pouvoir activateur le plus fort). MEK1 et MEK2 possèdent une activité tyrosine et sérine/thréonine kinase. A leur tour, elles activent très spécifiquement par double phosphorylation ERK1 et ERK2, elles même sérine/thréonine kinases (faisant partie de la famille des MAP kinase). ERK est le dernier maillon connu de la cascade.

Activées, ces protéines migrent vers le noyau par translocation et par phosphorylation, et activent de multiples effecteurs cibles : facteurs de transcription (c-Myc, c-FOS) ou d'autres protéines-kinases (p90RSK). Au final, un grand nombre de gènes voient leur expression stimulée, par exemple ceux de la cycline D1 et de cdk6, qui ont un rôle majeur dans l'initiation du cycle cellulaire en phase G1.

1.1.4. EFFETS CELLULAIRES DE L'ACTIVATION DE LA VOIE RAS-MAPK

Les effets conséquents à l'activation de la voie Ras-MAPK sont multiples, et jouent sur :

La prolifération cellulaire :

La plupart des facteurs de transcription cibles de la voie de Ras-MAPK sont impliqués dans le contrôle de la prolifération et de la différenciation cellulaire : c'est le cas par exemple de c-Fos, c-Jun et c-Myc.

Le cycle cellulaire et apoptose:

Elle est impliquée dans la régulation de la survie cellulaire en partie via l'activation de la kinase p90RSK. L'activation de la voie de Ras-MAPK entraîne l'inactivation des protéines pro-apoptotiques BAD et Bim-1 (membres de la famille Bcl-2). Elle produit aussi l'activation du facteur de transcription CREB et de la protéine Mcl-1, anti-apoptotiques, qui empêchent la dépolérisation mitochondriale (voie intrinsèque de l'apoptose). Elle entraîne également une phosphorylation de la Caspase 9, contribuant à son inactivation. Enfin, l'activation de ERK1/2 peut inhiber l'apoptose induite par les récepteurs Fas, TNF et TRAIL (voie extrinsèque).

L'invasion et migration cellulaire :

La transition épithélio-mésenchymateuse (TEM) est un processus rattaché à la dédifférenciation cellulaire, au cours duquel les cellules épithéliales tumorales changent de phénotype et acquièrent celui de cellules mésenchymateuses. Ceci leur confère des propriétés de migration, d'invasion et de résistance aux traitements antitumoraux. La voie de Ras-MAPK est impliquée dans la TEM induite par le TGF β . Elle joue un rôle dans l'invasion et la migration cellulaire par l'augmentation de la production de certaines métalloprotéinases permettant la dégradation de la matrice extracellulaire et régulant l'expression et la localisation des protéines des jonctions serrées entre cellules.

Angiogenèse :

Enfin, la voie de Ras-MAPK joue aussi un rôle dans l'angiogenèse, notamment par l'induction de l'expression de la protéine VEGF via Raf.

1.2. MECANISMES PRIMORDIAUX DE LA CANCEROGENESE

Il a été montré depuis plusieurs années que la voie RAS-MAPK avait un rôle prépondérant dans la tumorigenèse.

Dans l'article princeps « The Hallmarks of Cancer »¹⁰, publié en 2000, les auteurs mettent en lumière les différents mécanismes fondamentaux cellulaires et extracellulaires mis en place pour l'initiation et l'évolution des cellules tumorales. La figure 3 reprend les « caractéristiques » fondamentales que la cellule saine doit acquérir afin de devenir une cellule cancéreuse.

Parmi ces caractéristiques primordiales, on retrouve :

- **L'autosuffisance en signaux de croissance :**

Dans un tissu normal, l'homéostasie précise et minutieuse de la production et de la sécrétion des facteurs de croissance est essentielle. Ces signaux permettent aux cellules d'entrer et de progresser dans le cycle cellulaire. La régulation permet de garder l'intégrité architecturale et fonctionnelle des tissus. Les cellules tumorales profitent d'une prolifération chronique et continue en devenant autosuffisantes en signaux de croissance. Ce concept est proche de celui du mode de régulation dit autocrine, où la cellule émettrice d'un signal de croissance est aussi la cellule cible.

- **L'évitement des signaux inhibiteurs de croissance :**

Dans les cellules, des gènes « suppresseurs de tumeurs » permettent de réguler négativement la prolifération des cellules. Par exemple, les gènes *RB* (codant pour la « retinoblastoma protein » et ayant permis les premiers travaux de Knudson) et *TP53* (au centre de la stabilité génomique et appelé aussi « the guardian of the genome ») sont deux prototypes de gènes suppresseurs de tumeurs. Les cellules cancéreuses, par divers mécanismes, échappent à leur action.

- **La capacité à envahir et créer des métastases :**

L'envahissement des organes se fait en plusieurs étapes, d'abord les tissus adjacents, puis possiblement les sites distants atteints par des métastases. Un mécanisme fondamental est la perte de molécule d'adhésion sur la membrane couplée au « homing » des cellules (spécificité tissulaire).

- **La propriété de se répliquer indéfiniment :**

Les cellules cancéreuses ne meurent pas en évitant l'apoptose. De plus, elles ont la capacité de se diviser indéfiniment, et ainsi sont dites « immortelles ».

- **L'induction de l'angiogenèse :**

Les cellules tumorales ont, comme les cellules normales, un besoin d'apport de nutriments et d'oxygène et un besoin d'élimination des déchets métaboliques et de CO₂. L'angiogenèse, mécanisme permettant la néo-vascularisation, est stimulée dans l'environnement de la tumeur afin de remplir au mieux ces besoins.

- **La résistance au mécanisme de mort cellulaire :**

Se répliquer indéfiniment signifie aussi échapper à l'apoptose, étape pendant laquelle les cellules programment leur « mort ». Une des stratégies les plus communes est la perte de fonction de *TP53* qui joue sur le circuit inducteur de l'apoptose.

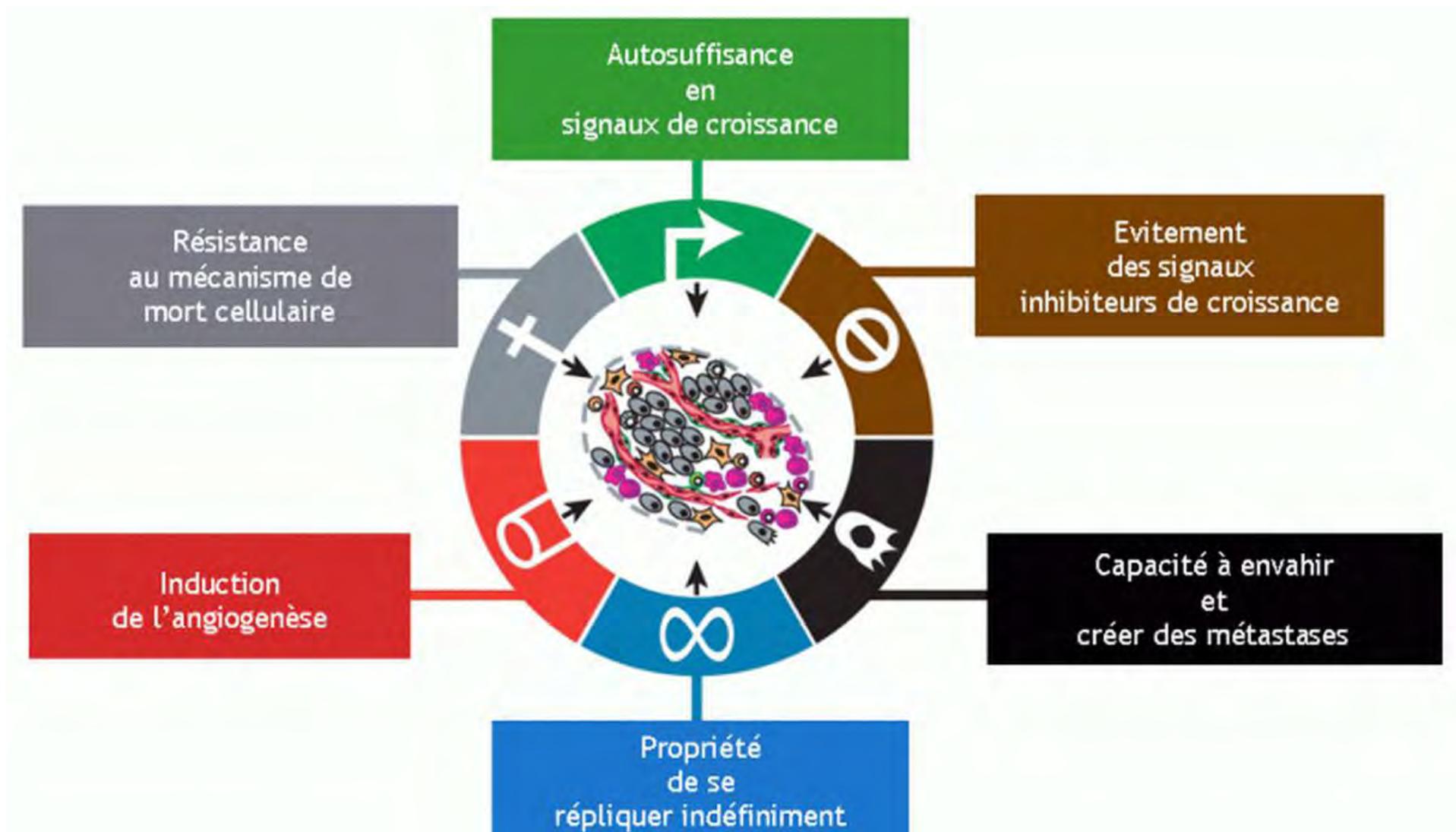


Figure 3. "The hallmarks of cancer", caractéristiques fondamentales des cellules cancéreuses (d'après Hanahan et Weinberg, 2000¹¹).

En 2011, Hanahan et Weinberg actualisent leur article en ajoutant quatre nouveaux «hallmarks»¹² (figure 4).

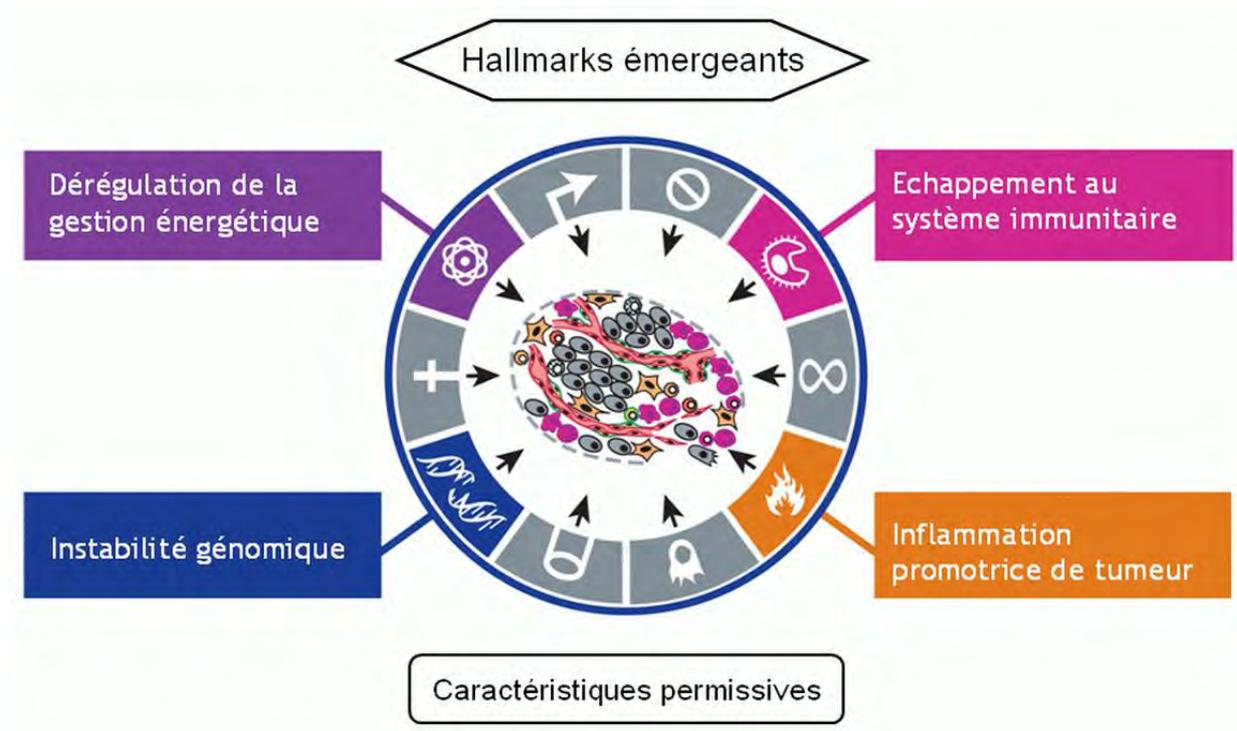


Figure 4. Les nouveaux « hallmarks », caractéristiques fondamentales des cellules cancéreuses (d'après Hanahan et Weinberg, 2011¹²).

Dans cette publication, les nouvelles caractéristiques se divisent en deux : les critères émergents et les permissifs.

- Les « Hallmarks » dits **émergents** sont :

- **La dérégulation de la gestion énergétique**

Ceci afin de soutenir la prolifération et l'augmentation du métabolisme des cellules tumorales. Il s'agit en particulier d'un passage de la respiration oxydative à la glycolyse (effet Warburg), avec augmentation des lactates, et par conséquent une dérégulation de l'expression de nombreux gènes.

Même si mis en évidence depuis longtemps, il n'est toujours pas démontré que cette dérégulation est indispensable à la conversion d'une cellule tumorale en cancer.

- **L'échappement des cellules tumorales au système immunitaire**
Comme pour la dérégulation du métabolisme, l'échappement au système immunitaire n'est pas clairement prouvé être indispensable à la tumorigenèse. Cependant, la perte de plusieurs antigènes de surface (dédiérenciation) est l'un des mécanismes d'échappement à la surveillance du système immunitaire.
- Les « Hallmarks » dits **permissifs** correspondent à :
 - **L'instabilité génomique**
L'instabilité génomique est causée par l'incapacité d'une cellule à maintenir son information génétique stable (déstabilisation des séquences répétées, erreurs de réplication). Elle peut engendrer des mutations aléatoires ou des réarrangements chromosomiques qui altèrent l'intégrité génétique des cellules et conduit à une progression de la maladie et une expansion des phénotypes tumoraux.
 - **L'inflammation promotrice de tumeur**
Le système immunitaire est en première ligne pour lutter contre les cellules tumorales, mais peut aussi de manière indirecte, renforcer tous les points précédents par la production de cytokines consécutive d'une modification du microenvironnement péri-tumoral.

Dans ce travail, les auteurs citent à plusieurs reprises la voie Ras-MAPK comme ayant une action directe ou indirecte sur les critères communs fondamentaux aux tumeurs lorsque un ou plusieurs des nombreux effecteurs de la voie de signalisation sont altérés et/ou déficients.

La voie RAS-MAPK peut ainsi jouer sur :

- **L'autosuffisance en signaux de croissance**
→ Par l'hyperproduction de facteurs de croissances ou l'altération (mutation activatrice, amplification, surexpression) de récepteur membranaire, en particulier EGFR.
- **L'évitement des signaux inhibiteurs de croissance**
→ La protéine Ras est le prototype de ce mécanisme. Par exemple, des mutations qui engendrent une réduction de l'activité GTPase permettant un rétrocontrôle négatif du signal transmis par Ras.

- **La résistance aux mécanismes de mort cellulaire**
 - ➔ L'augmentation de la quantité et de l'activité de Ras peut empêcher les mécanismes cellulaires de sénescence et/ou d'apoptose.
- **L'induction de l'angiogénèse**
 - ➔ Ras peut réguler positivement la production de facteurs angiogéniques
- **La dérégulation du métabolisme**
 - ➔ Ras peut entraîner une hausse de la glycolyse

1.3. DE LA MUTATION A LA TUMEUR

De nombreuses mutations somatiques sont retrouvées dans différents effecteurs de la voie RAS-MAPK dans les tumeurs humaines.

Dans 20 à 30% des tumeurs¹³, il existe une mutation ponctuelle activatrice de *RAS*, avec une prévalence forte dans les adénocarcinomes du pancréas (90%), du colon (50%), de la thyroïde (50%), des poumons (30%) et des mélanomes (25%). Ces mutations apparaissent le plus souvent sur *KRAS* (95% du total), puis sur *NRAS* ($\approx 15\%$), et plus rarement sur *HRAS*.

Des mutations récurrentes ou « hotspots » existent dans les gènes *RAS* qui mènent le plus souvent aux changements des acides-aminés suivant : G12, G13 et Q61. Ces altérations génétiques diminuent l'activité intrinsèque d'hydrolyse du GTP des protéines Ras et la réponse aux GAPs. Il en résulte une activation constitutive sous la forme GTP et *de facto*, la stimulation sous forme active des effecteurs en aval de la voie de signalisation (dont la voie des MAPKinases, mais d'autres voies peuvent être activées).

RAF, l'effecteur en aval de RAS, est aussi une cible d'altération génétique fréquente et environ 7% des cancers humains présentent une mutation du gène *BRAF*. On les retrouve par exemple dans les mélanomes (70%), les cancers de la thyroïde (30%) ou encore dans 15% des cancers du côlon. La plupart des mutations de RAF se trouvent dans son domaine kinase (V600E étant la plus connue) et engendrent une activation constitutive de la protéine, et par la même, des effecteurs en aval MEK et ERK.

Il est à noter qu'une seule altération d'un acteur de la voie RAS-MAPK semble suffire à l'activation anormale de la voie de signalisation et qu'il est fréquent qu'un seul des acteurs de cette voie ne soit touché.

Voici une figure résumant une partie des données précédemment décrites :

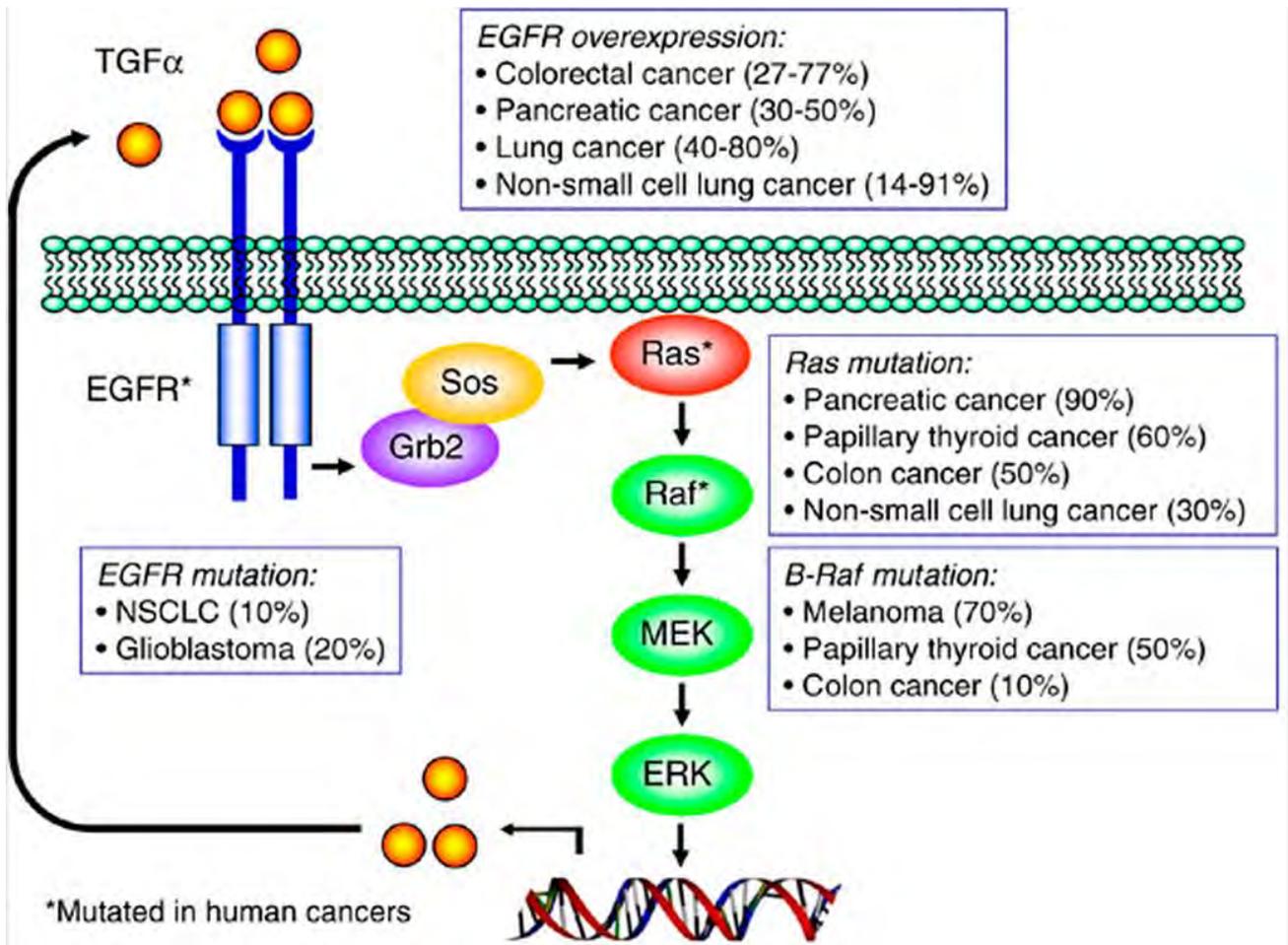


Figure 5. Effecteurs de la voie Ras-MAPK et leurs implications tumorales (Roberts PJ et al¹⁴, 2007).

1.4. LES RASOPATHIES : UN GROUPE HETEROGENE DE PATHOLOGIES

Comme nous l'avons vu, la voie de signalisation cellulaire des RAS/MAP Kinases (RAS/MAPK) joue un rôle majeur sur le devenir des cellules en participant à la régulation de la prolifération, de la différenciation et de la survie cellulaire. L'étude de modèles murins a montré que cette voie de signalisation joue un rôle important dans le développement des mammifères, notamment lors du stade embryonnaire¹⁵, et dans l'homéostasie des tissus chez l'adulte¹⁶. Jusqu'en 2006 et dans un premier article fondateur¹⁷, plusieurs maladies, de par leurs tableaux cliniques confondants et chevauchants, étaient regroupées sous l'acronyme SNCFC ou syndrome neuro-cardio-facio-cutané. Le point commun et central de ces pathologies est la dérégulation de la voie de signalisation cellulaire RAS-MAPK par une atteinte **constitutionnelle** d'un des acteurs de la voie (à distinguer des mutations somatiques vu précédemment). A cause de l'atteinte de la même voie de signalisation, de nombreux traits phénotypiques communs sont présents à des degrés variables. On retrouve notamment: retard de croissance, retard mental avec trouble de l'apprentissage, dysmorphie cranio-faciale, malformation cardiaque et anomalies cutanées.

Leur différence importante (essentielle dans la prise en charge du patient) se joue sur la prédisposition et la prévalence à certaines hémopathies ou cancers solides, qui varie considérablement d'une pathologie à une autre. Ces différents syndromes sont maintenant regroupés sous le terme de « Rasopathies ». Ils représentent un des plus grands groupes connus de syndromes malformatifs, et environ 1 / 1 000 individus partout dans le monde en sont atteints.

1.5. LES DIFFERENTS SYNDROMES COMPOSANT LES RASOPATHIES

Depuis l'identification de NF1 comme la première des Rasopathies dans les années 90, par l'inactivation de la neurofibromine, beaucoup d'autres syndromes ont été associés à des mutations à divers niveaux de la voie RAS-MAPK (figure 6).

Le tableau 1 récapitule les principaux syndromes regroupés sous le terme de Rasopathies, les gènes et protéines associés mis en cause, leur prévalence et clinique, et les cancers les plus fréquemment retrouvés pour chaque pathologie.

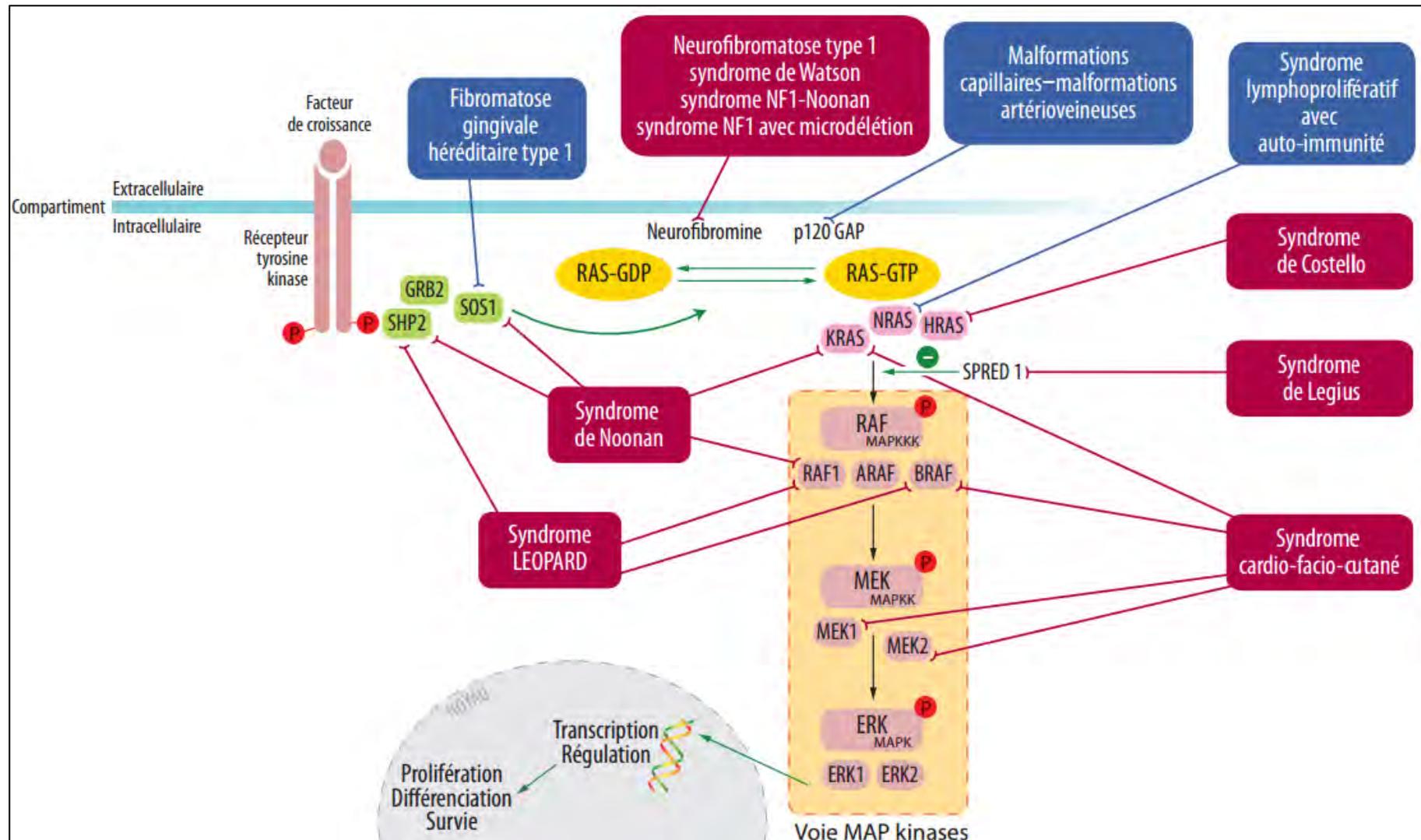


Figure 6. La voie de signalisation Ras-MAPK et les différentes RASopathies associées¹⁸.

Les mutations causales sont constitutionnelles et présentes sous forme hétérozygote (caractère dominant des Rasopathies). Elles mènent à l'activation de la voie à un degré moindre que les mutations somatiques associées à l'oncogénèse.

Bien que certaines mutations soient les mêmes que celles retrouvées dans certains cancers, les mutations les plus communes telles que la mutation V600E pour le gène *BRAF* n'apparaît pas dans le syndrome CFC, de même pour *KRAS* dont les mutations ne sont pas les mêmes dans le Syndrome de Noonan par rapport à celles retrouvées dans les tumeurs associées (du côlon par exemple). Une explication possible est la possibilité que certaines mutations constitutionnelles, les plus délétères, entraînent une mort embryonnaire et donc ne sont pas retrouvées¹⁹. Elles peuvent cependant apparaître de manière somatique.

Il reste encore beaucoup de mécanismes à découvrir afin de mieux appréhender ces différents syndromes. Les récentes avancées en matière de séquençage du génome facilitent l'analyse génétique des patients et permettent de mettre en évidence de nouvelles mutations causales.

Ainsi de nouveaux gènes sont régulièrement associés à ces différents syndromes, démontrant à nouveau que le chevauchement clinique peut s'expliquer par la multiplicité des cibles atteintes.

Par exemple, en 2015 il a été trouvé chez 50 patients brésiliens avec un phénotype de syndrome de Noonan, des variants dans les gènes *SOS2* et *LZTR1* (respectivement chez 4% et 8% des patients), mutations jusque-là jamais publiées²⁰ (non présentés sur la figure 6). Dans une autre publication récente, c'est le gène *RIT1* qui est mis en cause chez un enfant présentant un phénotype de Noonan associé à des problèmes hématologiques²¹. Une meilleure connaissance de la fonction de ces gènes, non connus pour participer de façon directe sur la voie Ras-MAPK, permettra de faire le lien entre ces différents effecteurs.

Ainsi, de nombreuses questions restent en suspens et le champ de la recherche est large pour les Rasopathies. Des recherches doivent être entreprises afin de mieux comprendre l'impact des différentes mutations sur la voie de signalisation. En effet, il a été suggéré qu'à l'inverse d'une activation de la voie RAS-MAPK commune à toutes les Rasopathies, certaines mutations de *BRAF* dans le syndrome Cardio-facio-cutané ou de *PTPN11* dans le syndrome LEOPARD ont un effet négatif avec baisse de l'activité. De plus, certains mécanismes semblent impliquer des effecteurs en amont de la voie RAS-MAPK ou bien d'autres voies de signalisation. Enfin, il reste à comprendre comment des mutations oncogéniques peuvent être tolérées en constitutionnel et ce qui définit le risque de malignité selon les différents gènes mis en cause

| Pathologies | Gènes (localisation chromosomiques) | Protéines (fonctions) | Prévalence | Caractéristiques phénotypiques | Cancers associés |
|---|---|--|------------------------|---|--|
| Neurofibromatose de type 1 (NF1) (OMIM: *162200) | NF1 (17q11.2) | Neurofibromine (RasGAP) | 1/ 2 500 | Cf clinique NF1 | Neurofibromes, astrocytomes, phéochromocytome, Leucémie myélomonocytaire juvénile (LMMJ), MPNST |
| Syndrome de Watson (forme allélique de NF1) (*193520) | NF1 (17q11.2) | Neurofibromine (RasGAP) | Inconnue | - Sténose pulmonaire - Macules couleur café au lait - Fonctions neurocognitives déficiences - Petite stature - Macroencéphalie - Nodules de Lisch | Cf NF1 |
| Syndrome de Legius (SL) (*611431) | SPRED1 (15q14) | SPRED1 (SPROUTY- related, EVH1 domain-containing protein 1) | < 200 cas rapportés | - Macules couleur café au lait - Macroencéphalie - Fonctions neurocognitives normales ou légèrement déficientes - Éphélides intertrigineuses | Aucune prédisposition au cancer solidement documentée |

| | | | | | |
|---|---|---|---------------------------------|--|--|
| Syndrome LEOPARD (SL) (*151100) | <i>PTPN11</i> (12q24) | PTP2C (Phosphatase) | < 200 cas rapportés | Phénotype similaire au SN - Apparition d'une lentiginose cutanée - Prédisposition au cancer inconnue | Neuroblastomes, leucémies myéloïdes |
| Syndrome de Noonan (SN) (*163950 / 609942 / 610733 / 611553 / 613224 / 613706) | <i>PTPN11</i> (12q24.1) <i>SOS1</i> (3p22.1) <i>RAF1</i> (3p25.1) <i>KRAS</i> (12p12.1) <i>NRAS</i> (1p13.2) <i>CBL</i> (11q23.3) Autres (<i>SOS2</i> , <i>LZTR1</i> , <i>RAIT1</i> ...) | SHP2 (Phosphatase) SOS1 (RasGEF) CRAF (Kinase) KRAS (GTPase) NRAS (GTPase) CBL (E3 ubiquitin ligase) | 1/1 000 - 1/2 500 | - Traits craniofaciaux dysmorphiques distinctifs - Défauts cardiaques congénitaux - Petite stature - Anomalies ophtalmiques - Défauts de coagulation - Fonctions neurocognitives normales ou légèrement déficientes -Cryptorchidie | LMMJ Prédisposition au cancer |
| Syndrome de Noonan avec cheveux anagènes fragiles (*607721) | <i>SHOC2</i> (10q25.2) | SHOC2 (Scaffolding) | 1/1 000 000 | Forme variante du SN - Phénotype similaire au SN - Cheveux cassants | / |
| Malformation capillaire - malformation artérioveineuse (*608354) | <i>RASA1</i> (5q14.3) | p120-RasGAP (RasGAP) | inconnue | - Malformations capillaires pouvant être associées à des Malformations artérioveineuses - Prédisposition au cancer incertaine | |

| | | | | | |
|---|---|---|----------------------------------|---|--|
| <p>Syndrome cardiofacio-cutané (CFC) (*115150)</p> | <p><i>BRAF</i> (7q34)</p> <p><i>MAP2K1</i> (15q22.31)</p> <p><i>MAP2K2</i> (19p13.3)</p> <p><i>KRAS</i> (12p12.1)</p> | <p>BRAF (Kinase)</p> <p>MEK1 (Kinase)</p> <p>MEK2 (Kinase)</p> <p>KRAS (GTPase)</p> | <p>Plus de 100 cas rapportés</p> | <ul style="list-style-type: none"> - Traits craniofaciaux similaires au SN mais pouvant être plus prononcés - Défauts cardiaques congénitaux - Troubles alimentaires graves - Petite stature - Anomalies ophtalmiques - Lésions cutanées - Fonctions neurocognitives déficientes de légères à graves - Hypotonie | <p>Prédisposition au cancer incertaine</p> |
| <p>Syndrome de Costello (SC) (*218040)</p> | <p><i>HRAS</i> (11p15.5)</p> | <p>HRAS (GTPase)</p> | <p>< 250 cas rapportés</p> | <ul style="list-style-type: none"> - Traits craniofaciaux similaires au SN mais pouvant être plus prononcés - Défauts cardiaques congénitaux - Troubles alimentaires très graves - Petite stature - Anomalies ophtalmiques - Lésions cutanées - Fonctions neurocognitives déficientes - Hypotonie - Prédisposition au cancer | <p>Rhabdomyosarcomes, neuroblastomes, ganglioneuroblastomes, cancer de la vessie</p> |

Tableau 1. Les principales pathologies regroupées sous le terme : Rasopathies.

2. LA NEUROFIBROMATOSE, UNE « RASOPATHIE » COMPLEXE

2.1. LA NEUROFIBROMATOSE DE TYPE 1

La neurofibromatose de type 1 fait partie d'un groupe hétérogène de maladies appelé : les « phacomatoses » (du grec : « *Phakos* » : tâche, soulignant l'aspect disséminé ou « en tâche » et « *Oma* » : tumeur). Les phacomatoses regroupent une quinzaine d'affections génétiques congénitales et héréditaires avec atteintes neuro-cutanées. Elles ont en commun des défauts de pigmentation cutanée, des modifications angiomateuses, des tumeurs et hamartomes localisés volontiers au niveau du tissu nerveux et responsables d'altérations neurologiques. En fait, bien qu'aucun tissu ni aucune fonction ne soient complètement exempts d'anomalies, c'est au niveau du système nerveux central et périphérique que siège l'essentiel de cette classe de pathologies.

Les principales maladies comprises dans ce groupe sont les Neurofibromatoses, la Sclérose tubéreuse de Bourneville, la maladie de Von Hippel-Lindau, et les phacomatoses angiomateuses (Rendu-Osler).

2.2. HISTORIQUE

Connue depuis l'antiquité²², la Neurofibromatose de type 1 a été décrite en 1882 par un médecin Allemand, le docteur Von Recklinghausen²³, qui donna son nom à cette nouvelle maladie.

Cette affection peut, de par son atteinte cutanée, amener à un préjudice esthétique lourd et handicapant socialement pour le patient, même si cela reste purement subjectif pour de nombreux patients.

Son diagnostic clinique est connu depuis longtemps, et a même été postulé pour la description physique de Quasimodo, le personnage de V.Hugo²⁴. Même si erroné, le cas de Joseph Merrick « Elephant Man », fut évoqué en 1909 comme une manifestation exceptionnelle de la NF1, avant que ne soit corrigé le diagnostic 90 ans plus tard, en soupçonnant une maladie encore plus rare : le syndrome de Protée²⁵. Cette erreur reste malheureusement trop souvent reprise sur internet ou dans les médias.

2.3. ÉPIDÉMIOLOGIE – PREVALENCE

Ce désordre génétique est une des maladies génétiques constitutionnelles les plus fréquentes. Avec une incidence estimée à 1 cas pour 2500 naissance, elle touche de 1/3000 à 1/4000 individus avec une répartition mondiale homogène²⁶.

Les hommes et les femmes sont atteints dans les mêmes proportions et avec une diversité d'expression équivalente.

La pénétrance (proportion d'individus possédant un génotype muté et qui exprime le phénotype correspondant) est très forte et proche des 100% à l'âge adulte.

2.4. PRESENTATION CLINIQUE

Les manifestations de la neurofibromatose sont hétérogènes, et dans une famille avec plusieurs personnes atteintes et portant la même mutation, la présentation clinique peut être variable. Il est rare que les patients présentent l'ensemble des signes caractéristiques de la maladie, qui atteint la peau, le système nerveux, l'œil et d'autres organes, avec un risque de développement de certaines tumeurs.

Les formes mineures et modérées, avec parfois seulement des signes cutanés, sont les formes les plus fréquentes. Les formes graves, dues à certaines complications secondaires ou à des cancers, ne représentent que 15 % des cas.

2.4.1. CRITERES DU NIH

Le phénotype des patients atteints de Rasopathies sont extrêmement variables. La NF1 ne déroge pas à cette constatation : les critères cliniques sont confondants et souvent non spécifiques, si pris de manière isolée.

Afin de simplifier le diagnostic de la NF1, un score clinique a été créé. Après examen clinique, lors d'auscultations réalisées par un spécialiste, le score est établi. S'il est suffisant, c'est-à-dire au minimum : 2, le diagnostic de NF1 est posé.

Apparu lors de la conférence de consensus du National Institute of Health (NIH) de Bethesda (U.S.A) en 1998, les critères utiles au diagnostic sont au nombre de sept (Tableau 2).

Ces critères peuvent être répartis en deux sortes : les non-spécifiques (1, 2, 3 du tableau 2) et les spécifiques ou quasi-pathognomoniques (4, 5, 6, 7). Pris ensemble, ces critères permettent une grande spécificité et sensibilité du diagnostic de la NF1²⁷.

La règle : deux critères parmi les suivants suffisent pour établir qu'un patient est atteint d'une neurofibromatose (tableau 2).

| Critères cliniques : 2 minimum sont nécessaires pour poser le diagnostic |
|--|
| 1- Un apparenté du premier degré atteint (parent, fratrie ou enfant). |
| 2- Au moins 6 taches café au lait de diamètre : <ul style="list-style-type: none">○ supérieur à 1,5 cm après la puberté○ supérieur à 0,5 cm avant la puberté. |
| 3- La présence de lentigines axillaires ou inguinales. |
| 4- Au moins deux neurofibromes quel que soit leur type ou un neurofibrome plexiforme. |
| 5- Un gliome du nerf optique. |
| 6- Au moins deux nodules de Lish ou plus (hamartomes iriens). |
| 7- Une lésion osseuse caractéristique (amincissement du cortex des os longs, dysplasie du sphénoïde, ou pseudarthrose). |

Tableau 2. Critères diagnostique du NIH (1998).

A partir de 5 ans⁶, la pénétrance de la maladie est élevée, et devient quasi-complète à 8 ans. Avec les critères du NIH, le diagnostic ne pose le plus souvent pas de problème et l'étude moléculaire n'est pas nécessaire. L'âge du patient reste un critère important.

Pour un enfant en bas âge, les symptômes cliniques se limitent souvent aux atteintes cutanées et le diagnostic peut être plus difficile à établir, voire confondu avec une autre RASopathie. Par exemple, les nodules de Lisch (hamartomes iriens), un des signes d'appel les plus spécifiques de la NF1, ne sont présent que chez 10% des enfants de moins de deux ans, mais retrouvés chez la majeure partie des adultes (jusqu'à 90% dans certaines études).

Un enfant sur deux, sans histoire familiale de NF1, rentre ainsi dans les critères diagnostique du NIH. De possibles échecs de diagnostic sont dus à l'évolution clinique tardive chez certains patients. Dans les formes incomplètes, partielles, ou en mosaïque, le diagnostic moléculaire trouve toute sa place et permet d'affirmer ou non l'atteinte. Son importance est grande, car il permet ainsi un suivi précoce et un dépistage rapide des apparentés.

2.4.2. CRITERE DU NIH : LES MANIFESTATIONS CLINIQUES EN DETAIL

Les atteintes cutanées (3 critères) :

Ce sont les signes les plus constants et les plus précoces, parfois l'unique point d'appel de la maladie. Ces manifestations sont diverses et se présentent sous la forme de taches pigmentées (aussi appelées taches café au lait ou TCL) ou de tumeurs cutanées bénignes. Elles ne relèvent pas toujours d'un caractère de gravité particulier en dépit de leur aspect non esthétique. Cependant elles peuvent avoir des conséquences indirectes, notamment dû à la compression nerveuse des neurofibromes.

- Les TCL sont de couleur marron clair, arrondies ou ovales, de taille variable (qui augmente lors de la croissance, sans relief (figure 7). La répartition est aléatoire et l'examen à la lumière de Wood aide à leur reconnaissance. Les TCL résultent d'une hyperactivité avec discrète augmentation du nombre de mélanocytes basaux²⁸.

Leur pénétrance est complète dès les premières années de vie, et sont au final présentes chez plus de 95% des individus NF1.

Le nombre et leur taille, fonction de l'âge, est un des critères principal de diagnostic, même si jusqu'à 25% des enfants de la population générale peuvent être porteurs de 1 à 3 TCL²⁹.



Figure 7. Taches café au lait de grande taille³⁰.

- Les lentigines (ou éphélides), désignent les TCL de petite taille avec un aspect de tache de rousseur et sont localisées au niveau des grands plis (au niveau axillaire, de l'aîne, et du cou). Le diamètre est inférieur à 3 mm et la prévalence de 50 à 80% à l'âge de 6 ans³⁰.



Figure 8 Lentigines axillaires³¹.

Les tumeurs cutanées retrouvées chez les patients NF1 sont de plusieurs types et de pronostics différents, correspondant à diverses formes histologiques :

- Les *neurofibromes superficiels* sont des tumeurs bénignes situés à la surface de la peau, de couleur clair ou foncée, molle, et qui ne dégénèrent pas en tumeurs malignes. Ils sont présents chez au moins 95% des adultes atteints.

Les neurofibromes profonds circonscrits comportent :

- Les *neurofibromes sous-cutanés*, présents chez 20% des malades, sont quelques fois douloureux à la palpation et ont un risque de transformation maligne
- Les *neurofibromes plexiformes*, retrouvés dans 25% des cas, sont des neurofibromes mixtes, et plus volumineux. Ils se situent préférentiellement au niveau de la paupière et des membres ou du tronc. Souvent présents dès la naissance, ils peuvent parfois dégénérer (Neurofibrosarcome).

- Les *neurofibromes plexiformes nodulaires* sont moins fréquents, et se regroupent en grappe, disposés le long de tronc nerveux, et développés à partir des racines nerveuses.

Ces atteintes cutanées ont un impact fort notamment sur le plan esthétique. Parfois, des atteintes ophtalmologiques secondaires peuvent être associées et situées sur la face (amblyopie).

Les conséquences psychologiques et sociales sont importantes, tout en étant très subjectives et dépendant de chaque patient. A l'adolescence, lors de l'évolution ou l'apparition des critères cliniques, le patient peut avoir besoin d'une aide psychologique.



Figure 9 Neurofibromes cutanés multiples³⁰.

Manifestations ophtalmiques (2 critères) :

- Les *nodules de Lisch* sont des nodules pigmentés de l'iris (hamartomes mélaniques iriens) quasi-pathognomoniques de la pathologie. Ils n'ont aucune conséquence physiologique mais sont un argument positif pour le diagnostic (particulièrement utiles dans les formes paucisymptomatiques). A l'inverse, leur recherche est inutile pour un patient adulte avec un diagnostic évident. Leur mise en évidence est réalisée par un examen ophtalmologique particulier par une lampe à fente. Leur nombre et leur taille augmentent avec l'âge³², et ils sont retrouvés chez majorité de malades après 16 ans (60-70%).

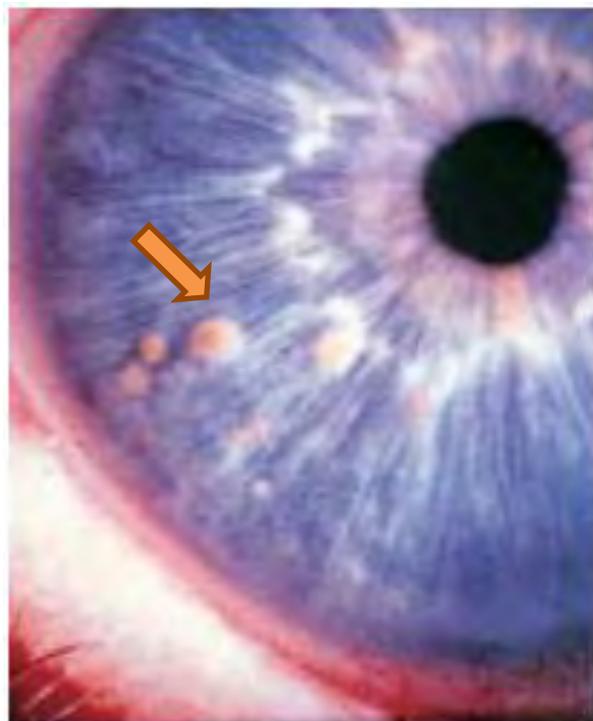


Figure 10 Hamartomes iriens ou nodules de Lisch³² (flèche orange).

- Les *gliomes des voies optiques* sont des tumeurs intracérébrales peu fréquentes, qui interviennent souvent dans l'enfance, avec une évolution imprévisible. L'incidence est difficile à estimer car de nombreuses formes restent asymptomatiques. On retrouve la présence de gliomes dans $\approx 15\%$ des cas.
Dans environ 1 cas sur 3, les patients sont symptomatiques. L'examen ophtalmologique est alors perturbé. La découverte se fait de manière fortuite lors d'un examen systématique chez les enfants³³. La période à haut risque de découverte est celle de la petite enfance (6 ans en moyenne).

En cas d'évolution, la baisse d'acuité visuelle peut être unilatérale, si la tumeur est située au niveau du nerf optique, ou bilatérale si l'atteinte est chiasmatique. Elle est peut être accompagnée d'un strabisme ou d'un nystagmus. Pour la plupart des patients, l'acuité visuelle est stable mais la surveillance se fait jusqu'à l'adolescence. Alors que la surveillance ophtalmique des signes d'évolution est primordiale pendant l'enfance, elle devient inutile par la suite (arrêt de la surveillance à 12 ans). Seuls les gliomes évolutifs ou symptomatiques seront traités. L'imagerie par IRM permet de faire le diagnostic en examinant attentivement les nerfs et chiasma optique. Cette lésion reste fortement associée à la NF1 et constitue un signe d'appel fort: près de 70% des patients avec un gliome des voies optiques sont des patients NF1.



Figure 11. Gliome du nerf optique droit³⁰ (flèche orange).

- Une exophtalmie, rare, peut se retrouver chez les patients NF1. Elle peut être la conséquence d'une dysplasie orbitaire (agénésie de la grande aile du sphénoïde) avec augmentation de taille du globe oculaire. L'exophtalmie peut aussi être liée à des neurofibromes orbitaires ou à un gliome des voies optiques.

Manifestations osseuses (1 critère) :

Le diagnostic de NF1 peut être fait au détour d'une consultation d'orthopédie pédiatrique. Les conséquences des anomalies osseuses peuvent devenir sévères et influencer grandement sur la qualité de vie future du patient. Même si l'incidence des atteintes osseuses est forte (jusqu'à 50% des cas), moins de 10% des patients auront besoin d'un traitement spécifique au cours de leur vie³⁴.

Les manifestations orthopédiques touchent entre autres le rachis et les membres, et sont souvent due à une dystrophie osseuse intrinsèque associée à une atteinte des parties molles. Parfois, ces atteintes sont associées à des neurofibromes plexiformes, ou en sont les conséquences.

- L'atteinte du rachis est la plus fréquente des atteintes orthopédiques, avec pour une grande majorité des cas une déformation vertébrale (cyphose, scoliose). La grande majorité de ces scolioses NF1 est similaire aux scolioses idiopathiques. Environ 4% des scolioses NF1 sont dystrophiques (avec érosion antérieure et postérieure des corps vertébraux).
- Les scolioses dystrophiques ont un potentiel évolutif d'autant plus sévère que la dystrophie est marquée. Des anomalies de courbure sagittale (cyphose, lordose) peuvent accompagner les atteintes du rachis. Ces atteintes du rachis peuvent entraîner un risque important de troubles de la statique.
- Les atteintes des membres consistent en des pseudoarthroses des os longs avec courbure, le plus souvent du tibia ou radius. Elles touchent 2 à 4% des enfants porteurs d'une NF1. Elles sont spécifiques, car on estime que 45 à 80% des pseudoarthroses congénitales sont dues à la NF1. Le traitement des courbures et une consolidation des os est primordiale. Le risque d'amputation existe encore aujourd'hui afin de permettre un appareillage adapté et fonctionnel.



Figure 12. Pseudoarthrose du tibia congénitale³⁰ (flèches orange).

- Les dysplasies cranofaciales sont rares mais peuvent provoquer une exophtalmie importante. La plus souvent retrouvée est la dysplasie du sphénoïde, qui touche le plus souvent une grande aile. Elle est évocatrice de la NF1 (50% des dysplasies du sphénoïde sont des NF1).

Pouvant se manifester à tout âge dans l'enfance, ces atteintes osseuses vont s'aggraver considérablement si elles ne sont pas prises en charge rapidement et précocement. Sans intervention sur le rachis, elles peuvent aller jusqu'à entraîner des complications cardio-respiratoires.

2.4.3. LES AUTRES ATTEINTES CLINIQUES :

Les manifestations neurologiques :

Ces signes sont liés d'une part aux tumeurs, et d'autre part à des anomalies développementales du système nerveux ou du squelette. Parmi elles, on retrouve des:

Manifestations tumorales :

- Des tumeurs cérébrales (gliomes du tronc cérébral, méningiomes), infiltrantes et/ou diffuses peuvent survenir. Leur nature et pronostic sont variables, mais sont en général peu évolutifs et parfois asymptomatiques. Même si leur agressivité est modérée, le pronostic vital est parfois en jeu à cause de leur mauvaise localisation.
- Les neurofibromes sont des tumeurs bénignes pouvant se former sur une ou plusieurs racines nerveuses de la moelle épinière. Les neurofibromes nodulaires, profonds et sous-cutanés sont susceptibles de comprimer les nerfs à partir desquels ils ont pris racine. Cette compression peut engendrer des neuropathies, des douleurs (en éclairs lors de leur pression, ou avec irradiation) ou des déficits sensitifs voire moteurs. Les neurofibromes intraduraux sont par ailleurs fréquents au niveau de la queue de cheval.
- Les tumeurs malignes des gaines nerveuses (*Malignant Peripheral Nerve Sheath Tumors* ou MPNST) forment la principale cause de mortalité de la NF1 à l'âge adulte (très rares avant 10 ans). Ils se développent à partir de neurofibromes nodulaires ou plexiformes, mais jamais à partir des atteintes cutanées. Elles sont de mauvais pronostic et les changements de taille, de morphologie, ou des retentissements neurologiques et sensitifs doivent conduire à une exploration immédiate des neurofibromes. L'exérèse chirurgicale précoce, complète et large, est le seul moyen d'espérer une évolution favorable.

Manifestations non tumorales :

- L'hydrocéphalie, qui atteint environ 2% des patients, est la conséquence d'une sténose de l'aqueduc de Sylvius, secondaire ou non à des tumeurs. De type non communicant, elle peut engendrer une hypertension intracrânienne aiguë avec céphalées, vomissements et troubles de la conscience. Les céphalées sont fréquentes chez un grand nombre de malades.

- Des OBNI (objets brillants non identifiés) sont mis en évidence à l'IRM chez 50 à 70% des enfants. Ce sont des images bien délimitées, hyperintenses (hypersignaux en T2), correspondant à des hamartomes. Leur signification est incertaine mais ils sont actuellement considérés comme n'ayant pas de lien avec des troubles cognitifs³⁵. Ces anomalies disparaissent vers l'âge adulte. Cependant, des difficultés d'apprentissage et des troubles cognitifs sont fréquents, on les retrouve chez la moitié des enfants³⁶. Cette légère déficience est faible et ne se discrimine pas avec ce qui est observé dans la population générale. Ces troubles influent majoritairement la coordination motrice, le repérage dans le temps et l'espace (troubles visuospatiaux) ainsi que la mémoire. Chez certains enfants, la scolarité peut être considérablement altérée, avec une élocution parfois difficile qui rend compliqué le dialogue avec les enseignants.

Les manifestations endocriniennes :

La NF1 augmente le risque d'apparition des phéochromocytomes (le risque relatif est multiplié par 1000 par rapport à la population générale). Ils surviennent dans environ 5% des cas à l'âge adulte. L'atteinte chez l'enfant reste exceptionnelle.

Cette tumeur peut s'accompagner de signes liés à l'hypersécrétion adrénérergique: hypertension permanente mêlée à des crises hypertensives, des palpitations, des céphalées.

On peut aussi observer des anomalies de la puberté : précocité pubertaire (parfois associée à un gliome du chiasma³⁷), ou un retard pubertaire dans environ 2% des cas.

Manifestations vasculaires :

L'atteinte des vaisseaux est générale même si elle est peu symptomatique. L'hypertension artérielle est fréquente et généralement secondaire à un rétrécissement de l'artère rénale, avec pour impact une diminution de l'irrigation du rein et une stimulation du système Rénine-Angiotensine-Aldostérone.

Les conséquences des vasculopathies sont rares mais peuvent être gravissimes, à type d'occlusion artérielle, formation de fistules artério-veineuses, rupture d'anévrisme cérébrale, toutes pouvant être fatales³⁸.

Manifestations malignes :

En accord avec le caractère de gène suppresseur de tumeur du gène *NF1*, la neurofibromatose prédispose à la survenue de tumeurs, pour la plupart bénignes. Il convient de noter que la surmortalité liée au cancer reste faible chez l'adulte NF1. Elle est cependant plus marquée chez l'enfant. Certaines tumeurs, plus rares, lui sont propres et ont une évolution maligne plus fréquente :

- Les gliomes des voies optiques, rare dans la population générale, voient leur fréquence particulièrement augmentée chez les patients NF1 (15 % des cas).
- Les MPNST (anciennement appelés neurofibrosarcomes), sont la principale cause de décès chez le patient NF1. L'âge médian au diagnostic est de 25 à 30 ans. Ils apparaissent à partir de neurofibrome (nodulaire isolé, plexiforme). Cette complication majeure survient chez environ 10% des patients³⁹. La survie à 5 ans est faible (de 30 à 50%) et dépend fortement de la précocité du diagnostic et de leur résecabilité.

D'autres cancers, rares dans la NF1, ont une prévalence accrue par rapport à la population générale : les glioblastomes, les leucémies (la leucémie myéloïde juvénile chronique, même si exceptionnelle, est souvent associée aux xanthogranulomes juvéniles, qui touchent 1 % des patients atteints), les rhabdomyosarcomes, les phéochromocytomes malins, les neuroblastomes.

Une étude menée récemment sur une large population a montré que le risque pour une femme de développer un cancer du sein avant l'âge de 50 ans est 5 fois plus élevé chez les patientes NF1 que dans la population générale⁴⁰.

Le tableau 3 regroupe les différents types de cancers rencontrés chez les patients NF1, et le risque associé.

| Type de cancer | Risque de développer un cancer pendant sa vie |
|---|---|
| Gliomes des voies optiques | 15 - 20 % |
| Autres tumeurs cérébrales | Plus de 5 fois la normale |
| MPNST | 8 - 13% |
| Tumeurs stromales gastrointestinales | 4 - 25% |
| Cancer du sein | 5 fois la normale avant 50 ans |
| Leucémies | 7 fois la normale |
| Phéochromocytomes | 1 - 5.7% |
| tumeurs carcinoïdes digestives | 1% |
| Rhabdomyosarcomes | 1,4 - 6% |

Tableau 3. Les différents tumeurs les plus fréquentes et leur risque associé³⁹.

2.5. DIAGNOSTIC

La fréquence des signes cliniques listés par le NIH sont reportées dans le tableau 2, avec leur âge d'apparition moyen.

| | Fréquence (%) | Âge |
|--------------------------------|---------------|----------------------|
| Peau | | |
| TCL | 99 à 100 | < 5 ans |
| « éphélides » | 50 à 80 | < 6 ans |
| Neurofibromes cutanés | 100 | < 30 ans |
| Neurofibromes nodulaires | 15 à 24 | enfance, adulte |
| Neurofibromes plexiformes | 30 à 39 | < 5 ans |
| Œil | | |
| Nodules de Lisch | 67 à 82 | > 6 ans |
| Gliome optique | 15 | petite enfance |
| Gliomes symptomatiques | 1 à 3 | |
| Squelette | | |
| Dysplasie sphénoïde | > 1 | enfance |
| Pseudarthrose | 3 à 4 | enfance |
| Scoliose nécessitant chirurgie | 2 à 4,4 | enfance, adolescence |

Tableau 4. Fréquence des différents critères diagnostiques du NIH et âge moyen d'apparition²⁸.

Si le diagnostic est parfois impossible dans la petite enfance, à cause d'une absence de signes cliniques et à une absence d'antécédents familiaux (50% des cas *de novo*) fréquents, le diagnostic est en générale facile pour les adultes NF1. Le diagnostic moléculaire reste exceptionnellement nécessaire.

2.5.1. ARBRE DECISIONNEL

L'arbre décisionnel qui en ressort pour le praticien est simple (figure 13) :

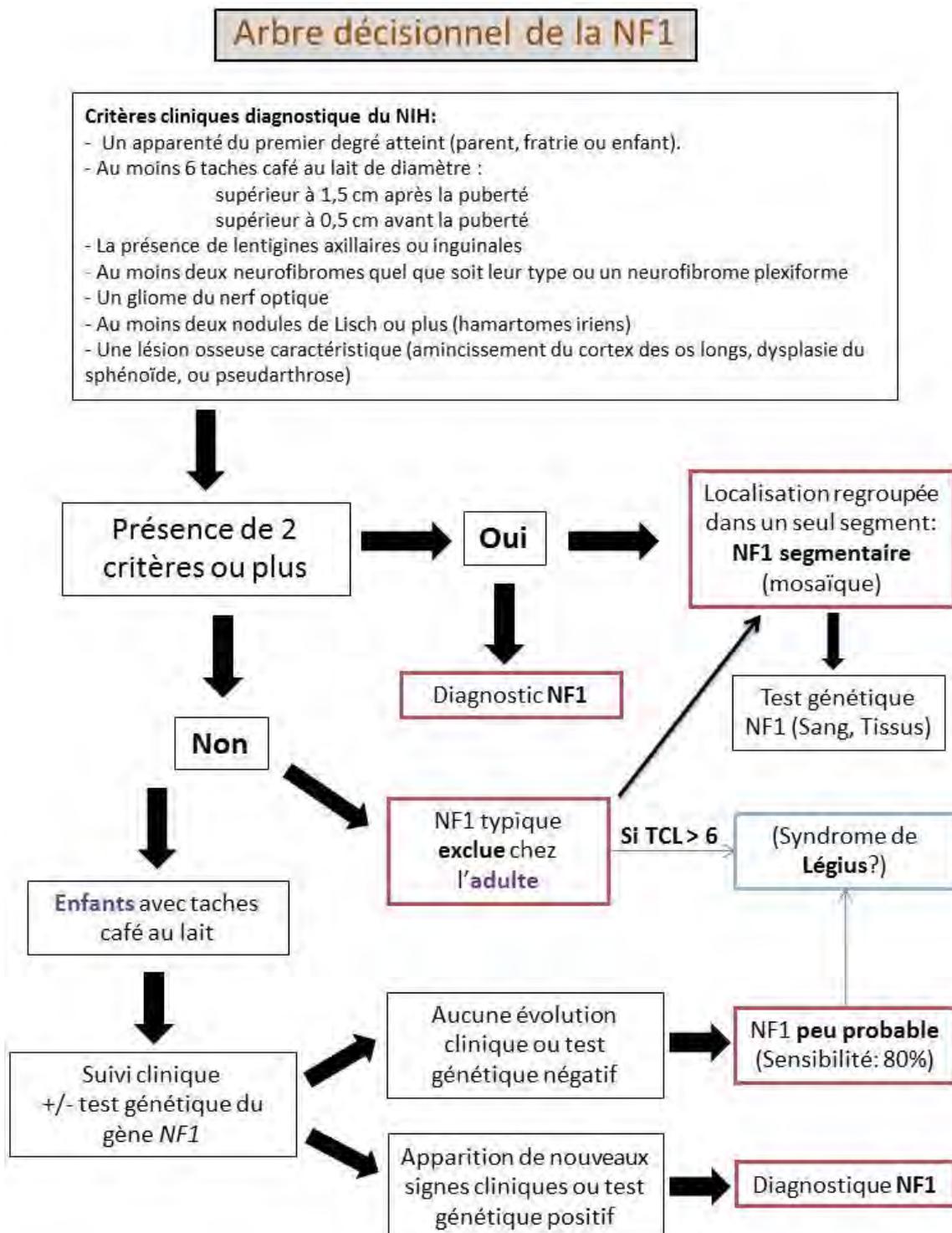


Figure 13. Arbre décisionnel de diagnostic de la NF1⁴¹.

Devant la présence de TCL, le diagnostic de NF1 est souvent le premier à envisager. Cependant, certains diagnostics différentiels doivent être évoqués afin d'éviter toute erreur et mauvaise orientation diagnostique.

Plus d'une centaine de maladies génétiques ont été décrites avec dans leurs signes cliniques des TCL, mais seul un petit nombre de pathologies peut être confondu avec la NF1.

Des syndromes comme ceux de Noonan, McCune Albright, LEOPARD, ont en commun seulement les tâches cutanées. Pour d'autres, les points d'appel tumoraux ou hypertrophiques peuvent tromper le médecin. Les pathologies qui peuvent potentiellement être considérées sont regroupées dans le tableau 5.

Il faut aussi différencier la NF1 des autres types de neurofibromatoses, classés initialement en 8 différentes pathologies par Riccardi⁴² en 1982, mais au sein desquelles seul la NF2 est intéressante dans notre étude:

- La neurofibromatose de type 2 a longtemps été confondue avec NF1 en une seule entité clinique. Elles sont pourtant différentes sur le plan clinique et génétique. La distinction des tumeurs propre à cette classe est essentielle. Les neurofibromes de la NF1 sont des tumeurs développées à partir de cellules conjonctives alors que les tumeurs de la NF2 sont des schwannomatoses, issues de cellules de Schwann qui sont des cellules gliales entourant les nerfs. Les gènes mis en cause dans la NF2 ne sont pas les mêmes et se situent sur différents chromosomes. La NF2 est dite « acoustique » car les schwannomes affectent préférentiellement sur les nerfs vestibulaires. La NF2 ne développe pas de TCL.
- Les autres formes NF3 à NF8 sont des formes frustres, ou segmentaires (NF5), dans lesquels les signes de la NF1 sont en parties présents (NF3, NF4), ou apparaissent tardivement à l'âge adulte (NF7). Ces différentes distinctions sont peu utilisées, sauf pour la NF2, qui est une pathologie bien délimitée.

Diagnostic différentiel de la NF1

| Syndromes | Signes cliniques associés |
|--|---|
| Autres formes de neurofibromatoses | |
| NF1 segmentaire ou en mosaïque | voir la clinique (section 2.5.3) |
| Neurofibromatoses de type 2 | Schwannomes vestibulaires bilatéraux, des nerfs crâniens, périphériques, des nerfs cutanés; méningiomes du système nerveux central, gliomes et cataractes juvéniles |
| Syndromes avec atteintes pigmentaires | |
| Syndrome de Légius | Voir la clinique (chapitre 3) |
| Syndrome de McCune-Albright | TCL, dysplasie fibreuse polyostotique |
| Syndrome LEOPARD | lentiginos multiples, hypertélorisme oculaire, surdit , atteintes cardiaques cong nitales |
| Syndromes hypertrophiques | |
| Syndrome angio-ost o-hypertrophique | h mangiomes cutan s, varices, h mi-hypertrophie |
| Syndrome de Prot e | Hyperostoses, hamartomes hypertrophiques, n vus  pidermiques |
| Syndromes tumoraux | |
| Lipomatose | Affecte le tronc ou les membres |
| Syndrome de Bannayan Riley Ruvalcaba | Lipomes multiples, h mangiomes, t ches pigment es |
| Fibromatoses | Tumeurs multiples des muscles, de la peau, des os |
| N oplasie endocrinienne multiple type 2 | Ph ochromocytomes, carcinome m dullaire de la thyro de, ganglioneuromatose du tractus digestif, morphologie marfano de |

Tableau 5. Diagnostic diff rentiel de la NF1.

2.5.2. LES FORMES MOSAÏQUES, UN DIAGNOSTIC PARTICULIER

Le mosaïcisme est un état dans lequel plusieurs populations de cellules (2 au minimum), porteuses de variants génétiques différentes, coexistent chez un même individu. La répartition et le nombre de cellules touchées dépendent du stade de développement embryologique au cours duquel l'anomalie est apparue. Les cellules filles issues de la cellule ayant acquis la mutation se propageront dans des lignées de cellules anormales, les cellules saines et leurs lignées ne seront pas porteuses du variant délétère.

Initialement décrite en 1956 sous le nom de neurofibromatose segmentaire ou de type 5, elle a été incluse par Riccardi dans sa classification des neurofibromatoses⁴². Elle a souvent une localisation limitée. Sa définition est donnée par les critères suivants : des TCL ou neurofibromes touchant un seul segment, sans histoire familiale connue, et sans signes systémiques.

Les formes en mosaïque donnent la plupart du temps un tableau clinique limité. Cependant, si l'anomalie s'est produite tôt dans l'embryogénèse, la forme clinique sera très proche d'une NF1 non mosaïque, car l'anomalie sera commune à une grande proportion de cellules. Dans ce cas de figure, la mosaïque passe souvent inaperçue et le patient sera probablement diagnostiqué comme ayant une mutation constitutionnelle. C'est le taux de mutations dans les cellules sanguines (leucocytes) qui déterminera la capacité ou non à détecter une anomalie.

Malgré les signes très localisés, les patients sont tout de même porteurs d'une anomalie du gène *NF1*. La descendance de ces patients ayant la même anomalie peut présenter un tableau clinique classique de la NF1 et donc différents du parent atteint⁴³.

Cette forme reste extrêmement rare et sa prévalence est estimée à 1/40 000 à 1/50 000⁴⁴.

L'intérêt de dépister ce genre de cas particuliers réside dans le fait que le conseil génétique peut être différent, pour le patient mais surtout pour les apparentés. En effet, le taux de mosaïcisme germlinal chez un des parents du sujet atteint d'une NF1 constitutionnelle détermine le risque de récurrence dans la fratrie de ce sujet. Une étude sur l'ADN spermatocytaire a démontré qu'un taux de 10% suffit à une récurrence⁴⁵.

2.5.3. LE DIAGNOSTIC GENETIQUE

Le diagnostic de la NF1 est principalement clinique. Les critères établis par la conférence de consensus de Bethesda suffisent dans la grande majorité des cas pour porter le diagnostic. L'analyse génétique n'influençant pas le suivi des patients, elle n'est donc pas proposée dans la plupart des cas. Cependant, lorsque les critères ne sont pas réunis, elle est indiquée dans plusieurs situations. L'utilité du diagnostic génétique est réelle dans le cadre d'une demande de:

- Diagnostic précoce chez les enfants présentant une clinique **incomplète**
- Diagnostic des formes **atypiques**, ou partielles dans le cas de **mosaïcisme**.
- Diagnostic **prénatal** ou **préimplantatoire** chez des patients porteurs d'une anomalie du gène NF1.

2.6. PRONOSTIC ET SUIVI

La morbidité et la mortalité liées à la NF1 résultent de la survenue des complications multisystémiques (tableau 6). L'espérance de vie des patients NF1 est inférieure de 10 à 15 ans à celle de la population générale⁴⁶, tandis que le risque de cancer (tumeur bénigne ou maligne) est augmenté de 2,5 à 5 fois par rapport la population générale⁴⁷. Les tumeurs malignes sont rares mais potentiellement de mauvais pronostic.

| Complications | % | Âge |
|--------------------------------|-----------|----------------------|
| Peau | | |
| Neurofibromes plexiformes | 30 à 39 | < 5 ans |
| Xanthogranulome juvénile | 1 à 2 | enfance |
| Œil | | |
| Gliome des voies optiques | 15 | petite enfance |
| Symptomatiques | 2 à 4 | |
| Squelette | | |
| Pseudarthrose | 3 à 4 | enfance |
| Scoliose nécessitant chirurgie | 2 à 4,4 | enfance, adolescence |
| Système nerveux | | |
| Difficultés d'apprentissage | 33 à 70 | enfance |
| Épilepsie | 6 à 7 | enfance |
| Hydrocéphalie | 1,5 à 2,6 | enfance |
| Cancers | | |
| Neurofibrosarcome | 3 à 4 | adolescence, adulte |
| Leucémie | < 0,1 | |
| Tumeur carcinoïde | 0,6 à 1,5 | |
| HTA | | |
| Phéochromocytome | < 1 | adulte |
| Sténose de l'artère rénale | 1 | enfance, adolescence |

Tableau 6. Complications justifiant le suivi des malades NF1 en fonction de l'âge²⁹.

Le suivi des patients est important, surtout pendant l'enfance, âge durant lequel de nombreuses complications à mauvais pronostic surviennent.

La détection précoce des complications est nécessaire, car le pronostic vital peut être en jeu lors d'atteintes de certaines localisations (neurologiques en particulier) et par les transformations malignes. Le pronostic fonctionnel est aussi dépendant de la localisation des lésions et de leur type. Aucun signe prédictif de l'évolution n'existe à ce jour.

L'expressivité et la gravité de la NF1 sont variables pour une même mutation, d'un sujet à un autre et au sein d'une même famille. Ainsi, chaque patient doit être suivi individuellement, sans tenir compte de l'évolution des autres porteurs de la mutation héréditaire.

Même pour les formes mineures, pour lesquelles la demande prioritaire du patient porte sur les atteintes cutanées, un suivi strict et régulier doit être effectué. Les principaux critères à évaluer, développés par le réseau NF-France, sont surtout cliniques, (résumés dans le tableau 7).

Les investigations systématiques paracliniques (biologique avec dosage des catécholamines ou IRM par exemple) ne sont pas nécessaires et les examens complémentaires sont proposés sur point d'appel clinique (apparition ou modifications de signes cliniques sur une courte durée). Seul examen systématique, l'examen ophtalmologique avec champs visuels est régulier jusqu'à l'âge de 12 ans.

La surveillance est fonction de l'âge : annuelle chez les enfants, elle devient plus espacée chez l'adulte, et fait appel à de nombreux spécialistes en cas de complications.

| A) Enfant |
|--|
| <p><i>Évaluation initiale</i></p> <p>Interrogatoire et examen de la famille (arbre généalogique, recherche de sujets atteints)</p> <p>Examen clinique complet :</p> <p>Poids, taille, périmètre crânien, TA</p> <p>Examen neurologique (mouvements anormaux)</p> <p>Examen dermatologique : taches café au lait, éphélides, neurofibromes sous-cutanés, neurofibromes plexiformes (taille, évolutivité).</p> <p>Examen orthopédique : cypho-scoliose</p> <p>Examen endocrinologique : anomalies pubertaires,</p> <p>Examen abdominal</p> <p>Examen ophtalmologique (acuité visuelle, lampe à fente, fond d'œil, champ visuel)</p> <p>Évaluation des troubles de l'apprentissage</p> <p>IRM cérébrale systématique avant 6 ans (dépistage du gliome optique).</p> <p>Examens complémentaires si anomalie clinique</p> <p>Information de la famille (pronostic, conseil génétique, éducation, résultats d'éventuels examens complémentaires)</p> <hr/> <p><i>Évaluation annuelle systématique</i></p> <p>Examen clinique complet (cf. supra)</p> <p>Recherche de signes de complications selon l'âge du patient</p> <p>Examen ophtalmologique.</p> <p>Évaluation des troubles de l'apprentissage</p> |
| B) Adulte |
| <p><i>Interrogatoire</i></p> <p>Antécédents familiaux et enquête familiale (examen initial)</p> <p>Céphalées</p> <p>Gêne esthétique</p> <p>Douleurs</p> <hr/> <p><i>Examen clinique</i></p> <p>Recherche des critères diagnostiques de neurofibromatose de type 1 (examen initial)</p> <p>Mesure de la pression artérielle</p> <p>Existence de neurofibromes plexiformes (taille, évolutivité)</p> <p>Neurofibromes cutanés gênant esthétiquement</p> <p>Existence de neurofibromes sous-cutanés (douleurs, taille, évolutivité)</p> <p>Existence d'une masse évolutive (cutanée ou intra-abdominale)</p> <p>Examen neurologique</p> <hr/> <p><i>Dialogue</i></p> <p>Explication répétée de la maladie, notamment en cas de prescription d'examens complémentaires</p> <p>Réponses aux questions</p> <p>Explications du résultat d'éventuels examens complémentaires</p> <p>Conseil génétique</p> <p>Visite tous les un à deux ans conseillée</p> |

Tableau 7. Critères d'évaluation initiale et de suivi d'un patient NF1²⁹

2.7. CONSEIL GENETIQUE

Définition : Le conseil génétique est un moment de rencontre entre le clinicien et les personnes touchées par une pathologie. Il se fait avec le « cas index », ou première personne d'une famille ayant été diagnostiquée comme atteinte de la maladie en question. Le conseil génétique prend aussi en considération les autres membres de la famille, même si il est préférable de voir la famille dans un 2^{ème} temps après le cas index.

Le conseil génétique est un acte médical qui permet de donner des informations sur la nature, la transmission, et les implications du désordre génétique aux patients, ainsi qu'une éducation sur les mesures à prendre afin de limiter les complications. Un échange avec le patient permet de lui donner le plus d'informations possibles afin de prendre des décisions, son désir de procréer par exemple, ou son souhait d'échanges sur la maladie avec ses proches, et permet de clarifier toutes questions qui resteraient en suspens.

La NF1 est une maladie se transmettant sur un mode autosomique dominant. La probabilité de transmettre l'allèle muté par un individu atteint à son enfant est de 50%, quel que soit son sexe. Dans la moitié des cas l'anomalie du gène NF1 apparaît *de novo*, c'est-à-dire qu'aucun des parents n'a transmis l'anomalie (apparition « à priori » sporadique par néomutation), ce qui complique le conseil génétique.

Le généticien, par des consultations successives, essaie :

- De réunir tous les éléments généalogiques, cliniques, et para-cliniques afin de poser le diagnostic avec certitude.
- D'informer le patient, en fonction de son âge, de ses capacités, et de ses demandes, sur sa maladie, sa transmission, ses évolutions et complications, et enfin la prise en charge qui en découle.
- De statuer si des tests génétiques sont nécessaires.
- De définir les sujets à risque au sein de la famille, et les sujets non à rissque sans nécessité de suivi.

Le conseil génétique doit être réalisé pour tous les couples concernés ayant un projet de procréation.

Lorsqu'une NF1 est suspectée, les parents du patient sont amenés à procéder à un examen médical approfondi à la recherche de signes cliniques évocateurs. Si des signes sont trouvés chez

eux, la puissance du diagnostic chez l'enfant devient plus importante. Si aucun des parents ne présente de signes clinique et n'est porteur de mutation, le risque d'avoir un second enfant atteint de NF1 reste supérieur à celui de la population générale⁴⁸, dû à la possibilité, exceptionnelle, d'avoir chez les parents une atteinte en mosaïque.

Le diagnostic prénatal est possible, en extrayant l'ADN des cellules fœtal obtenues par amniocentèse ou ponction des villosités chorales. Un diagnostic préimplantatoire est aussi possible⁴⁹ pour les couples à risque ayant pour projet d'avoir un enfant. Dans ce cadre-là, le conseil génétique tient un rôle important, les parents devant être bien informés du risque de transmission. Le projet doit être discuté en amont d'une grossesse, afin de procéder à l'identification de l'anomalie chez le parent atteint que l'on recherchera par la suite par des tests moléculaires. En France, quatre laboratoires (Paris, Nantes, Strasbourg, Montpellier) sont en charge de ces analyses.

2.8. PRISE EN CHARGE

Aucun traitement spécifique de la maladie n'existe actuellement. Le traitement des complications, le plus précocement possible, nécessite souvent une décision multidisciplinaire qui doit tenir compte des spécificités de la maladie et du patient. Les structures multidisciplinaires sont particulièrement adaptées à la diversité de la prise en charge.

Les troubles de l'apprentissage peuvent demander le concours d'orthophonistes ou psychomotriciens. De plus, il est important que l'enfant bénéficie d'une attention particulière et une aide adaptée à son cas par les enseignants, à partir d'une évaluation précoce de l'atteinte neurologique. Les médicaments classiquement utilisés par la population générale dans le traitement du déficit d'attention lié à une hyperactivité, de la dépression, de l'anxiété peuvent être prescrits chez les patients NF1. Le soutien psychologique joue aussi un rôle primordial afin que le patient accepte sa maladie invalidante.

Les méthodes de rééducation fonctionnelle ou de contention sont importantes pour les complications osseuses. Le traitement des pseudoarthroses reste difficile, et conduit parfois à l'amputation.

Les neurofibromes cutanés peuvent avoir un retentissement esthétique fort et impacter la vie du malade. Afin d'améliorer l'aspect visuel, ils peuvent être supprimés chirurgicalement ou par laser CO₂. Cet aspect du traitement est important pour les patients pour qui les marques défigurantes sont les moins acceptées des manifestations cliniques⁵⁰. Les neurofibromes

plexiformes sont difficiles à traiter chirurgicalement car ils envahissent les tissus nerveux et ont tendance à repousser après l'excision⁵¹.

Les gliomes des voies optiques requièrent un traitement par chimiothérapie, le traitement de 1^{ère} ligne⁵². L'évolution des gliomes est rare et imprévisible. Le pronostic visuel et vital peut être engagé en cas d'évolution rapide nécessitant un traitement chirurgical.

Des études ont montré une augmentation significative du nombre de tumeurs malignes secondaires à des traitements génotoxiques⁵³ (radiothérapie, chimiothérapie). Le traitement tumoral doit donc être spécifique et la prise en charge spécialisée compte tenu du risque de voir une seconde néoplasie apparaître. La radiothérapie n'est plus recommandée pour le traitement des gliomes des voies optiques chez ces individus génétiquement prédisposés aux cancers⁵⁴.

3. LE SYNDROME DE LEGIUS, UNE MALADIE DE DECOUVERTE RECENTE

3.1. INTRODUCTION/ HISTORIQUE

Le syndrome de Légius (OMIM: #611431) est caractérisé par des TCLs multiples sans neurofibrome ni aucune manifestation tumorale telle que retrouvée dans la NF1. Il n'y a pas de mutation du gène *NF1*.

En 2007, H.Brems et al⁵⁵ ont mis en évidence, par analyse de liaison génétique de 5 familles ayant un tableau clinique proche de celui de la NF1, des mutations dans le gène *SPRED1* (OMIM : *609291).

Ce gène mis en cause est porté par le chromosome 15 au niveau de la bande 14 du bras long. Il code pour la protéine SPRED-1 (Sprouty-related, EVH1 domain-containing protein 1). C'est une protéine qui régule négativement la voie de signalisation RAS-MAPK, par un mécanisme proche de la neurofibromine (codée par *NF1*).

Longtemps confondue avec la neurofibromatose de type 1, le tableau clinique NF1-like comprend de multiples TCL, des éphélides axillaires, et un dysmorphisme facial à type Noonan-Like. Environ 200 cas ont été décrits et étudiés jusqu'à présent dans la littérature.

Initialement appelé « NF1-like syndrome » (NFLS), la 13^{ème} réunion d'experts de l'European Neurofibromatosis Meeting de 2008 décida de nommer ce trouble : « syndrome de Légius », reprenant le nom de l'auteur sénior du premier article mettant en cause une anomalie du gène *SPRED1*³⁶.

3.2. CLINIQUE

Le petit nombre de patients jusqu'à présent diagnostiqués et le manque de recul sur cette nouvelle pathologie ne permet probablement pas l'accès au tableau clinique exhaustif du syndrome de Légius. Cependant, même si quelques signes sont trompeurs et se confondent d'avec la NF1, l'absence totale d'autres manifestations typiques de la NF1, permet sa distinction.

Atteinte cutanée :

Pour presque tous les patients (sauf pour 2 patients chez qui aucune atteinte cutanée n'a été retrouvée^{36,56}) :

- De multiples TCL sont présentes avec une répartition homogène sur tout le corps.
- Des éphélides se retrouvent aussi fréquemment au niveau axillaire ou au niveau du visage.

L'apparition de ces signes évolue selon l'âge des patients, et le nombre de TCL augmente avec l'âge, comme pour la NF1.

Macrocéphalie :

Plus ou moins impactant, la macrocéphalie est un signe courant, mais non présent chez tous les patients. Selon les cohortes étudiées, on retrouve une fréquence entre 3 %⁵⁷ et 40%⁵⁵.

Taille :

Les patients atteints du syndrome de Legius sont de petite taille dans la majorité des cas (31% dans la publication de Denayer³⁸ à 50% pour Brems³⁶).

Développement et apprentissage :

Des problèmes de développement sont fréquemment retrouvés chez l'enfant avec retard des apprentissages (langage).

Même si le nombre de patients étudiés reste faible, un certain nombre de cas avec des troubles de l'attention, de l'hyperactivité ou des retards moteurs ont été décrits.

Au final, il est plutôt acquis que le syndrome de Légius est à l'origine des problèmes cognitifs de certains patients, avec baisse du QI (1 cas décrit avec QI de 68⁵⁸) mais que ces atteintes restent moins importantes que pour les cas de NF1.

Autres signes

De nombreuses autres atteintes ont été décrites, parfois chez un unique patient : atteintes vasculaires (hémangiome), lipomes, clinodactylies, hypotonies, céphalées ou encore pectus excavatum (voir tableau 8).

| Signes cliniques | Etudes (avec leurs cohortes) | | | | | | | |
|---|------------------------------|-----------------------------|------------------------------|------------------------------|--|-----------------------------|----------------------------|--|
| | Brems [2007] (n=44) | Pasmant [2009] (n=18) | Spurlock [2009] (n=12) | Messiaen [2009] (n=40) | Muram- Zborovski [2010] (n=2) | Denayer [2011] (n=30) | Spencer [2011] (n=7) | Laycock- v.Spyk [2011] (n=10) |
| Macrocéphalie | 30% | 6% | 8% | 20% | 50% | 3% | | 30% |
| Petite taille | 7% | | | 3% | 0% | 27% | 14% | |
| Faciès Noonan-like | 11% | | | 3% | | 40% | 14% | |
| Anomalies thoraciques (pectus excavatum) | 7% | | | 8% | | 23% | 14% | 10% |
| Macules dépigmentées | 7% | | | 3% | | | | |
| Lésions vasculaires | 2% | | 8% | 5% | | | | |
| Lipomes | 32% | 11% | | 5% | | 7% | 14% | |
| Retards, troubles de l'attention | 18% | 28% | | 23% | 100% | 53% | 57% | 60% |
| Céphalées | | 22% | | | | 7% | | |
| Baisse d'audition | | | | 5% | | 3% | | |
| Convulsions | | | | 5% | | 3% | | 10% |
| Polydactylie | | | | 3% | | 7% | | |
| Scolioses | | | | | | 13% | | 10% |

Tableau 8. Différentes manifestations du syndrome de Légius : nombre de personnes par études chez qui les signes associés sont retrouvés, et leurs fréquences relatives (D.Stevenson et al, GeneReviews®).

Risque tumoral

Très peu de manifestations tumorales ont été décrites chez les patients diagnostiqués, avec en général, aucune des manifestations propre à la NF1 (nodules de Lisch, neurofibromes, MPNST).

Cependant, le cas d'un enfant de 11 mois a été décrit avec une leucémie aiguë myéloblastique⁵⁹. Pour conforter l'idée que le risque de leucémie est plus fort chez les patients atteints du syndrome de Légius, la même équipe (Pasmant et al) a recherché des atteintes du gène *SPRED1* chez 230 patients ayant une leucémie aiguë myéloblastique ou lymphoblastique⁶⁰. Une mutation constitutionnelle avec perte de fonction a été retrouvée chez un patient, associée à une perte d'hétérozygotie dans les cellules blastiques, démontrant bien que *SPRED1* dispose des caractéristiques d'un gène suppresseur de tumeur.

D'autres études devraient aider à connaître plus précisément l'incidence des atteintes hématologiques.

3.3. INCIDENCE / PENETRANCE

La pénétrance de la pathologie est forte, avec pour la plus grande majorité des patients, des signes cutanés isolés, même si l'évolution vers le phénotype complet selon l'âge n'est pas encore bien établie.

Il a été montré dans 2 cas (un homme de 60 ans et un enfant de 2 ans) avec un variant délétère du gène *SPRED1*, l'absence totale de TCL ou éphélides.

L'incidence n'est pas estimée avec certitude, dû à un recul encore trop faible et à un biais diagnostique important à cause de signes cliniques non spécifiques de la pathologie. Cependant, elle doit être faible et inférieure à l'incidence de la NF1.

3.4. CRITERES DIAGNOSTIQUES

La frontière clinique entre le syndrome de Légius et les autres pathologies partageant les mêmes signes cutanés aspécifiques est faible, ce qui complique le diagnostic initial. A l'heure actuelle, aucun signe n'est pathognomonique, et trop peu de patients âgés sont connus

et suivis pour permettre de trouver un élément ou une conséquence clinique caractéristique de la pathologie.

Cependant 3 points sont à prendre en compte dans le diagnostic pour inclure une majorité de patients tout en étant le plus spécifique possible (tableau 9) :

- Des anomalies de la pigmentation, consistant principalement en des TCL, ou des éphélides sont retrouvés.
- Une absence des autres signes non-cutanés de la NF1 (par exemple : les nodules de Lisch, les neurofibromes, les gliomes des nerfs optiques, des dysplasies osseuses etc...).
- Le trait doit être héréditaire, c'est-à-dire qu'un parent du patient doit posséder des signes cutanés (TCL) sans les autres atteintes de la NF1.

Malgré ces aides au diagnostic, la NF1 reste la pathologie la plus fréquemment confondue avec le syndrome de Légitus.

Dans deux études^{37,61}, il a été démontré que 2% des patients qui remplissent les critères cliniques diagnostic de la NF1 ont en fait un variant délétère du gène *SPRED1*. Ainsi, des erreurs peuvent s'opérer principalement quand le diagnostic de la NF1 n'est fait que sur les anomalies de la pigmentation et sa transmission héréditaire (2 des critères du NIH, tableau 9). Ceci est d'autant plus vrai que le patient est jeune, n'ayant pas encore développé tous les signes spécifiques à la NF1, dont l'incidence augmente avec l'âge.

Comme pour NF1, le diagnostic différentiel doit se faire sur un nombre important de pathologies possibles associant les TCL ou éphélides (LEOPARD-syndrome, McCune-Albright syndrome).

| Critères cliniques : 2 minimums sont nécessaires pour poser le diagnostic | Présents chez les patients NF1 | Présents chez les patients Syndrome de Légius |
|---|--------------------------------|---|
| ○ Un apparenté du premier degré atteint (parent, fratrie ou enfant). | X | X |
| ○ Au moins 6 taches café au lait de diamètre : ○ supérieur à 1,5 cm après la puberté ○ supérieur à 0,5 cm avant la puberté. | X | X |
| ○ La présence de lentigines axillaires ou inguinales. | X | X |
| ○ Au moins deux neurofibromes quel que soit leur type ou un neurofibrome plexiforme. | X | Absence |
| ○ Un gliome du nerf optique. | X | |
| ○ Au moins deux nodules de Lisch ou plus (hamartomes iriens). | X | |
| ○ Une lésion osseuse caractéristique (amincissement du cortex des os longs, dysplasie du sphénoïde, ou pseudarthrose). | X | |

Tableau 9. critères du NIH pour la NF1 et confusion possible avec le syndrome de légius

3.5. DIAGNOSTIC GENETIQUE

Il n’y a pas de critères préétablis afin de statuer de l’utilité de ces analyses en systématique. Cependant, avant de demander des tests moléculaires pour un patient, il faut prendre en considération :

- Les données cliniques et l’histoire familiale
- L’âge de l’individu (Plus l’âge est précoce, plus la différenciation avec la NF1 est difficile)
- Les différences de traitement, de suivi des patients NF1 et du syndrome de Légius
- Le poids psychologique du diagnostic sur le patient.

Si le doute subsiste sur un syndrome de Légius, le diagnostic se porte sur la NF1 jusqu’à preuve du contraire.

3.6. SUIVI ET TRAITEMENT

Afin d'établir les complications et les besoins de chaque patient ayant un syndrome de Légius confirmé par l'analyse génétique, la consultation de génétique initiale doit comprendre ces évaluations (intégrées souvent à la consultation NF1 qui peut précéder):

- Le statut développemental et neuropsychiatrique du patient selon son âge
- Une évaluation dermatologique à la recherche de signes cutanés
- Une prise de tension à la recherche d'une hypertension, si le doute avec la NF1 subsiste. L'hypertension n'a jamais été décrite à ce jour pour le syndrome de Légius.

Aucun traitement spécifique n'est pour l'instant étudié pour cette pathologie, de même qu'aucun essai clinique n'est en cours de réalisation (donnée : <https://clinicaltrials.gov/> de novembre 2015).

Le traitement des manifestations neuro-psychiatriques, comme les troubles de l'attention avec hyperactivité par exemple, peut être discuté avec des médicaments ayant fait preuve de leur efficacité.

N'ayant jusqu'à présent que peu de manifestations tumorales ou de complications engageant le pronostic vital, le suivi de ces patients sera moindre que pour les patients NF1. Une consultation de suivi régulière (tous les 3 ans) semble raisonnable.

3.7. CONSEIL GENETIQUE

Le syndrome de Légius se transmet selon un mode autosomique dominant. Pour les enfants d'un patient atteint, le risque d'hériter de l'allèle atteint et donc de développer la maladie est de 50% quel que soit le sexe.

Le risque pour les frères et sœurs d'un patient atteint est dépendant du statut des parents. Même si le taux de mutations *de novo* n'est pas encore précisément déterminé, il semble être assez élevé (6/23 patients avec un variant délétère *de novo* de SPRED1 dans l'étude de Messian et al²⁷).

Si aucun variant n'est retrouvé chez les parents du patient, deux solutions sont possibles, soit la mutation est apparue *de novo*, soit un des parents est porteur d'une mutation en mosaïque et donc non dépistée par l'analyse génétique (aucun cas décrit à ce jour).

Avec une pénétrance quasi-complète de la maladie, la priorité des analyses génétiques doit être donnée aux membres de la famille ayant des signes cliniques du syndrome de Légit, notamment les TCL.

Un examen minutieux et un interrogatoire précis doivent être réalisés par le généticien en charge de la consultation à tous les membres de la fratrie.

Aucun test prénatal n'est réalisé en France pour cette pathologie.

4. ETUDES MOLECULAIRES DES GENES *NF1* ET *SPRED1*

4.1. NEUROFIBROMATOSE DE TYPE 1 : DU GENE A LA PROTEINE

4.1.1. LE GENE *NF1*

Les gènes associés à la prédisposition au cancer sont classés en 3 catégories fonctionnelles :

- les oncogènes
- les gènes suppresseurs de tumeurs
- et les gènes de réparation de l'ADN.

Les gènes suppresseurs de tumeurs sont des régulateurs négatifs de la prolifération cellulaire. Le gène *NF1* est classiquement rangé dans cette catégorie.

4.1.2. CLONAGE POSITIONNEL

Le clonage positionnel consiste dans la caractérisation par des techniques indirectes (analyse de liaison) puis directes (clonage après cartographie d'une région candidate) d'une séquence nucléotidique (gène, séquence régulatrice) responsable d'un trait phénotypique.

Le gène *NF1* (OMIM :* 613113) a été identifié par clonage positionnel grâce à plusieurs travaux en 1990^{62,63} et localisé sur le **brin sens** du chromosome 17 au niveau centromérique, en **17q11.2** (figure 14).

4.1.3. STRUCTURE

Le gène *NF1* est un grand gène du génome humain avec **280 kb** (moyenne \approx 50 kb). Dans la version du génome de référence Hg19 (d'après le « Genome Reference Consortium »), le gène débute au nucléotide g.29421945 et se termine au nucléotide g.29704695.

La séquence codante du transcrit majeur contient **57** exons, auxquels peuvent s'ajouter **4** exons alternatifs selon l'épissage et appelés : 9br, 10a-2, 23a et 48a (figure 14).

Pour la suite de ce travail, il faut distinguer deux nomenclatures pour la nomination des exons du gène NF1. L'une, selon le « NF Consortium », appelle les exons de 1 à 49 en associant à certains exons des lettres (a,b,c) ou des numéros (1, 2). La deuxième et plus récente nomenclature correspond aux critères d'appellation du NCBI (National Center for Biotechnology Information) et nomme les exons selon une suite logique de numéros, allant de 1 à 58 pour le transcrit NM_000267.3 (transcrit de référence, d'après « the Reference Sequence database », du NCBI).

Par habitude et du fait d'un langage communément accepté dans la communauté, ce travail utilisera la nomenclature du « NF Consortium ».

Deux grands introns, l'intron 1 et 27b (1 et 35 dans la nouvelle nomenclature) ont une taille supérieure à 60 kpb.

Dans l'intron 27b mais sur le brin **antisens**, sont situés trois petits gènes⁶⁴ : *EVI2A*, *EVI2B* (Ecotropic Viral Intergration site) et *OMG* (Oligodendrocyte Myelin Glycoprotein).

Les gènes *EVI2A* et *EVI2B* sont les homologues humains des gènes murins *Evi-2A* et *Evi-2* qui codent pour une protéine associée au développement de leucémie chez la souris. Le gène *OMG* produit une glycoprotéine membranaire exprimée dans le système nerveux central durant le processus de myélinisation des cellules nerveuses. Il aurait un rôle sur l'inhibition de la plasticité du système nerveux à travers le récepteur PirB⁶⁵.

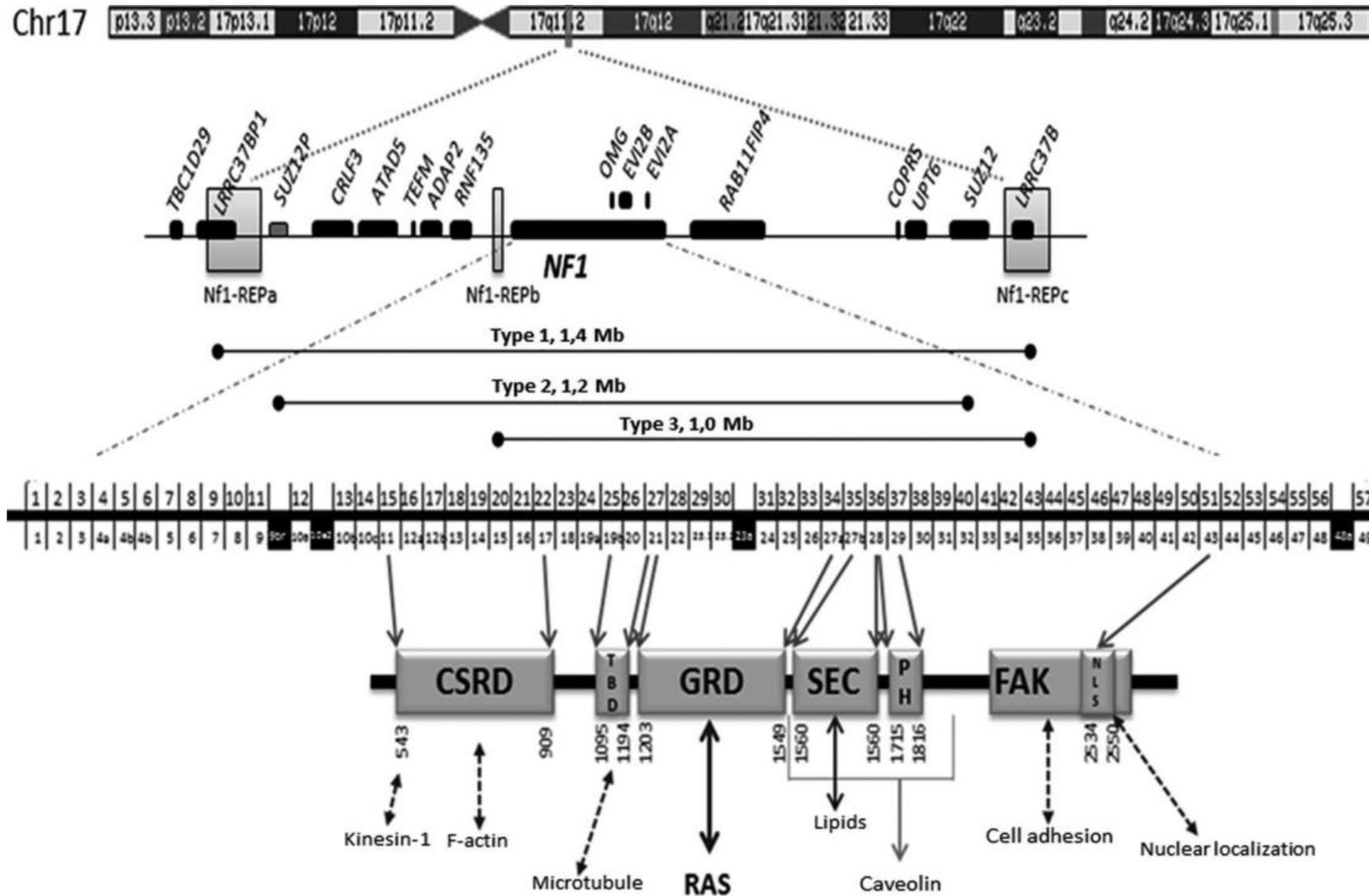


Figure 14. Schéma du gène NF1 et de la protéine neurofibromine (Abramowicz A et al, 2014⁶⁶).

De haut en bas : Localisation du gène *NF1* sur le chromosome 17 au niveau du locus 17q11.2. Visualisation des 3 types de délétion les plus fréquentes.

Schéma du gène et selon la nomenclature NCBI en haut, selon la nomenclature consortium en bas (exons alternatifs en bas).

Tout en dessous : domaines fonctionnels de la protéine et cibles de chacun.

4.1.4. LES PSEUDOGÈNES ASSOCIÉS AU GÈNE NF1

Par définition⁶⁷, un pseudogène désigne une séquence génétique (locus) inactivée (c'est-à-dire non codante) au sein d'un génome par l'accumulation d'altérations génétiques le rendant non fonctionnel.

Certains pseudogènes peuvent avoir une fonction⁶⁸ (résurgence, régulation) par la production d'un ARN. La frontière au fil des années et des connaissances entre gène et pseudogène est devenue moins tranchée.

Les pseudogènes appelés parfois « gènes fossiles » ou à tort « ADN poubelle », dérivent avec le temps par des altérations génétiques multiples, et sont ainsi une conséquence de l'évolution des organismes et de leur adaptation à l'environnement.

On peut classer les pseudogènes en trois types selon leurs mécanismes d'origine (figure 15) :

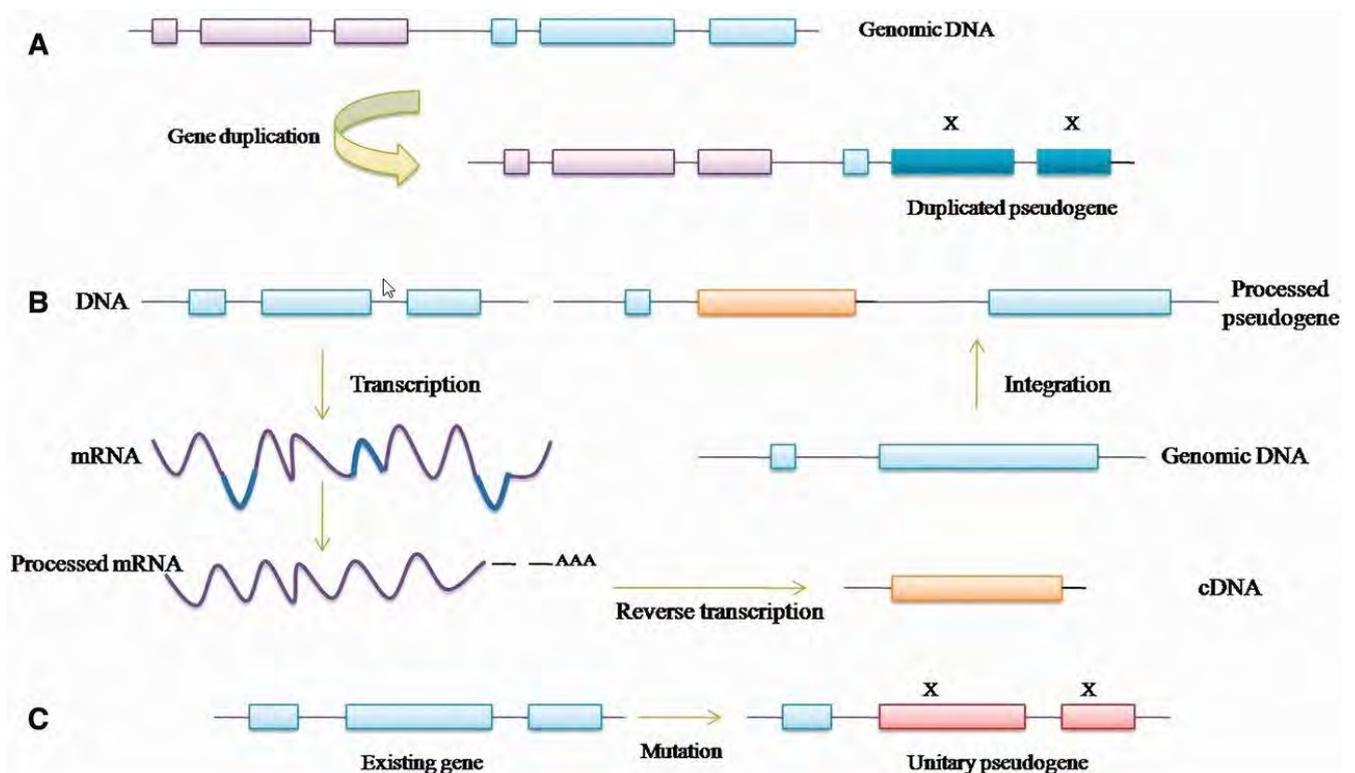


Figure 15. Mécanismes de création de pseudogènes (Kamalika S et al, 2013⁶⁸).

A) Pseudogènes non processés :

« Processé » provient du terme anglais « processed » qui est relatif à la maturation des ARN primaires et en particulier l'épissage. Le pseudogène peut coder un ARN nouvellement transcrit qui n'a pas encore subi l'épissage (« splicing ») et donc possède encore ses introns. Le premier mécanisme de création de pseudogènes correspond à une duplication d'un gène dans le génome, puis une des deux copies acquiert des mutations qui lui font perdre sa fonction d'origine (exons en bleus noté X) et ainsi devenir un pseudogène.

B) Pseudogènes processés :

Ce sont des séquences provenant de la transcription inverse d'ARNm (donc processés) par l'utilisation d'enzymes (transcriptase inverse) associées aux rétrotransposons. L'ADNc nouvellement synthétisé est inséré aléatoirement dans le génome. La majorité de ces séquences n'ont pas de fonction codante en terme protéique car elles perdent les éléments nécessaires à leur fonctionnalité, comme les séquences promotrices, mais jouent un rôle fondamental dans l'évolution et la plasticité du génome.

C) Pseudogènes unitaires :

Ce sont des séquences unitaires qui n'ont aucun paralogue fonctionnel. Au départ ces gènes ont des fonctions bien définies mais deviennent pseudogènes par accumulation à travers l'évolution d'altérations géniques délétères. Cette pseudogénisation pourrait être le fruit d'une pression de sélection ou de processus aléatoires.

D) Non notés sur la figure 15, certains pseudogènes proviennent d'autres espèces (virus, bactéries) et sont insérés dans le génome humain par transfert horizontal.

Une des bases de données les plus sérieuses pour les pseudogènes est celle portée par le projet GENCODE⁶⁹, grâce à l'expertise d'une équipe nommée HAVANA (Human and Vertebrate Analysis and Annotation) du Wellcome Trust Sanger Institute. Elle annote manuellement les prédictions *in silico* des pipelines Pseudopipe et Retrofinder (<http://pseudogene.org/Human/>).

D'après la dernière version disponible de Pseudopipe (build 79, GRCh38), **18158** régions pseudogéniques sont répertoriées à ce jour

Pour notre travail, nous avons utilisé la dernière version annotée sur Hg19 (Tableau des coordonnées et des noms en annexe 1). Elle nous a permis de regrouper les coordonnées des régions pseudogéniques de *NF1* dans un seul fichier. Huit régions pseudogéniques (notées NF1P1 à NF1P8) sont réparties sur les chromosomes 2, 14, 15, 18, 21 et 22. Toutes ces séquences sont issues d'un mécanisme de duplication

4.1.5. ARN MESSENGER

L'ARNm issu du transcrit de référence du gène *NF1* a une longueur de 11 à 13 kbp avec un ORF (open reading frame) de 8457 pb et 3.5 kb de région 3'UTR (« Untranslated region » ou région non traduite).

Selon les tissus, l'expression peut être différente et les transcrits varient en taille due à un épissage alternatif du pre-ARNm.

L'ARNm le plus commun est le transcrit NM_000267 qui a une longueur de 8.5 kb et contient 57 exons. Il est traduit en une protéine de 2818 acides aminés⁷⁰.

4.1.6. LES TRANSCRITS D'ÉPISSAGE

De nombreux transcrits ont été identifiés, avec l'inclusion ou non d'exons alternatifs qui n'altèrent pas l'ORF du gène *NF1*.

Le tableau 10 résume les différents exons alternatifs possibles et quelques-unes de leurs caractéristiques.

| Nom | Situé entre les exons | taille (AA) | Transcrit | action / rôle | Réf |
|--------------|-----------------------|-------------|--|---|-----|
| 9br | 9 et 10a | 10 AA | Pas de nom | N'affecte pas la fonction du domaine GRD. Exprimé fortement dans le SNC. | 71 |
| 10a-2 | 10a et 10b | 15 AA | Exclu : I Inclus : II | Ajoute un segment transmembranaire qui est normalement absent. localisé dans le domaine GRD. | 72 |
| 23a | 23.2 et 24 | 21 AA | Inclus : IV | Différentes localisations d'expression selon l'inclusion ou non | 73 |
| 48a | 48 et 49 | 18 AA | Inclus : III et IV | Très exprimé dans le fœtus, les tissus cardiaque et musculaire squelettique, permettrait le développement et la différenciation de ces tissus. N'affecte pas le domaine GRD. | 74 |

Tableau 10. Exon alternatifs du gène *NF1*.

Point sur l'exon alternatif 23a :

Les deux premières isoformes de *NF1*, appelées type I et II ont pour différence la présence ou non de l'exon 23a. Il est exclu (« skipped ») dans l'isoforme de type I et inclus dans l'isoforme de type II. Cet exon alternatif est localisé dans le domaine GRD (GAP-related domain) de la protéine, changeant les capacités de la neurofibromine à contrôler le signal de la protéine Ras.

Cet épissage particulier est tissu-dépendant. En effet, l'exon 23a est majoritairement inclus dans la plupart des tissus mais il est exclu dans les neurones du système nerveux central.

Bien que l'isoforme incluant l'exon 23a possède toujours des fonctions de Ras-GAP, il a été démontré par des études *in vitro* que son activité de régulation du signal Ras est dix fois plus faible comparée à celle de l'isoforme dont l'exon 23a est exclu. Cette baisse d'activité peut être la conséquence de l'ajout d'acides aminés basiques qui changent de manière significative la structure tertiaire du domaine GRD⁷⁵.

Des recherches sur des modèles murins dans lesquels l'exon 23a n'apparaît dans aucun tissu (*NF1*^{23a-/23a-}) ont montré que ces souris étaient viables et se développaient normalement. De plus, aucune tumeur spécifique ne se développait mais des différences dans l'apprentissage et la mémoire sont apparus.

Dans une étude récente⁷³ il est discuté du rôle de potentiels « gènes modificateurs » qui réguleraient ce transcrite et expliqueraient la variabilité phénotypique de la *NF1*.

A ce jour, il est acquis que la balance d'expression transcrite de type I / transcrite de type II est finement régulée et participe au bon fonctionnement neuronal.

D'autres transcrits naturels :

L'épissage naturel des exons 29 ou 30 ou les deux simultanément, peut aussi se produire mais au prix d'une altération du cadre de lecture⁷⁶. Ce décalage entraîne la création rapide d'un codon stop et la protéine produite est tronquée avant l'incorporation de son domaine GRD. Les fonctions de ces exons alternatifs n'ont pas été à ce jour déterminés.

Edition de l'ARNm (« Editing RNA »)

En dehors des mécanismes d'épissage classique, les transcrits du gène *NF1* peuvent subir une édition par des protéines formant un complexe appelé « éditosome » (APOBEC-1 par exemple). Par déamination, la cytosine localisé à la position c.3916 est transformée en uracile (c.3916C>U). Ceci génère un codon STOP prématuré (CGA → UGA) et la protéine produite après traduction ne contient pas le domaine GRD⁷⁷.

De plus, ce transcrit est instable et se dégrade rapidement par le mécanisme de «Nonsense-mediated decay » (dégradation des ARNm non-sens).

Cette édition entraine une expression non équivalente des différents allèles de *NF1* et une augmentation significative de ce mécanisme a été retrouvée dans les tumeurs, soulignant un possible rôle dans la tumorigénèse⁷⁸.

4.1.7. LA PROTEINE NEUROFIBROMINE

Le gène *NF1* code pour une protéine appelée **neurofibromine**, qui a un poids moléculaire estimé de 327 kDa. Elle est composée de 2818 acides aminés (pour le transcrit principal NM_000267.3).

Expression :

La protéine est exprimée de manière ubiquitaire. Cependant, elle plus fortement dans le système nerveux central, précisément dans les neurones, les astrocytes, oligodendrocytes et les cellules de Schwann⁷⁹ (transcrits de type I et II).

Les transcrits de type III (exon 48a inclus) et de type IV (exon 23a et 48a inclus) sont préférentiellement exprimés dans les muscles cardiaque et squelettique et semblent être essentiel à leur bon développement⁸⁰.

Fonctions / domaines fonctionnels :

La neurofibromine est une protéine multifonctionnelle qui possède plusieurs domaines fonctionnels (figure 14). Alors que toutes les fonctions ne sont pas élucidées, certaines sont bien définies et connues :

- *Activité GTPase*

L'activité primaire la mieux définie est l'activité « Guanosine-trisphosphate(GTP)ase activation protein » ou GAP.

Pour rappel, cette activité est liée à la région centrale de la protéine comprise entre les exons 20 et 27a de 360 acides aminés et connue sous le nom de GAP-related domain (GRD).

Cette fonction GAP permet d'inactiver la protéine Ras en accélérant la conversion (constitutive de Ras) de la forme active Ras-GTP à la forme inactive Ras-GDP (figure 16).

Cette régulation négative de l'expression de l'oncogène *Ras* par la neurofibromine prévient l'activation en aval de la voie MAPKinase et de la voie PI3K/Akt/mTOR, limitant la prolifération et la différenciation des cellules.

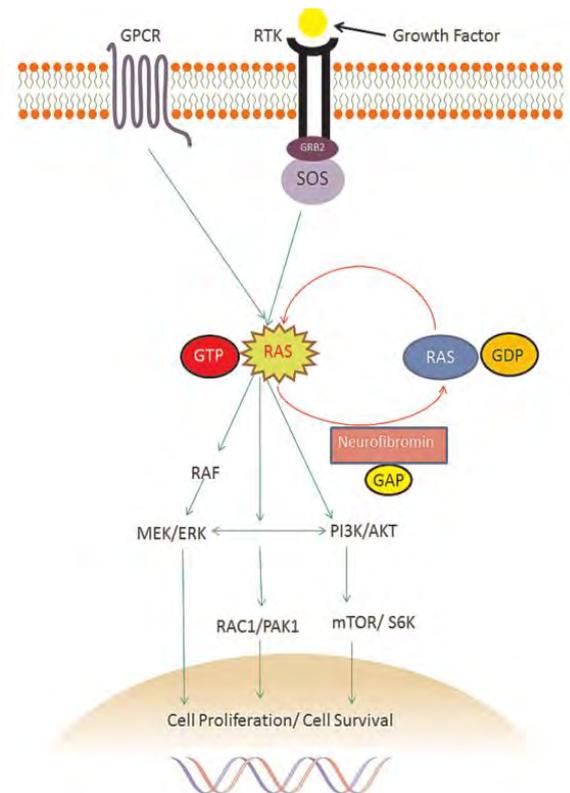


Figure 16. Rôle de la neurofibromine dans la voie de signalisation Ras.

- *Activation par la protein kinase C (PKC)*

La fonction Ras-GAP de la neurofibromine est activée par la phosphorylation de son domaine « cystein-serine rich » (CSRD) qui est porté par les exons 11 à 17⁸¹.

Il est à noter qu'un cluster de mutations peut exister à cet endroit⁸², ce qui démontre l'importance de ce domaine, et de la phosphorylation de la PKC qui permet de maintenir sa fonction normale.

- *Liaison à la Caveolin-1*

Le gène *CAV-1* qui code pour la protéine Caveolin-1, est un potentiel gène suppresseur de tumeur qui régule négativement la voie Ras-MAPKinase. PKC et les récepteurs aux facteurs de croissance.

Des études ont montré que la neurofibromine peut se lier à cette protéine membranaire en formant un complexe qui contrôle aussi la prolifération et la différenciation cellulaire⁸³.

- *Régulation de l'Adenylyl cyclase (AC)*

La neurofibromine régule de manière positive l'AC, une enzyme qui génère de l'AMP cyclique (AMPC). Cet AMPC a montré son importance dans les fonctions d'apprentissage et de la mémoire, tout en ayant une fonction possible de suppresseur de tumeur par la régulation de l'activité de Ras⁸⁴.

- *Motilité cellulaire*

Le rôle de la neurofibromine dans la motilité cellulaire est important pour le fonctionnement et le développement neuronal, tout comme dans sa fonction de suppresseur de tumeur.

La neurofibromine régule la dynamique et la réorganisation des filaments d'actines via la voie Rho-ROCK-LIMK2-cofillin.

Un déficit de neurofibromine déclenche, à travers cette voie Rho-ROCK-LIMK2-cofillin, l'altération de l'organisation du cytosquelette d'actine, ce qui promeut la motilité, l'invasion, et l'adhésion des cellules. Il en résulte au final la formation de larges agrégats cellulaires⁸⁵. Ce mécanisme pourrait être une des clés de la formation des neurofibromes chez les patients NF1.

- *Association à la focal adhesion kinase (FAK)*

La protéine FAK est une protéine ubiquitaire que l'on retrouve au niveau du complexe d'adhésion cellulaire dans le cytoplasme. L'activation de FAK permet l'activation en cascade des protéines de la famille des Src, qui ont un rôle important dans la régulation de la survie cellulaire (migration, prolifération...). La neurofibromine peut s'associer à la partie N-terminale de FAK et permettre la régulation des mécanismes de tumorigenèse⁸⁶.

- *Rôle dans la transition épithélio-mésenchymateuse (TEM)*

La TEM fait partie des mécanismes prépondérants dans la tumorigenèse et le développement de métastases. Des études ont montré que l'absence de neurofibromine induisait une expression anormale des facteurs de transcription relatifs à la TEM⁸⁷, démontrant une fois de plus le rôle de régulateur porté par la neurofibromine.

Tous ces mécanismes montrent bien les diverses associations et rôles fonctionnels de la neurofibromine, loin d'être encore toutes découvertes. La neurofibromine a prouvé son importance dans l'homéostasie des cellules, et la figure 17 montre le caractère central des anomalies de l'expression de la neurofibromine dans la tumorigenèse.

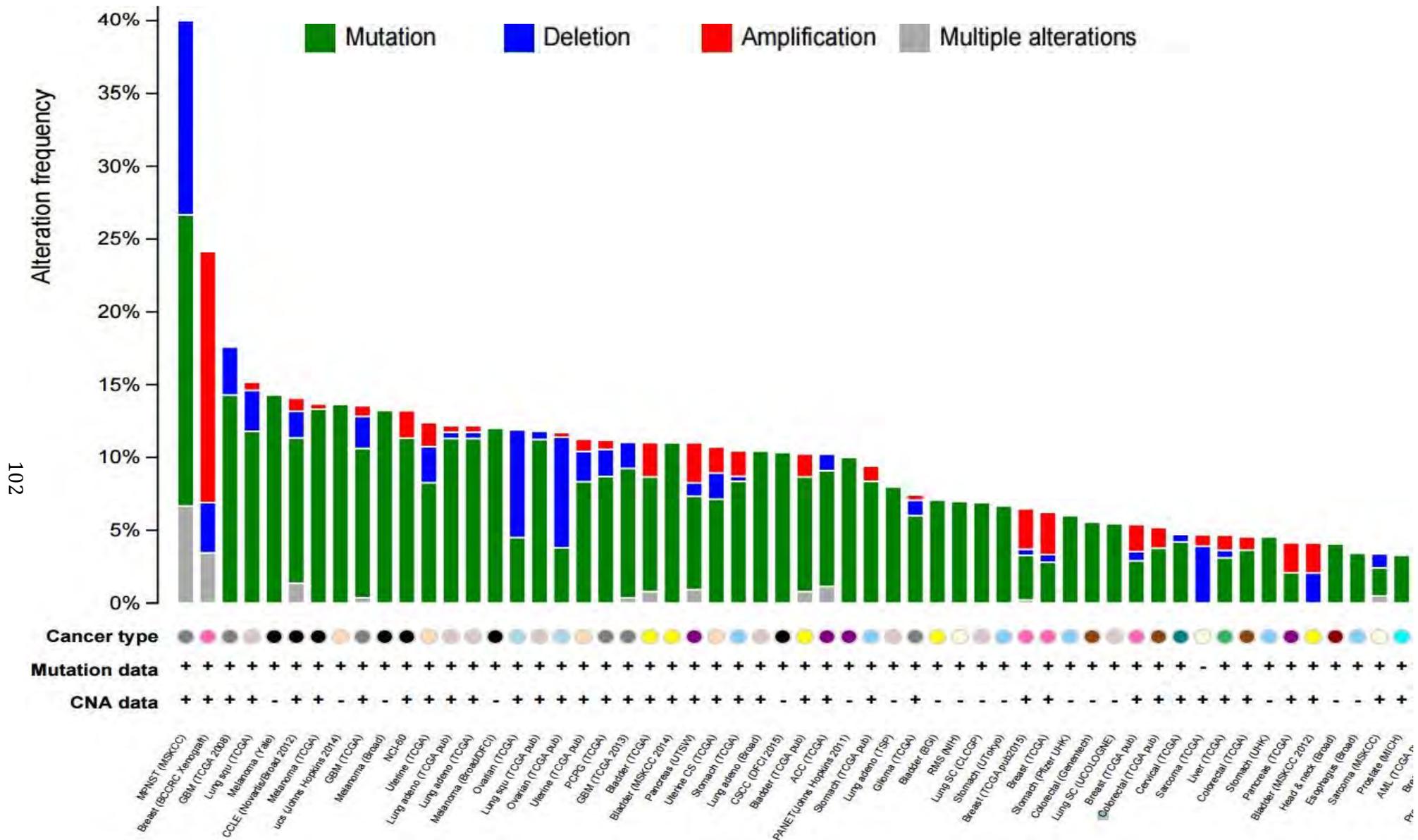


Figure 17. Fréquences des altérations génétiques du gène *NF1* dans les néoplasies humaines. (Source: The cBio Cancer genomics Portal; <http://www.cbioportal.org>, décembre 2015).

Régulation :

L'expression du gène *NF1* est régulée à différents niveaux et cette régulation est basé sur de multiples mécanismes. Il en résulte une variabilité de l'expression intracellulaires, qui peut être de l'ordre de la minute⁸⁸.

Parmi les mécanismes mis en jeu, on retrouve :

- Des facteurs épigénétiques : microARN et méthylation de l'ADN.
 - Les microARN sont des petites séquences non-codantes endogènes qui peuvent influencer l'expression post-transcriptionnelle de certains gènes cibles (inhibition de l'expression)⁸⁹.
 - Une étude a montré que le micro-ARN miR-10b peut se lier à la région 3'UTR de l'ARNm du gène *NF1* et réguler la tumorigenèse⁹⁰. Un article récent en conclut même que la hausse de l'expression de miR-10b pouvait en faire un marqueur indépendant de mauvais pronostic dans les gliomes⁹¹. Un autre micro-ARN, le miR-193b régule de manière négative l'expression de *NF1* et il est surexprimé dans les carcinomes épidermoïdes de la tête et du cou⁹².
 - La méthylation des promoteurs est une marque répressive, qui inactive la transcription d'un gène par la présence d'un groupement méthyle empêchant l'interaction entre la séquence nucléotidique et les facteurs de transcription. La méthylation permet aussi le recrutement de protéines dites MBD (méthyl-CPG-binding domain proteins) qui modifient les histones et ainsi la structure de la chromatine (eu/hétérochromatine)⁹³. Même si les études n'ont pu jusqu'à présent démontrer un rôle principal de la méthylation⁹⁴, elle a été mise en cause dans l'inactivation du deuxième allèle selon le mécanisme de « second-hit », chez un patient *NF1* ayant développé un astrocytome pilocytique⁹⁵.
- Une dégradation par le protéasome :
 - Le protéasome est un complexe enzymatique multiprotéique dont la fonction principale est de dégrader les protéines mal repliées, dénaturées ou obsolètes de manière ciblée.

Le complexe ubiquitine-ligase qui contrôle la régulation de la destruction par le protéasome de *NF1* a récemment été identifié⁹⁶ et correspond à un complexe Cullin 3 / kelch repeat and BTB domain-containing 7. Une dégradation protéasomale excessive de la protéine *NF1* mène à une baisse de son expression, comme cela a été retrouvée dans des gliomes⁹⁷.

4.1.8. ALTERATIONS ET MUTATIONS DU GENE *NF1*

La moitié des patients *NF1* hérite d'une mutation alors que la seconde moitié arbore une mutation *de novo* du gène *NF1*.

Le taux de mutation lors de la réplication d'un génome stable est estimé approximativement entre 10^{-8} et 10^{-10} /pb par génération cellulaire⁹⁸.

Le locus du gène *NF1* possède un taux de mutation spontané 10 fois supérieur en comparaison de la plupart des loci et il constitue l'un des locus au plus fort potentiel de mutation spontanée du génome humain.

Le taux de mutation d'une génération à une autre, et donc potentiellement entre 2 générations de sujets, a été estimé entre $3,1 \times 10^{-5}$ and $1,4 \times 10^{-4}$ par Hudson sur une population du sud-est du pays de Galles⁹⁹. Aucun effet de l'âge des parents sur ce fort taux de mutation n'a été démontré⁹⁷. Il existe par contre un fort biais lié au sexe sur l'apparition de mutation spontanée, plus de 80% des mutations étant d'origine paternelle¹⁰⁰ alors que les microdélétions auraient plutôt une origine maternelle¹⁰¹.

SNV (short nucléotides variation, ou mutation ponctuelle) du gène *NF1*

Plus de 1400 mutations du gène *NF1* ont été rapportées dans la « Human Gene Mutation Database » (<http://www.hgmd.cf.ac.uk>, décembre 2015). Une grande majorité des mutations sont dites « *perte de fonction* ».

Tous les types de mutations sont recensés : non-sens et faux sens, mutation au niveau d'un site d'épissage, délétions, insertions, avec décalage du cadre de lecture et translocations (voir figure 18 pour la répartition de ces mutations). La part la plus importante revient aux délétions.

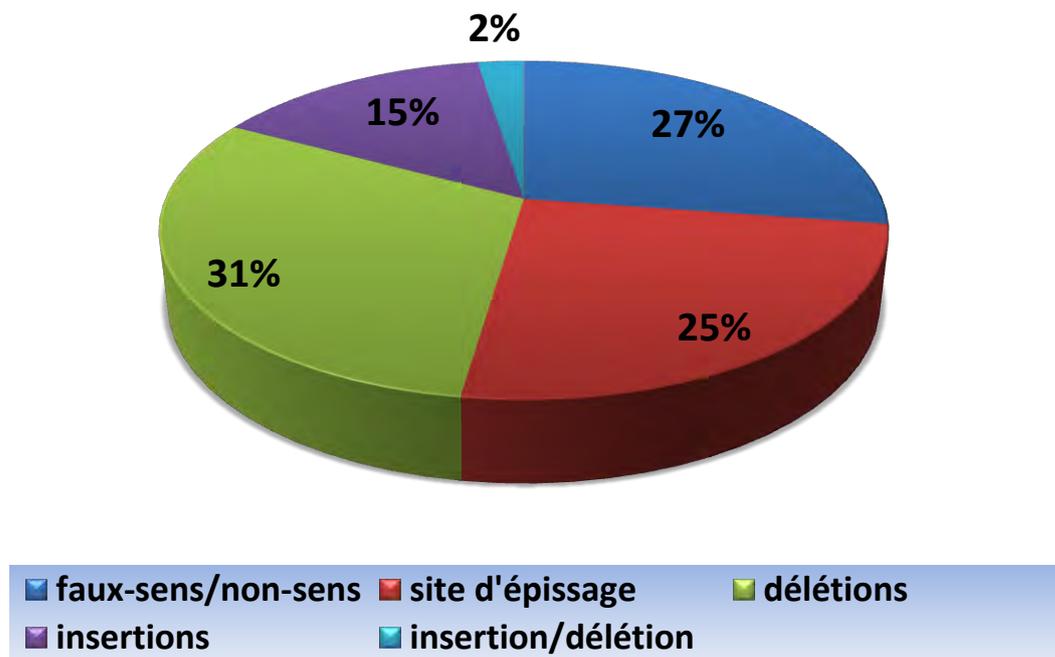


Figure 18. Répartition des mutations "Shorts nucléotides Variations (SNV)" du gène NF1.. (Source HGMV, decembre2015).

Si aucun « *point chaud* » (ou « hotspot ») de mutation n'est retrouvé, il est notable qu'une plus grande proportion de mutations non-sens ou faux-sens se trouvent dans le domaine GRD de la neurofibromine. Ceci supporte l'idée que cette structure joue un rôle crucial dans la pathologie¹⁰².

Réarrangements du gène *NF1*

Les patients présentant un CNV (Copy Number Variation) du locus du gène NF1 comptent pour environ 5% des patients NF1. Ce sont pour la majorité de larges délétions prenant tout le gène et les régions flanquantes¹⁰³. Les délétions/duplications intragéniques comptent pour environ 2% des patients NF1.

D'un point de vue clinique, le diagnostic moléculaire de ces grands remaniements est important car ils sont fréquemment associé à des manifestations cliniques sévères, comme un risque accru de MPNSTs, comparé au reste des patients NF1¹⁰⁴.

Trois grands types de délétions sont récurrents et forment la majorité des délétions retrouvées (figure 19) :

- **Type 1** : Le plus fréquent : 70 à 80% des patients¹⁰⁵. Les points de cassure du type 1 sont localisés dans des LCRs (low-copy repeats) nommés NF1-REPa et NF1-REPc. La taille de la délétion est de 1,4 Mb et entraîne la perte de 14 gènes.
- **Type 2** : Moins fréquent que le type 1 avec 9 à 20 % des larges délétions. Les points de cassures sont situés à l'intérieur du gène *SUZ12* et de son pseudogène *SUZ12P*. Un peu plus courte avec une taille de 1,2 Mb, 13 gènes sont supprimés (*LRRC37B* ne l'est pas en comparaison avec le type 1).
- **Type 3** : Plus rare que les deux autres types (1-4 %), cette délétion a une taille de 1 Mb et inclut la délétion de 9 gènes. Les points de cassures sont localisés dans les LCRs NF1-REPb et NF1-REPc.
- **Délétions atypiques** : une faible proportion des délétions (8-10%) sont dites atypiques car les points de cassures ne sont pas récurrents et sont donc hétérogènes dans leur taille, et le nombre de gènes emportés.

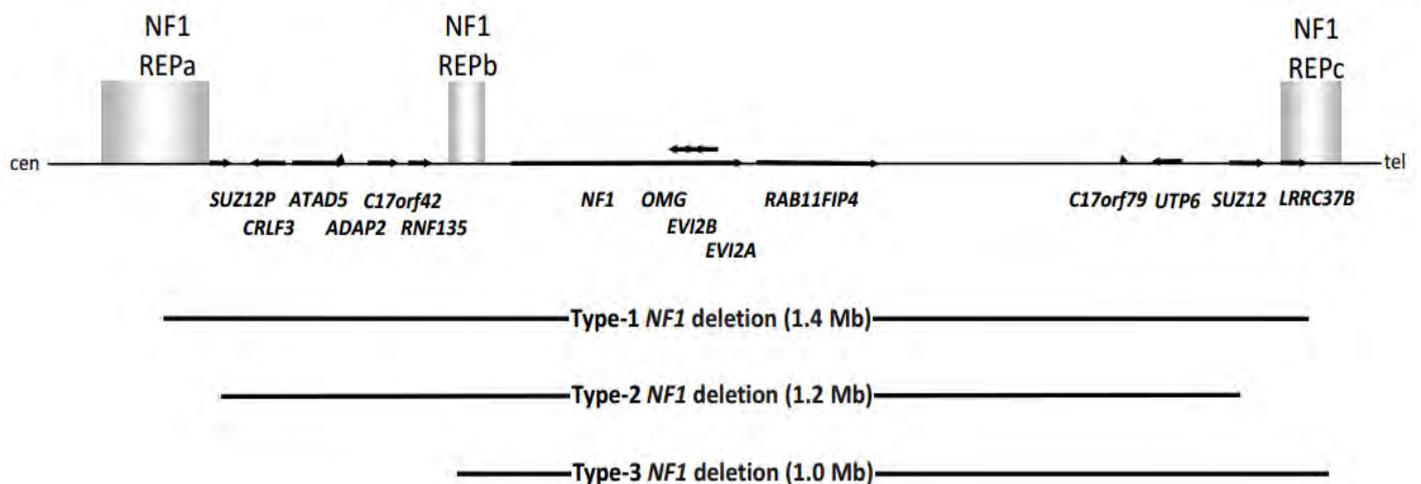


Figure 19. Schéma de la région du gène NF1 et les différents type de délétions récurrentes (Vogt J. et al, 2011¹⁰⁶).

Légende : Le sens de transcription de chaque gène est donné par les flèches horizontales.

Les LCRs NF1-REPa, NF1-REPb, NF1-REPc sont indiqués par des barres verticales.

« Cen » pour centromérique et « tel » pour télomérique.

Ces grandes délétions sont connues pour induire un « syndrome des gènes contigus », c'est-à-dire une association de signes cliniques dus à des délétions incluant plusieurs gènes voisins.

Le phénotype découlant de ces délétions est bien caractérisé pour les délétions de type 1 par l'association : d'une dysmorphie faciale, d'une grande taille, de larges mains et pieds, d'une scoliose, de retards de développement et des apprentissages et un haut risque de neurofibromes (spinaux, MPNST...).

Le phénotype est moins marqué pour le type 2, car le plus souvent la délétion est présente en mosaïque, et donc toutes les cellules ne sont pas atteintes par le réarrangement (NF1 segmental). Le mosaïcisme à contrario est peu retrouvé pour les délétions de type 1 (2-4% des cas¹⁰⁷).

Les patients ayant une délétion germinale de type 2 (très rares cas) ont cependant un phénotype similaire au type 1. Le gène *LRRC37B* qui différencie les deux types de délétions et dont les fonctions ne sont pas connues aurait donc un impact limité sur le phénotype¹⁰⁶.

4.1.9. RELATION GENOTYPE-PHENOTYPE

La grande variabilité des manifestations cliniques se retrouvent chez les patients d'une même famille partageant la même mutation¹⁰⁸.

Des exceptions existent cependant et dans certains cas, un rapprochement existe entre une altération génomique et une prédisposition à un trait phénotypique.

Il a été démontré que :

- Les patients ayant une mutation au niveau du site d'épissage ($\approx \frac{1}{4}$ des cas) ont un risque tumoral accru, en particulier pour les gliomes cérébraux et les MPNSTs¹⁰⁹.
- Les patients avec une délétion totale du gène NF1 et quelques gènes flanquants ($\approx 5-10\%$ des patients) ont un risque accru de MPNSTs au cours de leur vie¹¹⁰ et présentent un phénotype sévère.
- La délétion rare de 3 pb dans l'exon 17 (c.2970_2972delAAT p.Met992del) qui respecte le cadre de lecture, entraîne l'absence de neurofibromes cutanés chez les patients NF1¹¹¹ et un phénotype bénin.

- Récemment, une substitution dans l'exon 29 (c.5425C>T p. Arg1809Cys) a été rapportée comme donnant un phénotype intermédiaire. Sa particularité est l'absence de nodules de Lisch ou de neurofibromes chez les patients, mais une surexpression de lipomes¹¹². D'une manière plus globale, toutes les mutations faux-sens touchant le codon p.1809 (un des plus fréquemment altéré dans la cohorte des auteurs) ont été étudiées et un phénotype intermédiaire décrit sur une large cohorte¹¹³.

Ces notions sont à prendre en compte par le clinicien en charge du suivi des patients NF1.

4.1.10. GENES MODIFICATEURS

Comme nous l'avons vu, la NF1 est une maladie monogénique avec une faible corrélation génotype-phénotype. Dans une étude, des jumeaux monozygotes présentant une NF1 ont un haut degré de concordance pour les neurofibromes cutanés et le nombre de TCLs, mais ils étaient discordants sur certains autres traits¹¹⁴.

Il a ainsi été proposé que des gènes modificateurs puissent expliquer la variabilité phénotypique des patients NF1.

Même si aucun lien n'est clairement établi avec un gène modificateur précis, on a pu démontrer que :

- Certains gènes proches du gène *NF1* ont un rôle, comme pour la délétion polygénique sus-citée. Par exemple, des mutations somatiques dans le gène *SUZ12*, emporté par la délétion, ont été récemment retrouvées dans les MPNSTs¹¹⁵.
- Les gliomes des voies optiques seraient moins fréquents chez les populations Africano-américaines que chez les autres populations¹¹⁶.
- Le sexe des patients peut influencer sur le pronostic de la maladie. En effet, les garçons montrent un risque plus élevé de déficit de l'apprentissage, alors qu'une étude sur des souris femelles a révélé un pronostic inférieur pour les gliomes des voies optiques¹¹⁷.

- Sur un modèle murin, un article novateur a démontré qu'une région soumise à empreinte sur un chromosome pouvait interagir spatialement avec une région située entre les gènes *NF1* et *WSB1* (d'un second chromosome) et ainsi contrôler la transcription du gène *NF1*¹¹⁸.
- L'insertion de rétrotransposons, comme les séquences LINE1 ou *Alu*, représenterait une part non négligeable des altérations de NF1.

4.1.11. DETECTION DES ALTERATIONS DU GENE *NF1*

Jusqu'à présent, la sensibilité des techniques moléculaires classiques est de l'ordre de **95%** chez les patients correspondant aux critères diagnostic du NIH. Cette sensibilité élevée est atteinte grâce à la combinaison de plusieurs techniques complémentaires réalisées par étapes successives comme : la recherche d'hétéroduplex par dHPLC (HPLC en condition dénaturante), le séquençage par méthode Sanger des exons et l'analyse des transcrits par rt-PCR.

Les réarrangements du gène NF1 sont recherchés par M.L.P.A, qPCR ou des tests cytogénétiques (FISH, CGH-array)¹¹⁹.

Pour les patients chez qui aucune mutation ni réarrangement ne sont retrouvés, l'explication peut être donnée par : une mutation intronique profonde (c'est-à-dire loin des régions flanquantes d'un exon qui forment la majorité des mutations introniques), un mosaïcisme, ou un réarrangement chromosomique complexe¹²⁰.

Les mutations introniques profondes, si elles ont un effet sur l'expression, ne peuvent être identifiées que par l'analyse de la protéine ou de son transcrit (test de troncation des protéines ou séquençage Sanger des produits de rt-PCR) ou par une étude du génome entier. Cependant, elles sont rares (13/565 dans l'étude de la cohorte française¹¹⁷).

Pour aller plus loin, on peut imaginer qu'un des mécanismes de régulation soit altéré et entraîne une déficience en neurofibromine (cf : mécanisme de régulation).

4.1.12. STRATEGIE ACTUELLE DE DETECTION AU LABORATOIRE DE LYON

Actuellement, la recherche des altérations du gène *NF1* chez un cas index est réalisée en combinant successivement (figure 20) ;

- **Une recherche des réarrangements intragéniques ou extragéniques par technique MLPA.**

La technique MLPA (multiplex ligation-dependent probe amplification ou amplification multiplexe de sondes dépendant d'une ligation) est une technique développée et commercialisée par la compagnie MRC-Holland depuis 2002¹²¹. Elle consiste en une variante de la PCR multiplexe dans laquelle les sondes sont divisées en deux oligonucléotides qui nécessitent leur ligation pour que l'amplification ait lieu. Les amplicons sont détectés grâce à un marqueur fluorescent attaché aux amorces PCR et après séparation par électrophorèse (analyse de fragment sur séquenceur).

Entre 5 et 10% des patients *NF1* ont un réarrangement (délétion ou duplication) d'une partie ou du gène entier. Deux kits MLPA (P-081 et P-082) sont utilisés afin d'avoir une sensibilité optimale sur un nombre maximum d'exons. Un troisième kit (P-122) permet de préciser les points de cassure des régions flanquantes du gène *NF1* et les autres gènes ainsi touchés.

Pour les patients positifs avec un seul exon de modifié, on effectue la vérification par séquençage Sanger, pour exclure les mutations présentes au niveau des nucléotides sur lesquels les sondes de la MPLA viennent s'hybrider. Les patients MLPA-positifs ne continuent l'analyse par dHPLC.

L'analyse par MLPA est sensible et spécifique, demande peu de temps technique, mais requiert une analyse fine réalisée par une personne expérimentée. Elle permet une identification rapide des patients présentant un CNV du gène *NF1* et ainsi d'économiser les étapes suivantes de screening et de séquençage du gène qui sont longues et fastidieuses¹²².

- **Pour les patient MLPA-négatifs, un screening des mutations par dHPLC est réalisé.**

Le gène *NF1* est un grand gène qui possède plus d'exons que la moyenne des gènes du génome humain (≈ 60 vs 8.8 en moyenne¹²³). En conséquence, le séquençage immédiat de tous les exons avec les moyens actuels se révéleraient longs et coûteux, tout en demandant un temps technique important.

Afin de réduire le nombre de séquences à analyser par séquençage Sanger, une méthode de screening par dHPLC (Chromatographie liquide à haute pression en condition dénaturante) a été mise en place.

Après obtention par PCR des amplicons exoniques, une première étape de dénaturation-renaturation permet de créer des homoduplexes (brins d'ADN se réassociant avec une séquence similaire) et des hétéroduplexes (brins d'ADN se réassociant avec un mésappariement sur une ou plusieurs bases). Le mésappariement entraîne une instabilité qui réduit le temps de rétention dans la colonne chromatographique et permet la séparation. L'analyse des pics issus de la chromatographie permet de détecter les mauvais appariements (*missmatches*) et impose la vérification par séquençage Sanger des exons concernés.

La dHPLC permet de détecter des substitutions, des insertions ou délétions de petite taille de manière rapide et avec un haut débit¹²⁴. Cette méthode a aussi une grande sensibilité pour l'analyse des mutations du gène *NF1*, proche de 95 couplée à d'autres méthodes^{125,126}. Cette technique permet aussi de détecter les nombreux polymorphismes (SNP) du gène *NF1*. La limite intrinsèque de la méthode est la non détection des polymorphismes homozygotes.

Un défaut important de la dHPLC est la une mise au point difficile, qui demande du temps et beaucoup de patients tests dans le but d'affiner les conditions d'analyse, indépendantes et à mettre au point pour chaque réaction PCR.

De plus, l'entreprise Transgenomic qui développait les modèles dHPLC WAVE® permettant l'analyse, ne suit plus cette gamme de produits. Leur pérennité est donc aujourd'hui limitée.

- **Un séquençage Sanger ciblé pour les positifs à la dHPLC**
- **Ou séquençage de tous les exons pour les patients dHPLC-négatifs dont la clinique est évocatrice de la NF1.**

La méthode de séquençage de l'ADN développée par F. Sanger depuis 1977¹²⁷ est devenue la méthode de référence ou « gold-standard » pour l'analyse des séquences et la détection des substitutions et des petites insertions ou délétions.

Elle est basée sur l'interruption de l'élongation du brin d'ADN complémentaire par l'ADN polymérase grâce à l'incorporation de didésoxynucléotides (ddNTP) en lieu et place des désoxynucléotides normaux (dNTP).

Ces ddNTP n'ont pas de groupement -OH en 3', contrairement aux dNTP, ce qui empêche la synthèse enzymatique du brin d'ADN. Les séquences obtenues auront une taille dépendante du point d'insertion du ddNTP, et seront séparées par électrophorèse capillaire dans un gel de polyacrylamide.

A la place d'un signal radioactif, ils sont marqués depuis 1986 par un fluorophore différent à chaque ddNTP¹²⁸ qui permet la détection du dernier nucléotide inséré dans la séquence. Cette évolution a permis l'avènement des séquenceurs¹²⁹ dits de « première génération » (avec notamment le ABI310 en 1995, Applied Biosystems) sous l'impulsion du Human Genome Project (HGP) débuté en 1990.

Les fluorogrammes obtenus de manière automatisée et informatisée (pics de fluorescence par rapport à un temps, ou une longueur) sont par la suite traités par des algorithmes qui assignent des bases à chaque pic : c'est l'étape de « base-calling », développée grâce aux travaux de Ewing et Green^{130,131}.

Les bases ayant un score de qualité « Phred » satisfaisant (c.a.d dont la probabilité d'erreur est faible) sont ensuite alignées contre une séquence de référence (ou matrice).

Cette procédure de « mapping » est réalisée grâce à des logiciels d'alignement (notamment SeqScape® de Applied Biosystem) utilisant des algorithmes spécifiques (Algorithme de Smith-Waterman pour l'alignement local¹³² de SeqScape®). L'analyse des mutations peut alors se faire visuellement par le technicien ou le biologiste.

Le principal défaut du séquençage Sanger est sa faible sensibilité lorsque la proportion de séquences portant une mutation est faible, comme pour les mutations somatiques ou dans les cas de mosaïcisme (limite de sensibilité estimée proche des 15%)¹³³.

Son coût, sa relative lenteur et l'importance du temps tertiaire d'analyse, par des personnes compétentes, sont aussi à prendre en considération dans les limites de cette technique.

Les séquençages Sanger sont toujours vérifiés puis confirmés par une deuxième extraction ou un nouveau prélèvement.

L'exon 1 de *NF1* n'est pas analysé par dHPLC en raison de sa forte teneur en GC, qui demande une mise au point complexe mais est séquencé automatiquement pour tous les patients. Les régions UTR du gène ne sont pas séquencées par cette technique.

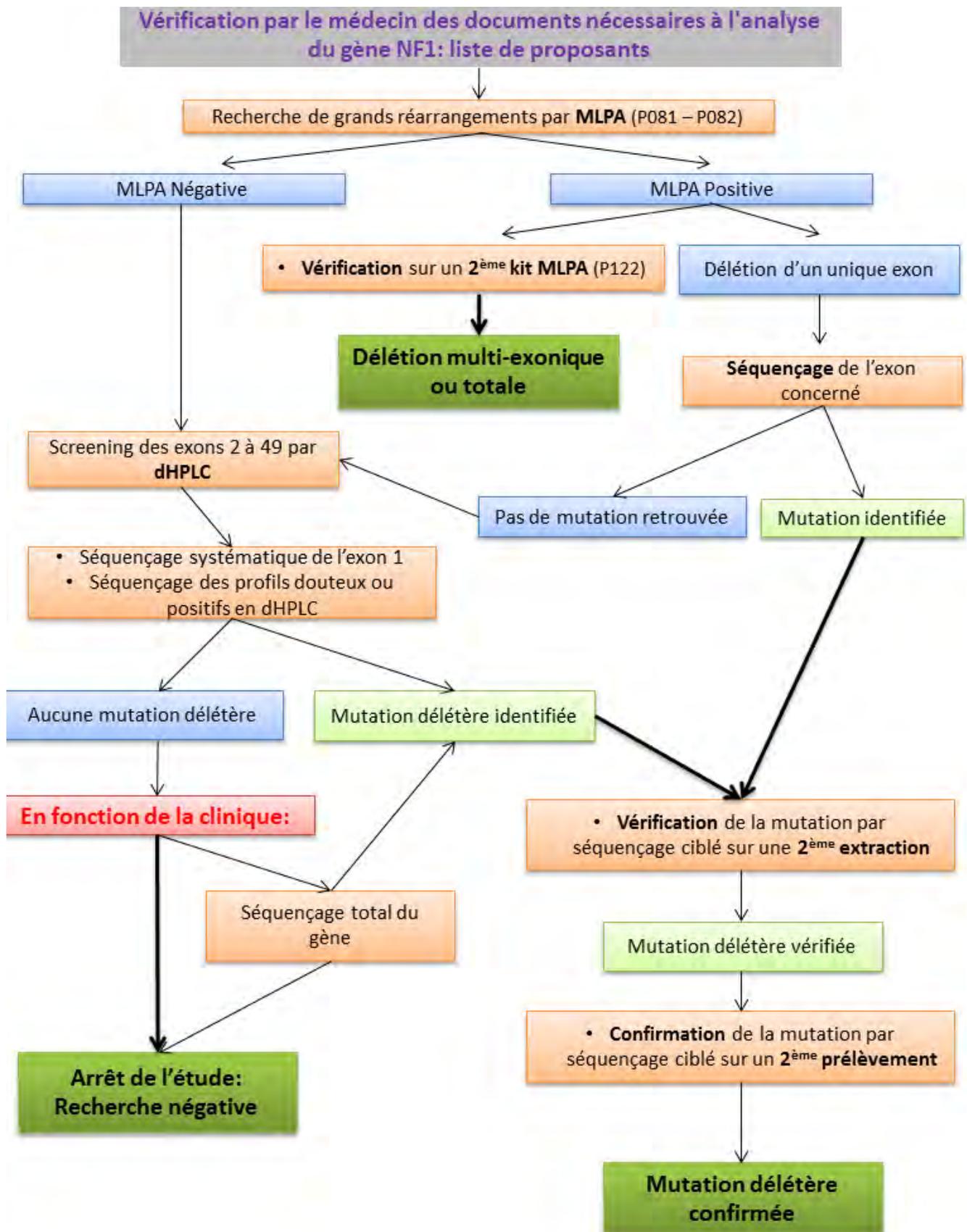


Figure 20. Stratégie du laboratoire pour l'analyse du gène NF1.

4.2. LE GÈNE *SPRED1*, UNE NECESSITE D'ANALYSE

Le gène *SPRED1* (Sprouty-related, EVH1 domain containing 1) a été découvert en 2001 chez la souris¹³⁴.

Les mutations de *SPRED1* sont retrouvées chez 1-4% de patients correspondant aux critères clinique de la NF1. Près de 15% des patients avec un syndrome de Légitus répondent à ces mêmes critères diagnostiques. Ce chevauchement phénotypique a longtemps était ignoré par les cliniciens.

Nos connaissances sur le gène ou la protéine *SPRED1* et leurs fonctions sont très limitées. Une simple recherche sur Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed>, au 31 décembre 2015) avec pour mot-clé « NF1 » donne **3 980** articles (premier item en 1982). Pour le mot-clé « *SPRED1* », seuls **117** articles ressortent de la recherche (premier item en 2001).

4.2.1. CLONAGE POSITIONNEL

Brems et son équipe ont effectué en 2007 des analyses de liaison pangénomiques dans deux familles. Ces familles exprimaient avec un trait dominant autosomique un phénotype composé de multiples taches café-au-lait, des lentigines axillaires, une macrocéphalie et une dysmorphie « Noonan-like ».

Aucune anomalie du gène *NF1* n'avait été retrouvée, ce qui amena Brems à effectuer une recherche de mutation sur d'autres loci (l'analyse de liaison au locus de *NF1* était négative). L'analyse de liaison a permis de localiser une région du chromosome 15 (*lod-score* 4.8). Le gène *SPRED1* était un bon gène candidat car connue depuis 2001 comme un gène régulant la voie de signalisation MAPK (de même que la neurofibromine). Son séquençage révéla des mutations non-sens délétères chez les probands des deux familles et non retrouvées chez les sujets sains.

Le gène *SPRED1* humain (*609291) est localisé au niveau du bras long du chromosome 15 au niveau de la bande 15q14.

4.2.2. STRUCTURE

Le locus possède une séquence d'une taille de 104.4 kb, et contient 7 exons. Le gène commence à la position g.38545052 et fini à la position g.38649450 et il est situé sur le brin sens (Hg19).

4.2.3. PSEUDOGENE

Aucun pseudogène n'a été rapporté dans la littérature à ce jour.

4.2.4. ARN MESSAGER

Le transcrit principal NM_152594.2 est composé de 7255 pb. La région 3'UTR est très large et compte 5585 pb.

Aucun autre transcrit n'est décrit dans la littérature.

4.2.5. PROTEINE SPRED1

La protéine SPRED1 est composée de 444 AA, et elle a un poids moléculaire d'environ 50kDa.

SPRED1 appartient à la famille des protéines SPRED, avec SPRED2 et SPRED3¹³⁵, qui sont caractérisées par leur domaine SPR (Sprouty-related C-terminal cysteine-rich).

Trois domaines fonctionnels ont été identifiés (figure 24) :

- Un domaine **EVH-1** (Ena/Vasp Homology 1) en position N-terminal.
- Un domaine **c-KIT-binding central** (KBD, non retrouvé chez SPRED3).
- Un domaine **SPROUTY-related** (SPR) en position C-terminal.

4.2.6. EXPRESSION

L'expression de *SPRED1* est forte dans le SNC au niveau de la zone germinale (surtout pendant la neurogénèse) et permet le maintien de structure des zones ventriculaires¹³⁶. *SPRED1* est essentiel dans le développement normal du cortex.

Son expression est aussi conséquente dans les poumons, le cerveau, la moelle épinière et la rate. Elle est faible dans le foie, le pancréas, la prostate, les reins, le cœur, le thymus, les muscles et la moelle osseuse¹³⁷.

4.2.7. ACTIVITE FONCTIONNELLE

Ces dernières années, de nombreuses études fonctionnelles ont permis de prendre conscience des fonctions multiples de la protéine *SPRED1*.

- **Régulation négative de la voie Ras-MAPK**

SPRED1 est phosphorylée par des kinases et régule l'activation de la cascade Ras-MAPK en réponse à divers facteurs de croissances. Cette régulation n'affecte pas les voies Akt ou Rac.

SPRED1 est constitutivement associée à Ras sans prévenir son activation (Ras-GTP) ou la translocation membranaire de Raf. Cependant, *SPRED1* inhibe l'activation des MAPKinases en supprimant la phosphorylation et l'activation de Raf¹³¹.

Le domaine SPR est nécessaire à la localisation membranaire de *SPRED1* et à la co-localisation de Ras.

- **Effets et interactions en lien avec d'autres partenaires**

D'autres interactions ont été identifiées, comme l'inhibition de la kinase TESK1 (testis-specific protein kinase)¹³⁸, l'augmentation du temps de rétention du récepteur FGFR1 (fibroblast growth factor receptor like-1)¹³⁹, l'inhibition de l'activité de la kinase DRK1A¹⁴⁰, ou encore son rôle comme substrat de SHP2¹⁴¹.

De plus, il a été montré récemment que SPRED1 se lie à la neurofibromine et permet son recrutement au niveau de la membrane cellulaire où la neurofibromine exerce sa fonction de régulateur négatif de Ras-GTP.

Il est acquis que des mutants non-sens de SPRED1 sont incapables de se lier à la neurofibromine et empêche son recrutement, démontrant bien le rôle de vecteur tenu par Spred1¹⁴². Le domaine EVH-1 de SPRED1 interagirait avec 16 AA en position N-terminal et 20 AA en position C-terminal du domaine GRD de la neurofibromine.

Cette découverte récente conforte bien l'idée que la neurofibromine et SPRED1 ont un rôle similaire ou du moins complémentaire dans une même voie. Ceci est cohérent avec le chevauchement phénotypique qui résulte dans l'anomalie de l'une ou l'autre protéine.

Ainsi, le rôle fondamental du domaine EVH-1 permet de considérer le phénotype modéré du syndrome de Légius (comparé à NF1) comme une conséquence de la redondance de ce domaine chez *SPRED2* et *SPRED3*, qui compenserait pour partie la perte d'expression de *SPRED1*¹⁴³.

- **Gène suppresseur de tumeur**

Les gènes *SPRED* sont des supposés gènes « suppresseur de tumeur », de par leur rôle dans les mécanismes de régulation de la voie Ras-MAPK et plusieurs autres études dans le domaine du cancer.

Des équipes ont ainsi rapportés la baisse de l'expression de *SPRED1* et *SPRED2* dans les carcinomes hépatocellulaires¹⁴⁴ ou dans les leucémies aiguës myéloblastiques (cf : clinique du syndrome de légius, point 3.2).

Dans une autre étude, la surexpression de SPRED1 dans les cellules tumorales de souris induit une baisse de la tumorigénèse¹⁴⁵.

4.2.8. ALTERATION DE SPRED1

SNV (short nucléotides variation, ou mutation ponctuelle) :

Toutes les mutations et polymorphismes retrouvés dans diverses études sont répertoriés dans la base de donnée « Leiden Open Variation Database » (<http://www.lovd.nl/SPRED1>).

A ce jour (décembre 2015), 153 variants uniques sont inscrits (figure 21).

Tous les types de variants sont représentés, avec une prédominance de substitutions (62%) et de délétions (26%).

| variant | number | location | | | | percentages |
|-------------------------------------|------------|----------|------------|-----------|----------|-------------|
| | | 5'start | coding | intron | 3'stop | |
| substitutions | 165 | 1 | 150 | 13 | 1 | |
| deletions | 69 | 7 | 57 | 5 | 0 | |
| duplications | 23 | 0 | 23 | 0 | 0 | |
| insertions | 3 | 0 | 3 | 0 | 0 | |
| insertion/deletions | 2 | 0 | 0 | 2 | 0 | |
| totals | 262 | 8 | 233 | 20 | 1 | |

Figure 21. Variants de *SPRED1* répertoriés

(https://grenada.lumc.nl/LOVD2/mendelian_genes/variants_statistics.php, décembre 2015).

L'impact sur les protéines est de type « perte de fonction » avec une répartition homogène entre substitutions, modifications du cadre de lecture (frame shift), et faux-sens (figure 22).

| variant | number | percentages |
|--|-----------------------------|-------------|
| substitutions | confirmed: 52 predicted: 20 | |
| deletions | predicted: 1 | |
| frame shifts | confirmed: 45 predicted: 43 | |
| no protein variants | confirmed: 7 | |
| nonsense | confirmed: 51 predicted: 24 | |
| translation initiation variant | predicted: 2 | |
| silent | confirmed: 5 | |
| unknown | 12 | |
| total | 262 | |

Figure 22. Variants protéiques de *SPRED1*.

En bleu les variants confirmés, en rouge les variants prédits.

(https://grenada.lumc.nl/LOVD2/mendelian_genes/variants_statistics.php).

Pour les mutations somatiques, contrairement à *NF1* souvent muté dans les cancers, une seule anomalie a été rapportée par Brems et son équipe. Cette mutation somatique a été retrouvée dans une culture de mélanocytes prélevés dans une TCL d'un patient avec un syndrome de Légius. Cette expérience avait déjà été réalisée sur des mélanocytes dérivés de TCL de patients *NF1*¹⁴⁶, démontrant la présence d'une inactivation bi-allélique dans les deux contextes: *NF1* ou syndrome de Légius.

Réarrangements du gène *SPRED1* (CNV) :

L'analyse des CNV peut se faire par différentes méthodes, dont la MLPA, la CHG-array, ou encore la qPCR¹⁴⁷.

Dans une large étude sur 510 patients *NF1*-négatifs¹⁴⁷, 40 sur ces 510 patients avaient une altération du gène *SPRED1*. Dans ces 40 altérations, **10%** de CNV ont été retrouvés (4/510 au total). La proportion de CNV semble donc être modérée chez les patients avec un syndrome de Légius, et proche de celle retrouvée pour les patients *NF1*.

4.2.9. RELATION GENOTYPE-PHENOTYPE.

Alors que de nombreuses mutations déjà été décrites, aucune relation génotype-phénotype n'a encore été identifiée. On peut remarquer sur la figure 24 un réel continuum de mutations tout le long du gène, sans cluster ou hotspot de mutation, ce qui révèle que n'importe type de mutation peut entraîner les mêmes conséquences phénotypiques.

Pour les quelques cas de CNV rapportés, aucune modification du phénotype n'a été mise en évidence, même si la délétion emporte d'autres gènes voisins (comme *TMC05A*, *FAM98B*). Cependant, un phénomène de « syndrome des gènes contigus » peut exister si des régions flanquantes trop importantes sont délétées. Un patient avec une grande délétion de 6.6 Mb emportant 29 gènes additionnels, on a pu observer une fente labio-palatine. Cette dysmorphie avait déjà été rapportée avec une délétion de cette même région, mais ne touchant pas le gène *SPRED1*¹⁴⁵.

4.2.10. DETECTION DES ALTERATIONS DU GENE *SPRED1*

Depuis la découverte du chevauchement clinique partiel entre les patients NF1 et le syndrome de Légius, il est devenu essentiel de ne pas manquer le diagnostic entre les deux pathologies.

Dans son article de 2012, Brems¹⁴⁸ et son équipe recommande :

- 1) de tester *NF1* en premier lieu pour les patients ne présentant que les signes cliniques confondants (TCL, lentigines),
- 2) puis si le résultat est négatif, d'analyser *SPRED1*.

Ces recommandations viennent du fait que des mutations pathogéniques de NF1 sont bien plus fréquemment retrouvées dans les cas sporadiques (sans critères phénotypiques propres à NF1) que pour *SPRED1* (43% vs 1.3% pour Messiaen¹⁴⁹).

Ces recommandations sont suivies par notre laboratoire.

4.2.11. STRATEGIE ACTUELLE DE DETECTION AU LABORATOIRE DE LYON

Actuellement, la recherche des altérations du gène *SPRED1* chez un cas index est réalisée en combinant successivement (figure 23) :

- **Un séquençage Sanger de tous les exons pour les patients.**

Le gène *SPRED1* étant relativement court, tous les exons sont séquencés directement (sans les régions UTR, mais avec les régions flanquantes introniques).

- **Une recherche des réarrangements intragéniques ou extragéniques par technique MLPA pour les patients négatifs en Sanger.**

Dans le cas de *SPRED1*, le séquençage rapide et peu coûteux est effectué avant la MLPA, car plus économique (la proportion de CNV étant relativement faible).

Vérification par le médecin des documents nécessaires à l'analyse du gène SPRED1: liste de proposants NF1-négatifs

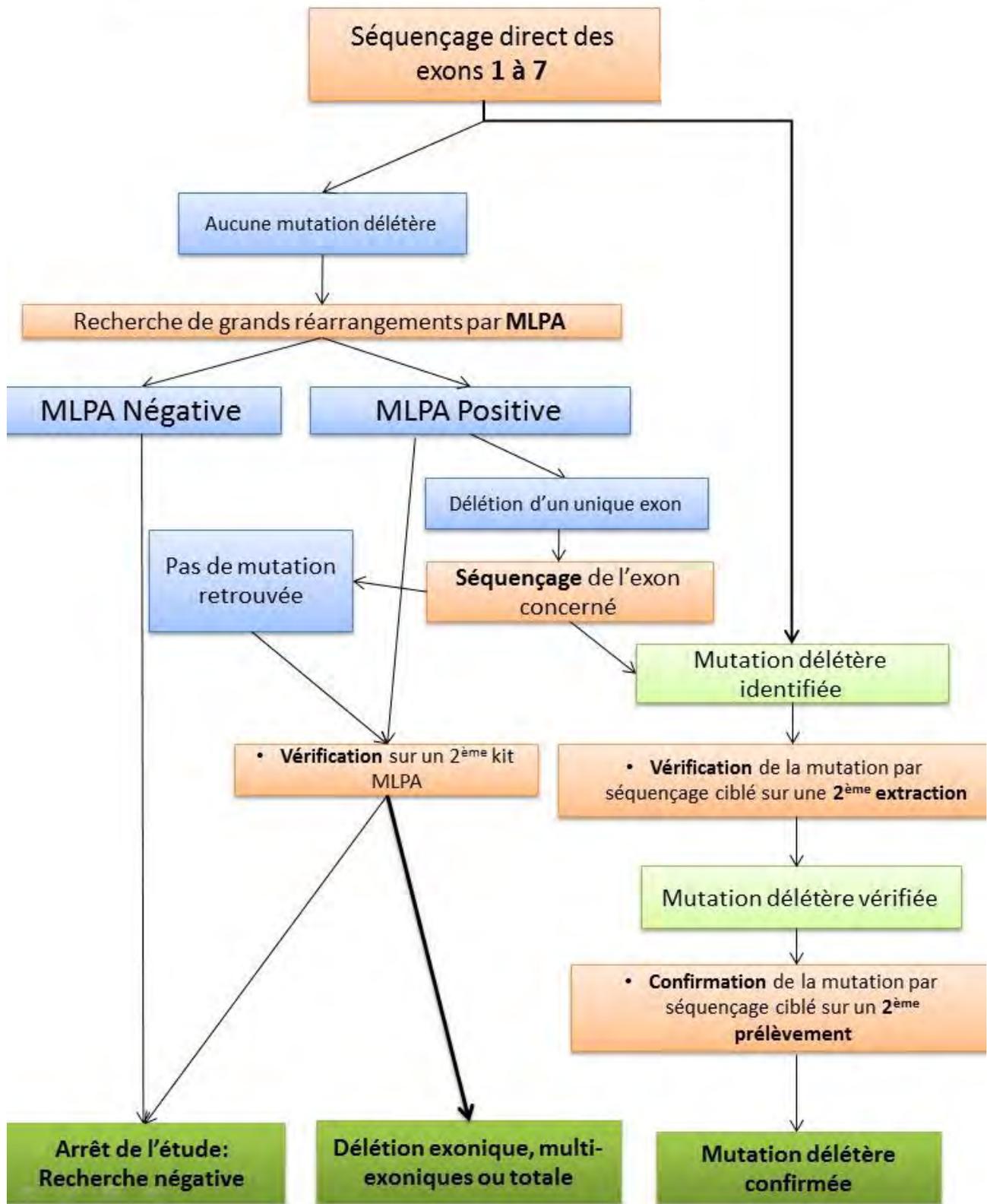
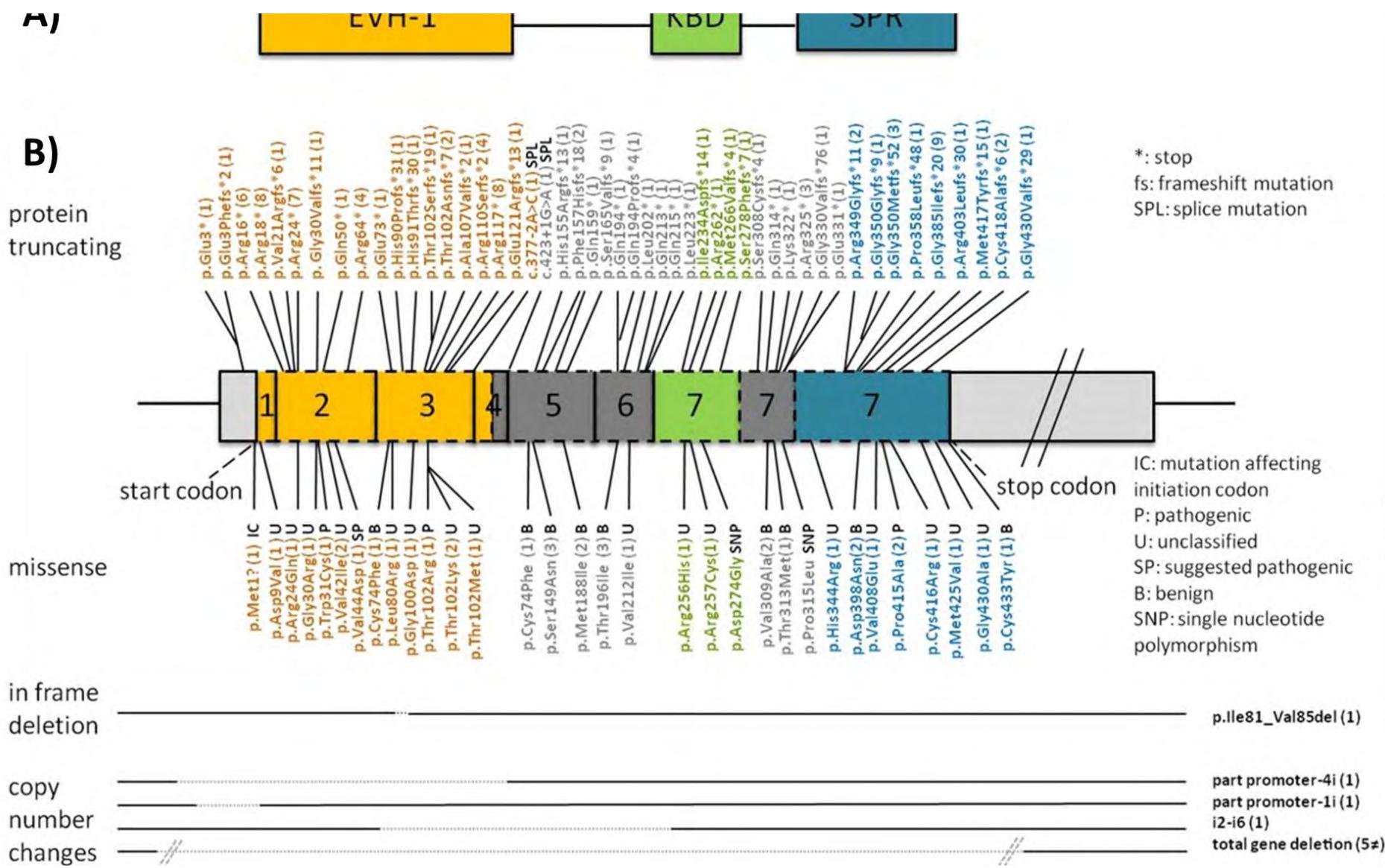


Figure 23. Stratégie du laboratoire pour l'analyse du gène SPRED1.

Figure 24. Schéma du gène SPRED1 et ses altérations génétiques (Brems et al, 2013¹).



123

Légende : A : domaines fonctionnels du gène SPRED1 (EVH-1 : domaine EVH-1 ; KBD : domaine c-KIT-binding ; SPR : domaine SPROUTY-related) B : mutations et CNV associés et leur localisation sur les 7 exons du gène : les mutations donnant une protéine tronquée sont listées au-dessus (* : codon stop ; fs : mutations frameshift ; SPL : mutations touchant le site d'épissage). les mutations faux-sens et les délétions sont listées en-dessous (IC : mutations faux sens affectant le codon d'initiation ; P : mutations pathogéniques prouvées ; U : mutations non classées ; SP : mutation suggérée pathogénique ; B : variant bénin ; SNP : polymorphisme. Le nombre entre parenthèses indique le nombre de patients indépendants porteurs de la mutation

5. NGS

5.1. INTRODUCTION

Le projet de séquençage du génome humain fut accompli de manière partielle en 2001^{150,151} grâce à un travail collaboratif de nombreuses dans des institutions publiques et privées. Réalisé par séquençage Sanger, ce premier séquençage encore incomplet a coûté plus de 2 milliard de dollars.

Le séquençage Sanger, qui demande de nombreuses interventions manuelles, limite la réalisation du séquençage d'un génome complet et son utilisation à large échelle. Après la publication du génome humain en version « complète » par le International Human Genome Sequencing Consortium en 2004¹⁵², le National Human Genome Research Institute (NGHRI) fonda une initiative dotée de 70 millions de dollars ayant pour but de promouvoir le séquençage d'un génome humain pour 1 000 \$ à l'horizon de 10 ans. Grâce à cet élan impulsé par les agences publiques (Américaines en particulier), de nombreuses technologies de séquençage dites à « haut-débit » (*high-throughput sequencing* (HTS)) apparurent. Cet avènement fut permis grâce aux développements et améliorations: des techniques microfluidiques, de nouvelles techniques d'amplification clonale et de l'informatique (puissance des microprocesseurs, et développement d'algorithmes nouveaux).

Ces nouvelles technologies permettent de séquencer en quelques jours de nombreuses régions cibles comme un **panel** de gènes (*targeted-sequencing* (TS)), toutes les régions codantes (exome ou *whole-exome sequencing* (WES)) voir l'ensemble du génome (génome entier ou *whole-genome sequencing* (WGS)).

Ce séquençage haut-débit ou massivement parallèle (*massive parallel sequencing* (MPS)) permet de réaliser le séquençage de plusieurs patients en une seule série (ou *run*), dans un délai minime. Une impressionnante masse de données est générée par ce nouveau mode de séquençage. Il pose un défi nouveau aux généticiens. Plus qu'auparavant, ils doivent s'intégrer dans une équipe aux compétences multiples, afin de traiter et analyser ce flux massif de données. Les sciences de la vie se voient donc transformer depuis quelques années par cette révolution moins coûteuse et plus exhaustive. De grandes avancées sont réalisées, de la recherche fondamentale à la médecine, avec une évolution rapide des pratiques et des connaissances.

5.2. UN SCHEMA GENERAL DE FONCTIONNEMENT

De nombreuses plateformes commerciales ont été créées à partir de 2006, avec des différences dans les technologies employées, mais un protocole commun et général de fonctionnement (figure 25).

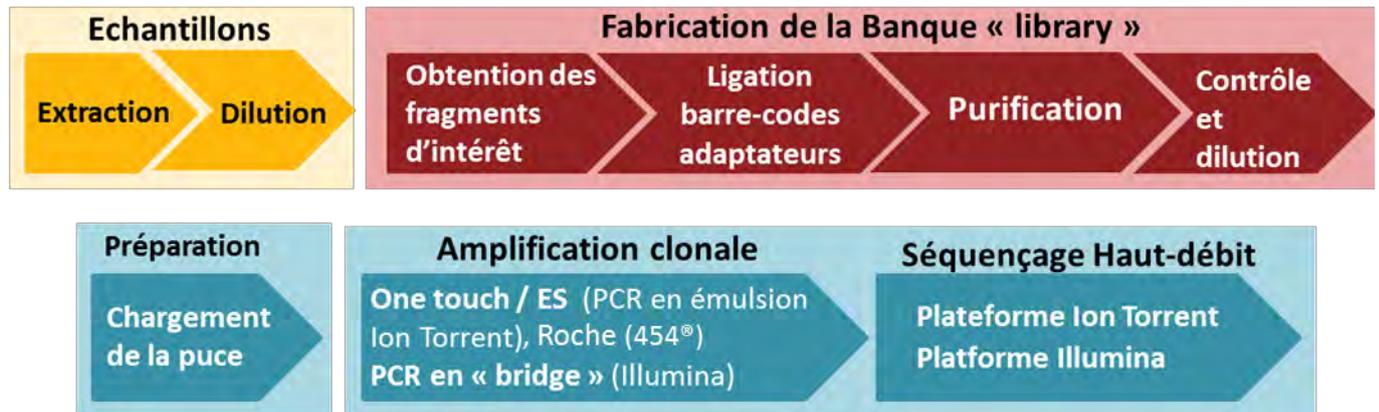


Figure 25. Protocole commun au séquençage de 2ème génération.

Le schéma général de séquençage inclut plusieurs étapes similaires pour les séquenceurs de 2ème génération, avec dans l'ordre :

1. Préparation des échantillons (*template preparation*)

Elle permet la création finale d'une **librairie** (*library*, ou encore « banque ») qui regroupe les fragments d'ADN, ou amplicons, à séquencer.

Cette étape peut utiliser des stratégies variées et indépendantes du fabricant de séquenceur. De nombreuses entreprises de biotechnologie (souvent de petites taille) proposent leurs propres développements, parfois spécifiques d'une plateforme, afin de sélectionner les régions d'intérêts que l'on veut séquencer. La sélection de ces régions commence par une PCR multiplexe, une fragmentation de l'ADN, ou encore est faite par capture.

Généralement, plusieurs patients sont séquencés en même temps, et les séquences d'ADN sont chargées sur une même puce (*chip*). Les fragments, mélangés, doivent lier à leur bout un « **code-barre** » (*barcode*). Ce sont des séquences nucléotidiques de petite taille (≈ 20 nt), uniques et qui sont attribuées unitairement à un seul patient. Les codes-barres se lient à une ou aux deux extrémités d'un

fragment (*single* ou *dual indexing*). Ils permettent lors du traitement bioinformatique, de trier et d'attribuer les séquences au patient source (phase de démultiplexage).

Les fragments d'ADN doivent aussi acquérir des **adaptateurs** (*adapters*) qui sont nécessaires aux étapes suivantes d'amplification clonale et de séquençage.

Ce sont des courtes séquences nucléotidiques qui sont liées aux extrémités 5' et 3' de tous les fragments. Ces adaptateurs sont identiques pour tous les patients (figure 26).

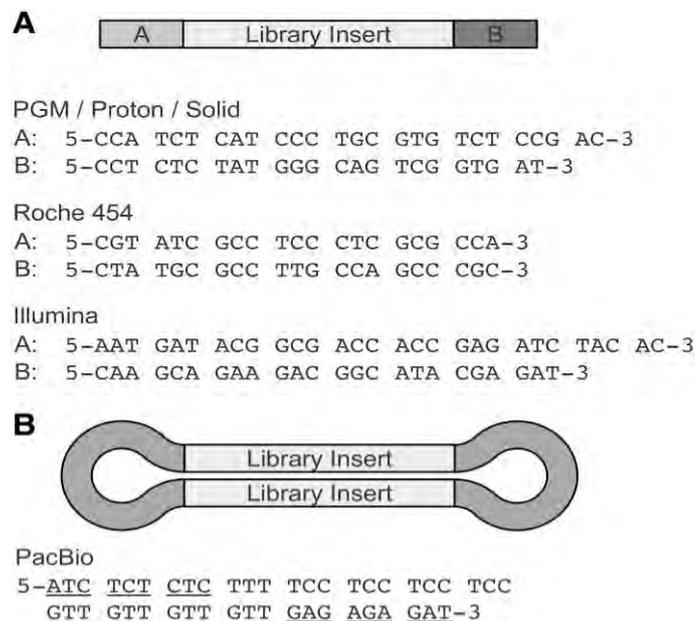


Figure 26. Séquence des adaptateurs de plusieurs plateformes¹.

Après la ligation des codes-barres et adaptateurs, une étape de purification est nécessaire afin d'éliminer les résidus d'ADN génomique, d'amorces, de nucléotides, de barre-codes et d'adaptateurs, et d'enzymes.

Cette purification se fait grâce à des billes magnétiques sur lesquelles les amplicons viennent s'hybrider. Elles permettent de sélectionner par la taille les fragments d'intérêt selon leur concentration.

Le lavage des billes ayant lié des fragments est réalisé par de l'éthanol.

Les fragments purifiés de tous les patients sont ensuite rassemblés de façon équimolaire dans un seul tube (*pooling*), qui est la banque finalisée.

2. **Amplification clonale** (*clonal amplification*).

Cette étape est dépendante de la plateforme de séquençage. Elle se fait soit directement dans le séquenceur (Illumina), soit de manière externe (One Touch™ de Life Technologies).

Elle permet l'amplification du signal (fluorescence pour la plupart) émis lors du séquençage en multipliant la quantité de fragments d'ADN purifiés composant la librairie.

L'amplification est dite « clonale » car les fragments initiaux vont générer un grand nombre de leur propre copie. Lors du séquençage, les signaux seront suffisants pour être détectés et capturés.

3. **Séquençage en parallèle par répétition de cycles** (*cyclical rounds of massively parallel sequencing*).

Cette étape est également dépendante du séquenceur. Elle se fait de manière parallèle et simultanée pour tous les patients. Les fragments simples brins sont immobilisés sur un support. La génération du brin complémentaire par polymérisation se fait par incorporation de base grâce à une polymérase. Une étape de lavage (wash) permet d'éliminer les résidus, puis le signal émis à l'incorporation de chaque base, est capturé (scan).

Le séquençage de deuxième génération se fait donc par polymérisation grâce à une succession d'étapes de *wash-and-scan*.

4. **Analyse des données, étape bio-informatique**

Cette étape est dépendante des caractéristiques du séquenceur et de la technologie choisie pour créer la banque. Elle permet de traiter les signaux, de filtrer, de trier et d'annoter les données permettant au biologiste d'interpréter les variants retrouvés.

Cette étape cruciale est peut être manuelle ou automatisée (utilisation de *pipelines* automatisés). Elle demande des compétences en programmation et en bio-informatique. Elle sera présentée plus en détail au chapitre 5.5.

La stratégie de fabrication de banque et les plateformes de séquençage sont basées sur des propriétés spécifiques selon les applications voulues. La combinaison des deux détermine la qualité, la quantité, et les biais inhérents aux séquences résultantes, et conditionne la qualité des résultats finaux.

5.3. LES PLATEFORMES ACTUELLES

De nombreuses plateformes, opérantes depuis quelques années déjà ne seront pas détaillées ici: le séquenceur 454 de Roche et sa technologie issue du pyroséquençage¹⁵³ ; le séquenceur Heliscope de Helicos¹⁵⁴ ; la technologie SOLiD de ABI¹⁵⁵ et d'autres. Ces machines furent parmi les premières commercialisées et malgré certains atouts, deviennent déjà obsolètes (Roche a annoncé l'arrêt du suivi de sa plateforme 454 pour 2016). Nous détaillerons la spécificité des plateformes les plus communément utilisées à ce jour, ainsi que les récents développements.

5.3.1. LES PLATERFORMES ILLUMINA



Figure 27. Séquenceurs de 2ème génération d'Illumina en 2016 (hors MiniSeq).

Depuis la commercialisation du Genome Analyzer II en 2006, la société Illumina/Solexa augmenta largement le débit tout en réduisant les coûts de ses machines. Illumina est à ce jour un acteur largement dominant sur le marché du NGS.

Une large gamme existe (gammes MiniSeq, MiSeq, NextSeq 500, HiSeq), les séquenceurs sont produits et optimisés selon le débit et l'utilisation voulue.

Parmi les premiers automates : le MiSeq est pensé pour être un automate de paillasse rapide, avec des runs courts (minimum de 4h), pour des applications de séquençage ciblé ou le séquençage de petits génomes (infectiologie). Le HiSeq 2500 à l'inverse, trouve son utilité pour les applications de très haut débit (jusqu'à 1 Tb en 6 jours).

Depuis 2014, Illumina produit deux nouveaux séquenceurs : le NextSeq 500 et le HiSeq X Ten. Le NextSeq 500, proche du MiSeq, est un automate de paillasse permettant des runs courts et rapides mais avec un meilleur débit (jusqu'à 120 Gb, permettant le séquençage de génomes complets en moins de 30h).

Le NextSeq 500 se distingue par sa stratégie de détection à deux longueurs d'ondes (*two-channel*) au lieu de quatre jusqu'à présent pour les autres séquenceurs:

- La cytosine est marquée en rouge.
- La thymine en vert.
- L'adénine en jaune (mélange rouge et vert).
- La guanine n'est pas marquée.

Cette stratégie permet une réduction du temps d'acquisition (deux images requises au lieu de quatre) et du traitement de l'image. Le débit du séquençage est donc augmenté avec un taux d'erreurs similaire à la stratégie *four-channels*.

Quant au HiSeq X Ten, il est utile dans les projets de séquençage de génomes complets (WGS) avec des performances très haut débit: 1.8 Tb en trois jours ou 18 000 génomes par an (couverture de 30X).

Absent de la figure 27, le dernier automate de la gamme est le MiniSeq™. Avec un débit moindre que le MiSeq mais pour un prix divisé par deux, il est disponible depuis début 2016.

5.3.1.1. La technologie Illumina

L'amplification clonale se fait par « *bridge-PCR* » de fragments d'ADN liés à des adaptateurs¹⁵⁶. La stratégie de séquençage implique une terminaison cyclique et réversible du nucléotide séquencé (*Cyclic Reversible Termination* ou CRT). Le séquençage se fait par les deux extrémités des fragments d'intérêt.

Amplification clonale par *bridge-PCR* :

L'étape d'amplification clonale se fait à l'intérieur de l'appareil, après chargement sur la puce (*flowcell*) des fragments d'intérêt auxquels sont fixés les barre-codes et adaptateurs.

La *flowcell* est composée d'une lame de verre sur laquelle sont fixés des adaptateurs complémentaires à ceux liés aux extrémités des fragments d'ADN. Lors du chargement de la librairie sur la puce, les fragments d'ADN s'accrochent de manière aléatoire. La distance entre deux brins d'ADN se fixant sur la *flowcell* est dépendante de la quantité de brins chargés et suit un modèle probabiliste.

Une fois les brins fixés, l'amplification clonale est réalisée sur milieu solide par un type de PCR particulier : la *bridge*-PCR. Elle se déroule en plusieurs étapes, répétées en cycles :

- L'adaptateur de l'extrémité non fixée sert d'amorce à la synthèse du brin complémentaire par la polymérase. On obtient un ADN double brin.
- Une étape de dénaturation permet l'élimination du brin fixé initialement.
- Le brin synthétisé à la première étape se courbe grâce à la complémentarité de l'adaptateur fixé sur l'extrémité libre, formant ainsi un « pont ».
- La synthèse du brin complémentaire se fait à nouveau.
- Une étape de dénaturation permet d'obtenir deux copies simple brin, indépendant.

Au bout d'un certain nombre de cycles, on obtient un grand nombre de copies du brin initialement fixé sur la flowcell, dans un voisinage proche. On parle alors de la création d'un « *cluster* ». Les *clusters* contiennent une séquence d'ADN unique, portée par de multiples copies, à un endroit concis.

Le séquençage Illumina :

Après la génération de *clusters*, tous les brins *reverse* sont éliminés, laissant ainsi les brins *forward*, sur lesquels la première étape de séquençage va se faire. Une fois terminée, un pont est à nouveau fait par les brins *reverse* initiaux (le brin nouvellement synthétisé a été éliminé), et le séquençage par polymérisation des brins *reverse* peut commencer.

La détection des nucléotides, un à un, se fait par cycles qui regroupent les étapes suivantes:

- Incorporation de la base (grâce à une polymérase).
- Lavage des bases en excès.
- Capture d'image (fluorescence).
- Clivage chimique et initiation du cycle suivant.

En résumé, à l'incorporation d'une base, la polymérisation est mise en pause par des 3'-O-azidomethyl-dNTPs couplés à un fluorophore. Cette pause permet un lavage et donc l'élimination des bases non incorporées ainsi qu'une capture optimale de l'image de la puce (scan de la *flowcell*) par une caméra CCD (coupled-charge device). Le fluorophore et le blocage en 3' sont ensuite éliminés par clivage chimique (hydrolyse) et le processus est ainsi répété.

L'image capturée des *clusters* est analysée, et leur position sur la puce est calculée. Pour chaque image, à chaque position de cluster, la base incorporée est déterminée. La succession de

cycles détermine la séquence d'intérêt, qui est contenue dans un même et unique *cluster*. Une flowcell peut produire des millions de *clusters*.

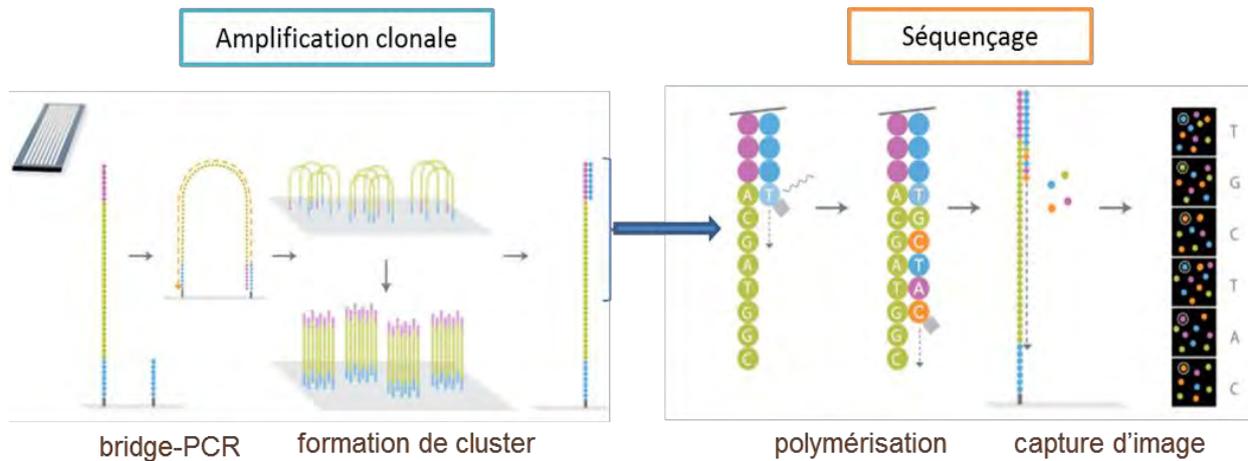


Figure 28. Technologie Illumina: amplification clonale et séquençage (Illumina NGS Methods Guide).

5.3.1.2. Avantages et inconvénients

Les avantages de la technologie Illumina sont multiples :

- Le **temps technique** pour la préparation des bibliothèques est **minime**, car l'amplification clonale se fait directement à l'intérieur du séquenceur. Des cartouches avec tous les réactifs sont vendues prêtes à l'emploi, ce qui facilite la préparation du run.
- **Débit** le plus important actuellement.
- Les séquences obtenues sont de **grande qualité**, grâce à l'arrêt de l'élongation du brin séquençé qui permet une capture précise de l'image de la puce.
- Le **coût** par base est devenu **compétitif**, actuellement le moins chère du marché.
- La **gamme** illumina est **diversifiée**, permettant de choisir le séquenceur dont le laboratoire a besoin selon son débit et l'utilisation escomptée.

Les points négatifs sont :

- La **taille** des séquences est relativement **courte**, pas plus de 200 pb pour beaucoup d'automates (300 pour le MiSeq).
- Technologie de « wash-and-scan » : le **temps** nécessaire pour un run est **important** (jusqu'à 60h) à cause de la capture de l'image de la puce. Les séquenceurs sont ainsi vite saturés, et la fréquence des runs est limitée.

5.3.2. LES PLATEFORME LIFE TECHNOLOGIES/THERMOFISHER (ION TORRENT)



Figure 29. Séquenceurs de 2ème génération de Life Technologies en 2016.

Life Technologies commercialise depuis 2010 plusieurs séquenceurs utilisant la technologie par semiconducteur Ion Torrent. Le séquenceur de paillasse avec un débit moyen est appelé PGM™. La préparation de la librairie et les étapes du séquençage sont conceptuellement proches de la plateforme de pyroséquençage du 454 (Roche®)¹⁵⁷.

Un second séquenceur commercialisé en 2012, le Ion Proton, a un débit bien supérieur au PGM™ (2 Gb pour le PGM™ contre 10 Gb pour le Proton). Les fragments séquencés ont cependant une taille inférieure à ceux du PGM™ (200 pb contre 400 pb pour le PGM™). Le PGM™ est surtout utilisé pour de petits projets de séquençage ciblé (*targeted sequencing*) alors que le Proton est capable de réaliser les projets d'exome ou de petit génome.

Le travail présenté dans cette thèse a pu être réalisé au CHU de Lyon avec le séquenceur de paillasse PGM™.

La dernière génération est apparue au court de l'année 2015, avec les séquenceurs S5 et S5 XL. Ce sont des versions améliorées des séquenceurs précédents, mais avec une prise en main plus simple, et un débit plus élevé. La grande nouveauté vient de l'utilisation d'une cartouche contenant tous les réactifs, similaires à celles employées par Illumina. Les nouvelles puces ont une capacité augmentée, et la durée du séquençage est diminuée. Il est possible d'avoir des résultats d'un run en 24h, ce qui est utile dans les applications cliniques qui demandent une prompt réactivité des services (en infectiologie par exemple¹⁵⁸).

L'amplification clonale est réalisée de préférence sur l'automate Ion Chef, qui réalise toutes les étapes jusqu'au chargement de la puce de manière autonome. D'après le site internet du constructeur, la durée des manipulations nécessaires n'excéderait pas 45 min par run. L'utilisation manuelle du One Touch™ reste possible.

5.3.2.1. La technologie Ion torrent

L'Amplification clonale par PCR en émulsion :

L'étape d'amplification clonale se fait par une stratégie novatrice *d'émulsion-PCR* (PCR en émulsion). Cette stratégie a besoin de billes (*beads* ou *Ion sphere particles (ISP)*), sur lesquelles sont liés les fragments ADN à amplifier. Les billes sont ensuite emprisonnées dans des émulsions H/E (huile dans eau). Plusieurs configurations sont possibles :

- Une seule bille et un seul fragment d'ADN doivent se retrouver dans la gouttelette d'émulsion, permettant une réaction PCR individualisée (chaque goutte est son propre microréacteur). Dans cette situation idéale, de nombreuses copies du fragment intégré sont produites et recouvrent l'ISP : elle est dite **monoclonale**.
- Si deux fragments d'ADN se trouvent dans une même gouttelette, l'ISP aura des fragments liés qui n'ont pas la même séquence, ce qui perturbera le séquençage (signaux émis discordants). L'ISP ne sera pas valide, elle est dite **polyclonale**. Cette polyclonalité diminue la spécificité de l'analyse et consomme la place (des puits) sur la puce.
- Si aucun fragment ne rentre dans la gouttelette, aucune amplification ne se fera : l'ISP sera **non clonale**. Les deux dernières configurations sont donc à éviter le plus possible. Le ratio fragment d'ADN / billes nécessaire est précis (1/1) et l'étape de fabrication de l'émulsion est très importante et délicate. Cette étape est automatisée grâce à l'automate OneTouch2.

Après l'émulsion, l'amplification clonale peut commencer. Elle nécessite pour cela plusieurs étapes (figure 30) :

- Grâce aux adaptateurs A et P1, la ligation d'une biotine et d'un adaptateur B par PCR est réalisée (fusion-PCR). L'amorce A permet l'amorçage de l'élongation, et l'addition d'une biotine à son extrémité. L'amorce P1 elle, est suivie de la séquence supplémentaire B, qui permet aux amplicons de s'accrocher aux ISPs.
- Une fois les amplicons liés aux ISPs, la séquence B sert d'amorce et une nouvelle réaction PCR se déroule à la surface de l'ISP. Une phase de dénaturation est nécessaire pour décrocher le brin nouvellement synthétisé. Il se fixe à nouveau

sur l'ISP et un nouveau cycle PCR commence. A la fin, de nombreux fragments nouvellement synthétisés recouvrent les ISPs (polyclonale ou monoclonale).

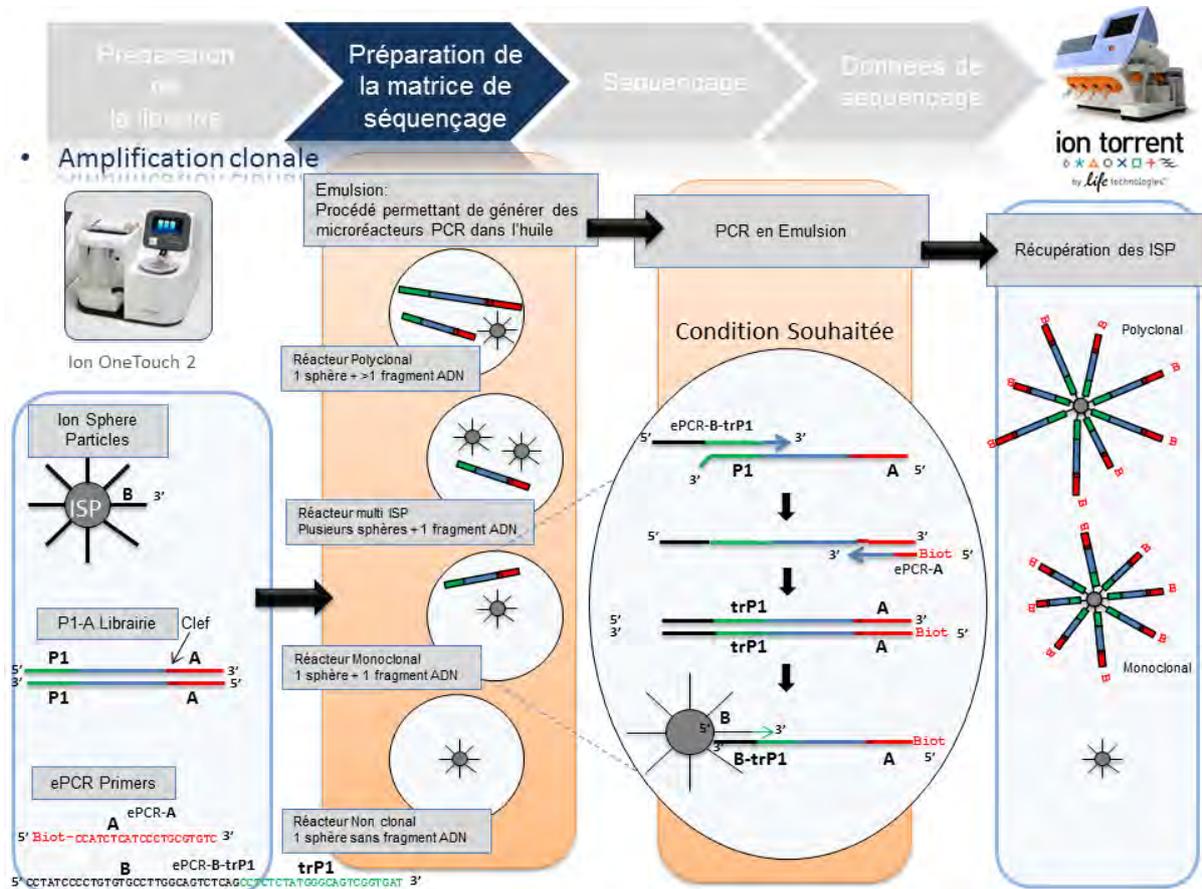


Figure 30. Amplification clonale par PCR en émulsion (<http://www.biorigami.com/>).

Enrichissement des ISPs :

Après l'étape de PCR en émulsion, les billes ayant fixées des amplicons doivent être sélectionnées. Après cassage de l'émulsion, des billes magnétiques « coatées » avec de la streptavidine sont rajoutées. Par une réaction biotine-streptavidine, seules les ISPs avec des fragments ADN sont récupérées à l'aide d'un aimant. Les ISPs non clonales sont éliminées (seules restent les mono et polyclonales).

Cette étape d'enrichissement est réalisée par l'automate Ion OneTouch ES (*Enrichment System*) (figure 31).

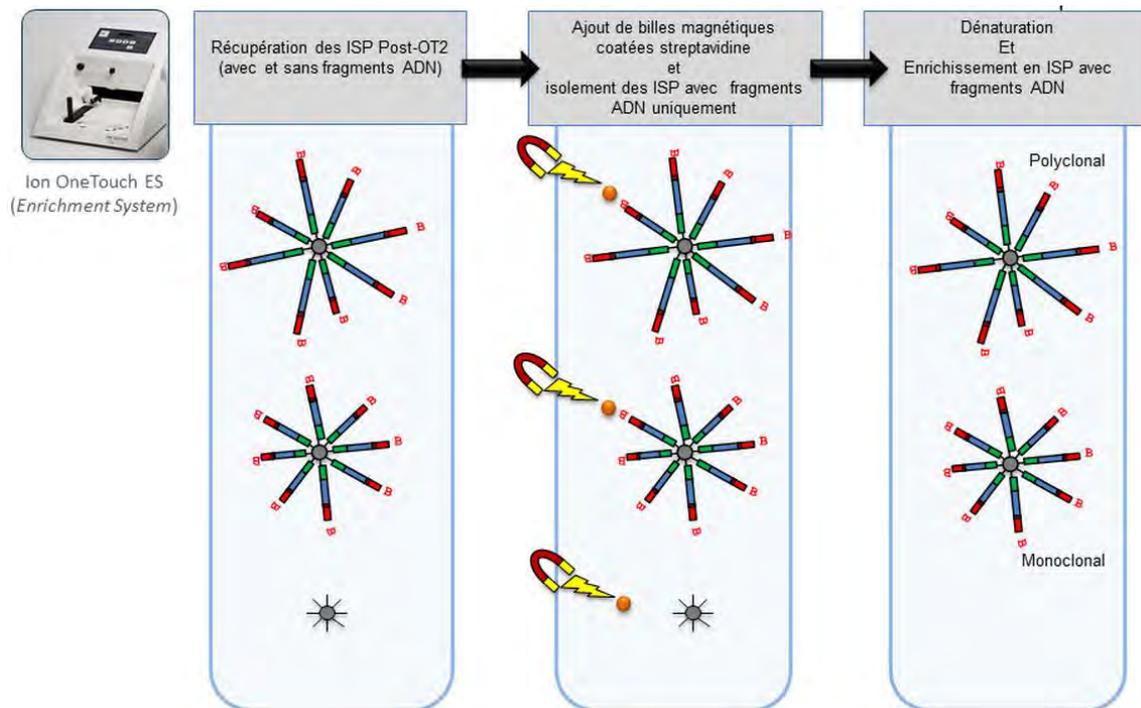


Figure 31. Enrichissement des ISPs. (<http://www.biorigami.com/>)

Contrôle de l'amplification clonale :

Afin d'estimer l'efficacité de l'amplification clonale et de l'enrichissement, un contrôle est effectué pour estimer la quantité d'ISP qui sont couvertes par des amplicons.

Pour cela, deux sondes marquées par un fluorochrome issu de la gamme Alexa Fluor (produit d'une filiale d'Invitrogen) sont utilisées. La première sonde, l'AF488, est complémentaire de la séquence B qui se retrouve sur toutes les billes. Le signal à 488 nm (cyan-vert) permet donc de quantifier toutes les billes présentes, avec ou sans amplicon. La seconde sonde, l'AF647, est complémentaire de l'adaptateur A, séquence à l'extrémité libre de tous les amplicons. Après lavage des billes, le signal à 647 nm (rouge) permet de quantifier toutes les ISP qui ont fixées des amplicons.

La lecture du signal des deux longueurs d'onde est réalisée grâce au fluorimètre Qubit®. Le rapport d'intensité : $\frac{647}{488} nm$ permet l'estimation du pourcentage d'ISP ayant fixées un ou plusieurs amplicons. Le contrôle du pourcentage d'ISP monoclonale n'est pas possible à cette étape.

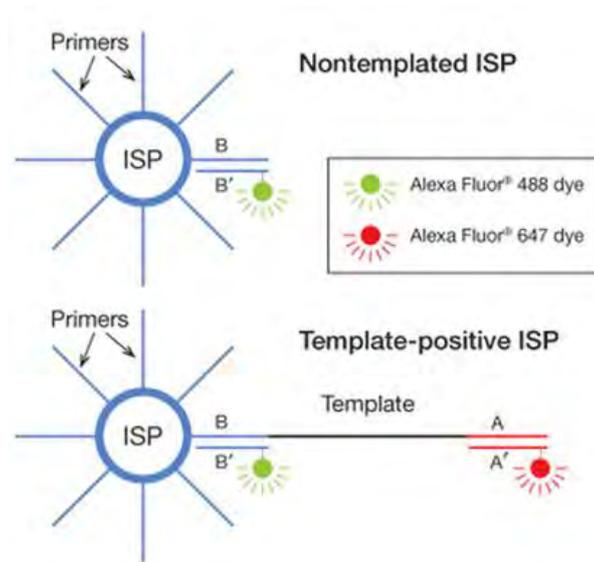


Figure 32. Principe de dosage des ISPs.

Le séquençage Ion Torrent :

Les billes chargées sont ensuite distribuées unitairement dans des micropuits (*microwells*) dans lesquels est réalisé le séquençage **par synthèse** (*sequencing-by-synthesis*). Le séquençage Ion Torrent fait appel à la technologie des semi-conducteurs CMOS (*Complementary Metal Oxide Semiconductor*). Ils permettent la détection d'une variation de pH induite par la libération d'ion hydrogène H⁺ (proton) lors de l'extension du brin d'ADN (élongation)¹⁵⁹.

Au niveau de chaque puits (une puce peut en avoir plusieurs millions), ce changement de pH est converti en signal voltaïque. Le signal électrique ainsi créé est proportionnel au nombre de bases incorporées lors d'un flow. L'addition séquentielle de nucléotides dans un même puits, permet la discrimination des bases incorporées et donc *in fine*, le séquençage des amplicons.

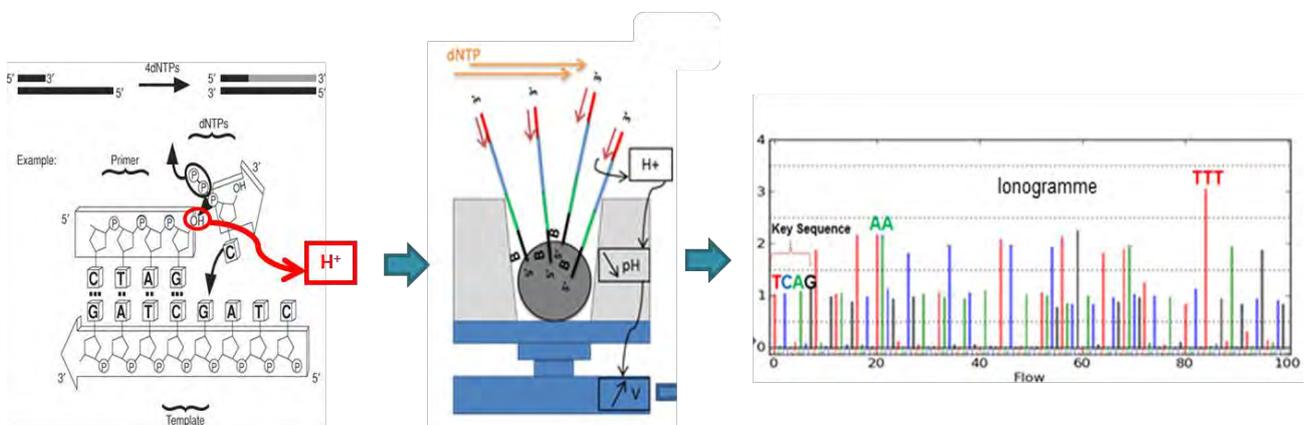


Figure 33. Schéma conceptuel du séquençage Ion torrent.

5.3.2.2. Avantages et inconvénients

Les avantages de la technologie Ion Torrent sont :

- Elle ne nécessite **pas d'étapes de scan** de la puce (capture d'une image), ce qui réduit drastiquement le temps du run.
- Le coût par base est modéré.
- Les runs sont **courts** (2 à 8h).
- Taille des séquences jusqu'à **400pb**.

Les points négatifs sont :

- De multiples étapes manuelles (semi-automatisables) nécessaires pour l'amplification clonale et le chargement de la puce.
- Un fort taux d'erreur dû à la technologie, surtout dans les régions **d'homopolymères**.
- Débit moyen, plus faible qu'Illumina (tableau 11).

5.3.3. PERFORMANCES SPECIFIQUES AUX DEUX TECHNOLOGIES :

Selon les désirs des équipes, il existe plusieurs puces aux capacités différentes. Elles permettent un confort d'utilisation et un ajustement aux besoins (voir tableau 11).

Des problématiques de séquençage sont communes aux deux technologies : les régions fortement répétées (triplets, homopolymères) et les régions riches en nucléotides G et C

Les insertions et délétions sont les erreurs les plus fréquentes de la plateforme Ion Torrent¹⁶⁰. Bien connue, cette erreur est inhérente à la technologie. Elle découle d'une corrélation imparfaite entre le nombre de bases incorporées et la différence de voltage générée.

A partir d'un certain nombre de bases répétées insérées, la différenciation se fait mal. Le taux d'erreurs augmente ainsi drastiquement dans les répétitions homopolymériques (à partir de 6 bases répétées). Ces erreurs sont plutôt constantes pour une même région.

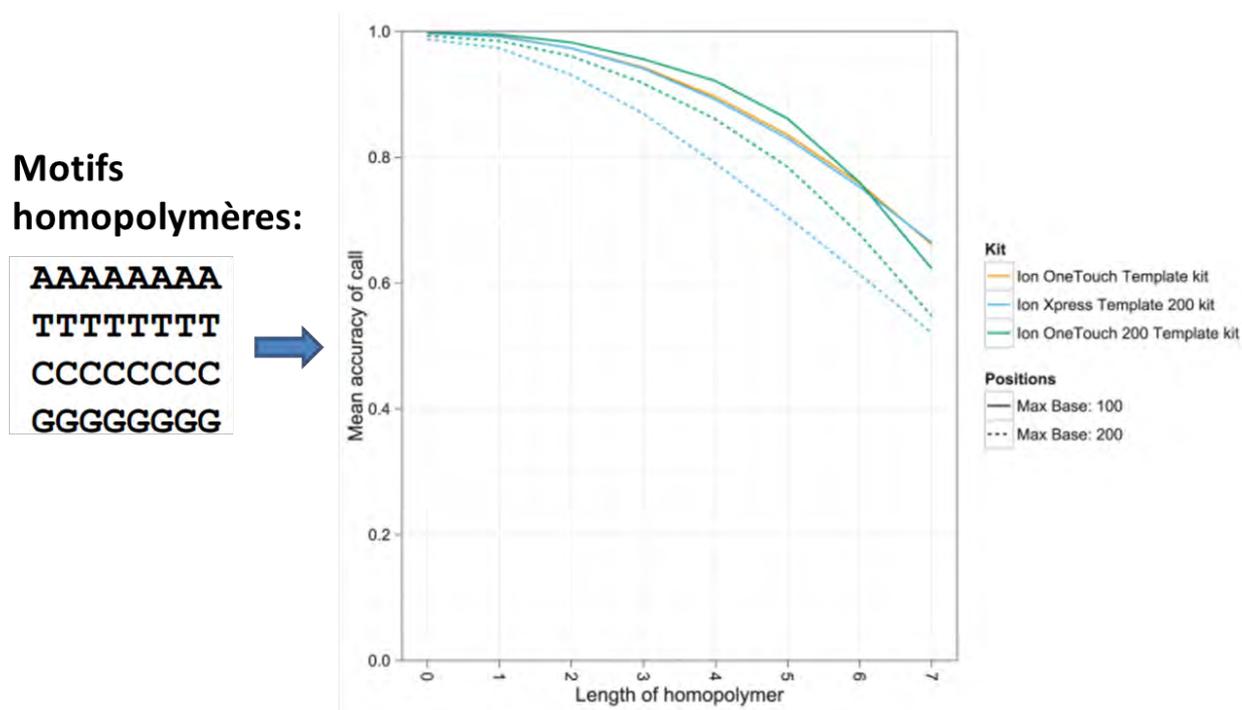


Figure 34. Etude sur plateforme Ion Torrent: précision moyenne des variants appelés en fonction de la longueur de l'homopolymère (Bragg LM et al, 2013).

Pour Illumina, les erreurs les plus fréquentes sont de type « substitution ». Elles restent à un taux minime, et sont produites de manière aléatoire tout le long des séquences. Dans de longs homopolymère (>9 nucléotides), un surplus d'erreurs apparaît aussi. Le mauvais point d'Illumina par rapport à Life Technologies se situe sur la longueur des séquences produites, seulement 300 pb uniquement pour le MiSeq, les autres plateformes se contentent de lectures (ou *reads*) de 150 pb.

| Critère | Plateforme Illumina | | | Ion Torrent | | | | PacBio Rs |
|--|--|-------------|---------|--|---------------|----------------|-------|---|
| | MiSeq | NextSeq 500 | HiSeq X | Ion PGM™ | Ion Proton™ | S5 | S5 XL | PacBio Rs II |
| Séquenceur | | | | | | | | |
| Technologie d'Amplification Clonale | Bridge PCR | | | PCR en émulsion | | | | / |
| Technologie de séquençage (signal) | Terminaison réversible de l'élongation | | | Semi-conducteur (variation de pH) | | | | Séquençage en temps réel |
| Durée du run de séquençage | 4- 24 h | 4 -55 h | < 3 J | 2.3 – 7.3 h | | 2.5 -4 h | | 20 min |
| Nombre de barre-codes disponibles (Multiplexage) | 1 – 96 (combinaison d'adaptateurs) | | | 1 - 96 | | | | 1 - 48 |
| Longueur de lecture (pb) | 2 x 300 | 2 x 150 | 2 x 150 | 200 – 400 | | | | ≈ 10 -20 kb |
| Nombre de séquences lues | 25 M | 400 M | 6 G | 0.4 – 80M | | | | / |
| Données générées (nombre de bases séquencées) | 15 Gb | 120 Gb | 1800 GB | 0.05 – 2 Gb | Jusqu'à 10 Gb | 0.6 Gb – 15 Gb | | Jusqu'à 1 Gb |
| Type d'erreurs | Substitution | | | Indels, régions d'homopolymères | | | | Indels |
| Taux d'erreurs global | 0.32 % - 0.8 % | | | ≈ 1 % avec la chimie précédente / ≈ 0.4% avec chimie Hi-Q | | | | Très dépendante de la couverture et du traitement bioinfo |
| Coût / Mb | Faible (0.007 – 0.25 \$) | | | Moyen (0.8 – 3 \$) | | | | Moyen (1\$) |

Tableau 11. Comparaison de deux technologies de séquençage de 2ème génération et une de 3ème génération

5.4. SEQUENÇAGE DE 3EME GENERATION

Des séquenceurs dits de «3^{ème} génération » sont commercialisés depuis plusieurs années. Leur caractéristique fondamentale est qu'aucune phase d'amplification clonale n'est nécessaire. Le séquençage se fait directement à partir de l'ADN, après fragmentation en long fragments et ligation des barcodes.

La création de banque est donc plus simple et plus rapide. De plus, les erreurs d'amplification clonale sont évitées (erreurs dues à la polymérase notamment).

Deux exemples de machines performantes avec des technologies innovantes sont rapidement développés ici.

5.4.1. PACIFIC BIOSCIENCES :

Un des pionniers du séquençage de 3^{ème} génération avec une première machine sortie en 2010, avec en 2013 une version améliorée : le RS II.



Figure 35. Séquenceur PacBio RS II.

5.4.1.1.

5.4.1.2. La technologie Pacific Biosciences

La technologie du RS II permet le séquençage en temps réel d'une unique molécule d'ADN (*Single-molecule real-time* ou SMRT).

Fabrication des banques :

La préparation de la librairie implique la ligation de barcodes en épingle à cheveux (*ligated hairpin*) à des fragments d'ADN simple brin issus d'ADN digéré ou d'ADNc.

Les SMRT-*bell* ainsi créés sont des structures linéaires et circulaires. Le même fragment d'ADN étant séquencé dans les deux sens *forward* et *reverse*. Cette technologie permet le séquençage deux types de fragment d'ADN (figure 36) :

- des fragments de grandes tailles (*Continuous Long Read* ou CLR) avec un seul passage de la polymérase.
- ou des séquences plus courtes mais générées en grand nombre grâce à plusieurs passages de la polymérase.
- Un mélange des deux stratégies est aussi possible.

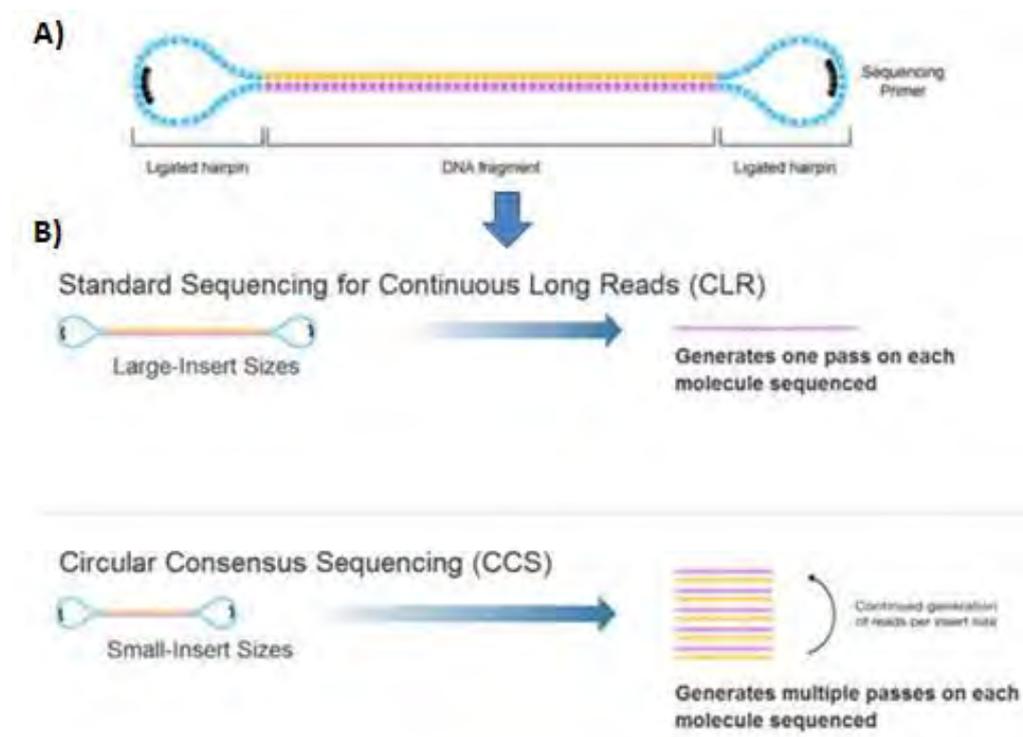


Figure 36. SMRTbell de Pacific Biosciences. (source : www.pacb.com).

Légende :

- A) Séquence d'ADN avec ses adaptateurs
- B) Les deux stratégies de séquençage possible : création de longues séquences uniques (CLR) ou génération de multiples courtes séquences d'un même brin d'ADN.

Séquençage :

L'amplification clonale n'est pas nécessaire sur les plateformes de 3^{ème} génération.

Le séquençage se fait par **synthèse** dans des puits ayant un volume de l'ordre du zeptolitre (10^{-21} litre) appelés *zero-mode waveguides* (ZMWs) et dans le fond desquels sont immobilisés des polymérases. La polymérisation, par l'incorporation de bases ayant un phosphate marqué, se fait de manière continue. Elle peut être suivie en temps réel par l'enregistrement vidéo du signal fluorescent.

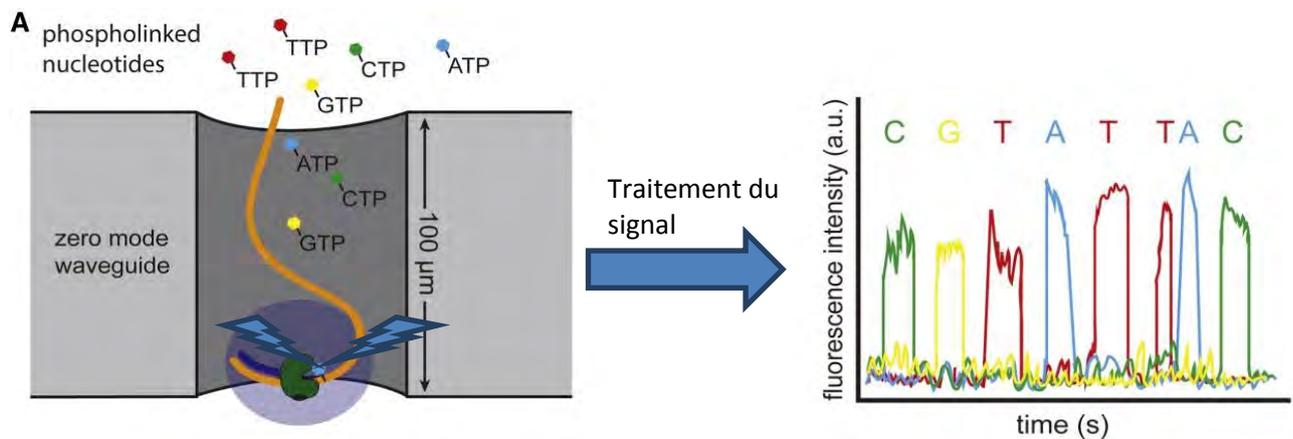


Figure 37. Technologie de la plateforme Pacific Bioscience de 3^{ème} génération.

- A) Une polymérase est ancrée au fond de chaque puits (ZMW). Les quatre versions des nucléotides avec un phosphate-marqué sont présents sur le schéma. Ils permettent la polymérisation en continue du brin d'ADN. L'incorporation de la base augmente son temps passé dans le puits, permettant la détection d'un signal fluorescent qui est capturé par vidéo (en bas).

5.4.1.3. Performances du PacBio RS II

La dernière chimie permet, grâce à des puces avec plus de 150 000 puits, de générer jusqu'à 1 Gb de données en quatre heures. Les séquences peuvent atteindre une taille jusqu'ici jamais atteinte de 60 kb.

Malheureusement, dans cette plateforme de séquençage novatrice, le taux d'erreur reste élevé ($\approx 11\%$), spécialement pour la stratégie des CLR. Ces erreurs, en majorité des indels, sont cependant réparties sur toute la séquence de manière aléatoire. Les séquences étant de très grande taille, une hausse de la couverture ou une stratégie de CCS permettent une précision accrue du séquençage¹⁶¹. Ce taux d'erreurs élevé est aussi contrebalancé par le fait qu'un même fragment d'ADN est séquençé plusieurs fois. Au final, par un traitement informatique adapté, le taux d'erreurs à l'appel de variant est comparable aux plateformes de 2^{ème} génération et proche des 1%.

5.4.1.4. Utilisation de la technologie Pacific Biosciences

En évitant l'étape d'amplification clonale, les séquences riches en GC posent peu de problème d'amplification. Ainsi la première étude de répétition par séquençage haut débit des triplets CGG du gène *FMR1*, responsable du syndrome de l'X fragile (OMIM #300624), s'est faite sur une plateforme RS II ¹⁶².

Dans plusieurs utilisations, la stratégie de CLR est très utile et permet des études jusque-là impossibles avec une étude classique par séquençage Sanger ou par des séquenceurs de 2^{ème} génération. Elle permet de compléter les trous (*gap*) de génomes incomplets, dans des régions inatteignables au séquençage issu de séquences courtes¹⁶³.

L'étude des variants structuraux est aussi facilitée grâce aux amplicons de très grande taille, complétant un peu plus l'étude du génome humain et démontrant un peu plus sa complexité structurale¹⁶⁴.

Des études sur le transcriptome sont parues étudiant la structure entière de longs ARN, impossible jusque-là, permettant une mise à jour majeure des annotations jusqu'ici connues¹⁶⁵.

En plus de produire de longues séquences non biaisées, une caractéristique essentielle du séquençage SMRT est le monitoring en temps réel de la réaction de polymérisation. Il permet la collecte des signaux des bases intégrées ainsi que de leur cinétique enzymatique. Des profils cinétiques différents sont produits par la polymérase selon le type de nucléotide mais aussi selon le changement épigénétique subi par l'ADN. En analysant ces profils, il est possible de produire deux types de données : celles relatives au séquençage des nucléotides, et celles relatives aux sites de méthylation ou d'acétylation de l'ADN¹⁶⁶.

Toutefois, son débit moyen et son coût à la base limite actuellement l'emploi de cette technologie dans les études de génome (tableau 11).

5.4.2. OXFORD NANOPORE

Oxford Nanopore est le leader du développement et de la commercialisation d'une stratégie de séquençage basée sur les nanopores. Cette technologie évolue rapidement et a fait d'importants progrès ces dernières années¹⁶⁷.



Figure 38. MinIon de Oxford Nanopore.

5.4.2.1. Technologie Oxford Nanopore

Le concept du séquençage est basé sur la différence variable de courant ionique que des nucléotides peuvent induire lorsqu'ils passent à travers un canal étroit (pore). La faisabilité a été démontrée dès 1996 par la détection d' α -hémolysine lors de sa translocation à travers un nanopore¹⁶⁸.

Séquençage Oxford Nanopore :

Sur la puce actuelle, des centaines de micro-puits indépendants sont répartis, chacun contenant une membrane double couche synthétique perforée par des nanopores biologiques. Une protéine motrice fait passer la molécule d'ADN à travers les pores, et le changement de courant électrique du au passage des bases est mesuré et traité.

La préparation de la librairie est minimale. Elle comprend la fragmentation de l'ADN et la ligation d'adaptateurs.

La phase d'amplification par PCR n'est pas nécessaire. Le design de la librairie permet le séquençage des deux brins d'une unique molécule d'ADN, ce qui augmente la précision.

Le séquençage se fait comme une lecture continue des bases lors de leur translocation à travers le nanopore. Chaque base possède sa propre signature cinétique, et le signal est détecté en temps réel. Comme pour la technologie Ion Torrent, aucune étape de capture de signal lumineux n'est nécessaire. L'ADN n'est pas altéré et peut être récupéré à la fin du run.

Les cytosines portant un groupement méthyle peuvent aussi être distinguées des autres bases¹⁶⁹. Ainsi, comme pour le séquenceur RS II, des données de séquençage et d'épigénétique peuvent être obtenues en même temps.

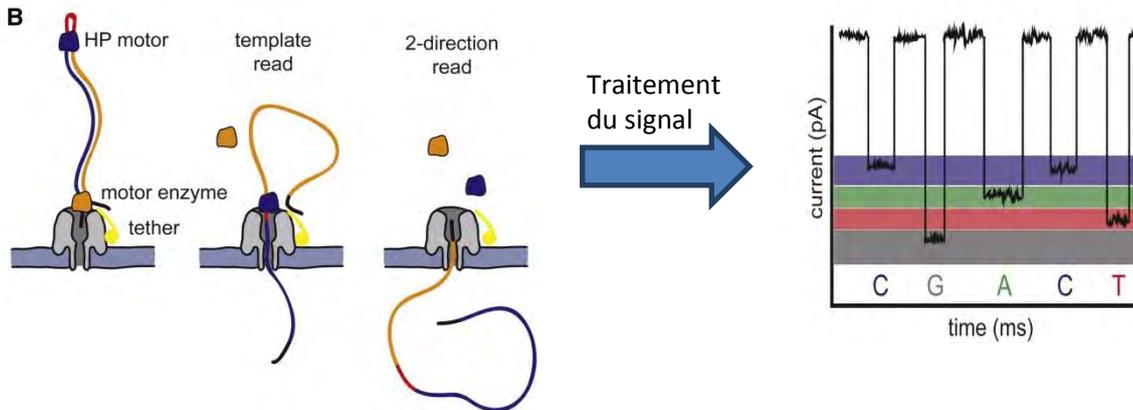


Figure 39. Technologie de la plateforme Oxford Nanopore de 3ème génération.

- B) **Plateforme Oxford Nanopore.** Le brin d'ADN est lié à deux adaptateurs. Le premier adaptateur est lié à une enzyme moteur et à un point d'ancrage. Le deuxième adaptateur est un oligonucléotide en épingle à cheveux qui est lié à la protéine moteur HP. Le changement de courant induit par le passage de nucléotides à travers le pore est utilisé pour la discrimination des bases. Le séquençage des deux brins d'ADN est possible (*two-directions reads*).

5.4.2.2. Performances du MinIon

Depuis 2014, le premier séquenceur sur le marché est le MinION (2000 micropores). La miniaturisation de cet automate est extrême, car le MinION est un automate portable alimenté par une prise USB et tenant dans la main.

Un run de 18h produit plus de 90 Mb de données, avec des séquences d'une taille moyenne de 6 kb mais pouvant aller au-delà de 60 kb.

Le taux d'erreur reste très élevé, pouvant monter jusqu'à 38% d'erreur initiale¹⁷⁰. Les erreurs viennent du chevauchement du signal produit lors du passage successif des nucléotides. En augmentant la force d'entraînement, la résolution peut être drastiquement augmentée et les erreurs de détection évitées¹⁷¹.

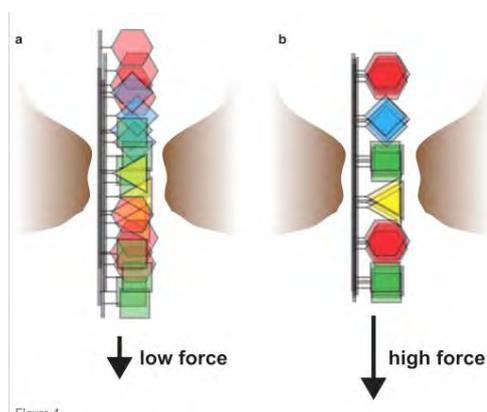


Figure 40. Influence de la force appliquée sur la résolution du séquençage par nanopore.

En améliorant l'analyse du signal (par des algorithmes d'espérance-maximisation), la résolution s'améliore et des taux d'erreur de 4.9 % pour les insertions, 7.8% pour les délétions et 5.1% pour les substitutions ont été rapportés récemment¹⁷².

Malgré les évolutions récentes, le faible débit ne permet pas encore au séquençage par nanopore de concurrencer réellement les autres plateformes de séquençage. Les applications sont similaires à celles du PacBio. Le véritable avantage de cette stratégie est son faible coût.

En conclusion, les séquenceurs de 3^{ème} génération ont un bel avenir devant eux et promettent déjà des performances bien supérieures à celles de la génération précédente. Leur évolution est rapide et une multitude d'autres stratégies existent et s'améliorent rapidement.

Dans un futur proche, les séquenceurs de 3^{ème} génération promettent de combiner :

- une faible taille (celle d'un disque dur externe pour le MinION)
- un séquençage immédiat, ou en temps-réel
- une longueur de lecture inégalée (> 60kb)
- une fabrication de banque facilitée (pas d'amplification clonale)
- et un très faible coût.

Nul doute que ces nouvelles technologies remplaceront rapidement les séquenceurs de 2^{ème} génération, et permettront de nouvelles analyses encore peu réalisées actuellement.

5.5. L'ANALYSE BIO-INFORMATIQUE

La quantité de données générées par le séquençage haut-débit a rendu l'analyse des résultats manuelle impossible. Il est nécessaire de développer des outils bio-informatiques dédiés et adaptés.

Le traitement des données se fait en une succession d'étapes, du traitement du signal de base lors du séquençage, à l'annotation finale des variants retrouvés. Certaines étapes sont réalisées directement par l'automate de séquençage, quand d'autres peuvent être réalisées de manière manuelle ou semi-automatisée grâce à l'utilisation de scripts informatiques ou des logiciels commerciaux.

L'analyse bio-informatique comprend aussi le contrôle qualité : de la qualité des séquences à l'appel de variant. Les principales étapes de l'analyse bio-informatique (et de la validation biologique) sont décrites dans la figure 41.

Certaines étapes nécessitent des ressources logistiques importantes et demandent des investissements initiaux très importants. Pour travailler avec des données NGS, de nombreux besoins doivent être pris en compte afin de proposer une analyse de qualité :

- Besoins en capacité de calculs ;
 - o station informatique puissante avec plusieurs cœurs CPU et une mémoire vive suffisante.

- Besoins en logiciel spécialisé :
 - o Le traitement des données est impossible avec les logiciels de bureautique.
 - o Besoin de logiciels académiques, qui tournent le plus souvent sur Linux.
 - o Scripts (programmation en perl, python, R, java).
 - o Logiciels commerciaux (souvent sur windows).

- Besoins en capacité de stockage :
 - o Stockage rapide et temporaire sur la station de travail.
 - o Stockage à moyen terme pour les essais, les runs non validés etc.
 - o Stockage à long terme, d'une très grande capacité, en plusieurs exemplaires.

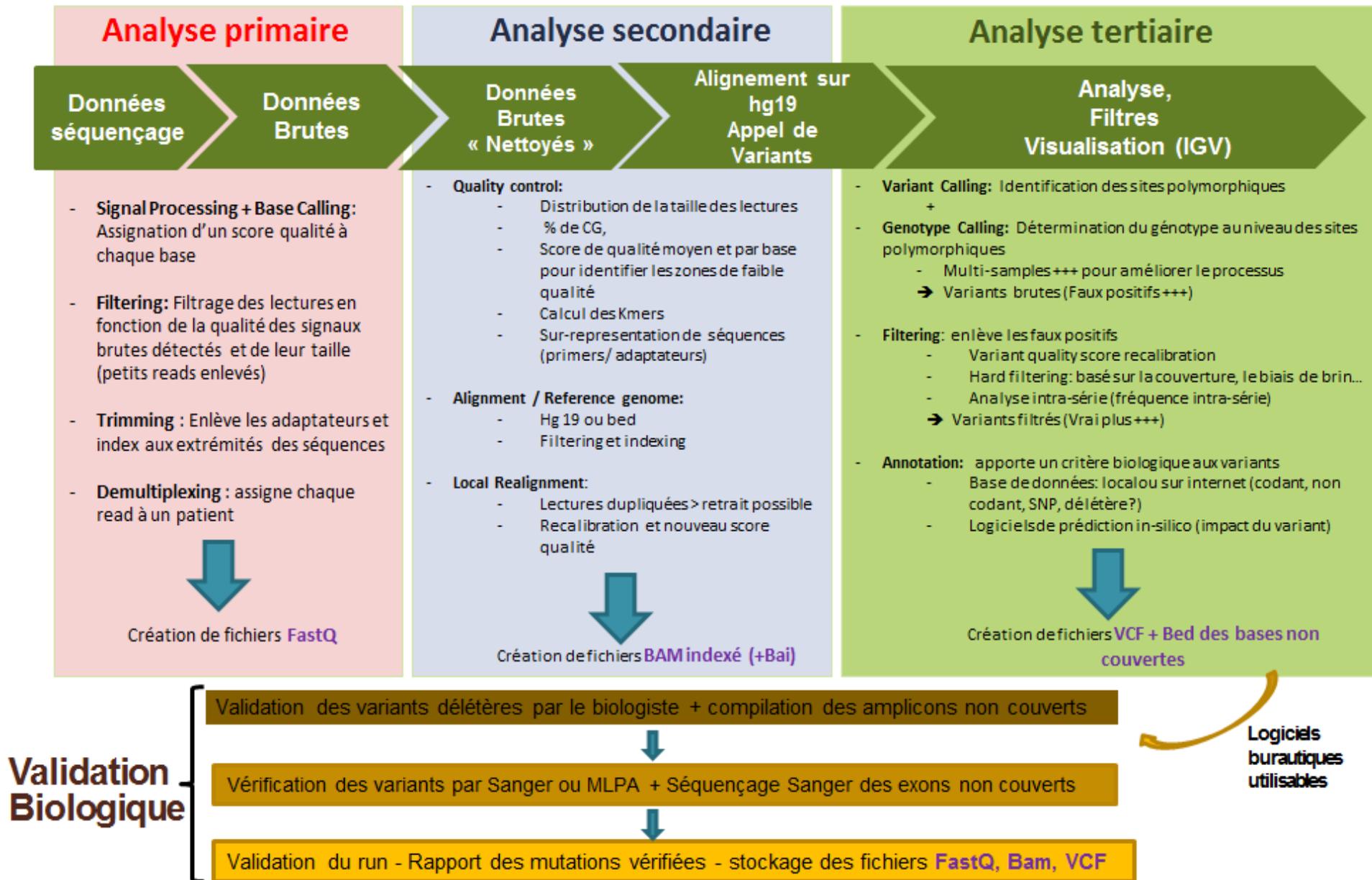


Figure 41. Principales étapes de l'analyse bio-informatique

5.5.1. ÉTAPE D'ALIGNEMENT

L'étape d'alignement est une étape importante dans le processus de traitement des données. Elle permet, à partir de l'ensemble des séquences produites, d'identifier les zones de concordance et de donner une information globale et structurée.

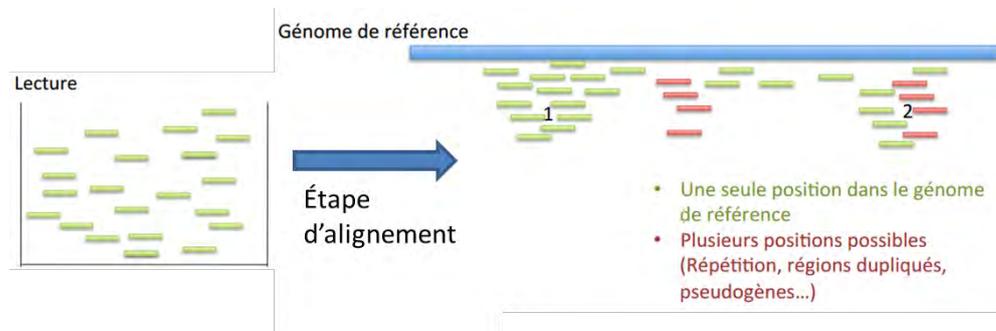


Figure 42. Schéma de l'alignement sur un génome de référence.

(En vert les *reads* alignés a un seul endroit, en rouge les *reads* pouvant etre s'alignés a plusieurs endroits,et qui ne seront pas gardés)

L'alignement de novo permet de reconstruire un génome inconnu à l'avance, ou pour lequel aucune séquence de référence n'existe. Il est utile dans le domaine des pathologies infectieuses, ou dans les sciences environnementales. De nombreux microorganismes sont découverts et peuvent être caractérisés génétiquement grâce aux nouvelles techniques de séquençage, par l'assemblage de courts *reads* et l'utilisation d'algorithmes optimisés¹⁷³.

Les *reads* peuvent être alignés sur un génome de référence. Des algorithmes nouveaux ont été créés afin de prendre en charge les caractéristiques nouvelles des séquences produites par NGS : une grande quantité de séquence (millions), de tailles diverses (35 à 400 pb), des erreurs de séquençages différentes selon les plateformes.

De très nombreux outils ont vu le jour et les améliorations sont continues (gestion des homopolymères et des séquences répétées, des indels...).

| Software | Function |
|---|---|
| MAQ, Bowtie, SSAHA2, GATK, BWA, SOAP2, MOSAIK, and SAMtools | Analysis, short-read alignment, and SNP calling |
| BreakDancer, Pindel, Dindel, PEMer, and VariationHunter, | Structural variant (INDELS and CNVs) calling |
| SNVMix, VarScan, SomaticSniper, TIGRA | Cancer-specific genome assembly and variant calling |
| SAMtools, EagleView, MaqView, Tablet, MapView, and IGV | Alignment viewers |
| PairedScope | Visualization of paired-end data |
| BLAST, BLAT, phyloP, and PHAST | Analysis of evolutionary conservation |
| SIFT, PolyPhen-2, SNPs3D, PMUT, TopoSNP, PANTHER, Align GVGD, MAPP, PhDSNP, nsSNPA, and Parepro | Prediction of the effect of amino acid substitution |

Figure 43. Exemples de logiciels académiques pour l'alignement, la visualisation, l'appel de variant et leur interprétation¹⁷⁴.

Plateforme Ion Torrent :

Life Technologies applique à travers leur solution Torrent Server un programme d'alignement optimisé par leurs soins qui se nomme TMAP (Torrent Mapping Alignment Program). Il intègre plusieurs algorithmes largement utilisés pour l'alignement :

- BWA-short¹⁷⁵
- BWA-long¹⁷⁶
- SSAHA¹⁷⁷
- Super-maximal Exact Matching¹⁷⁸

BWA est un des algorithmes les plus utilisés dans l'alignement de séquences multiples (multiple sequence alignment (MSA)) sur un large génome comme celui du génome humain. Il utilise la transformée de Burrows-Wheeler (méthode de réorganisation des données) et autorise les mismatches et les gaps. Il est rapide et demande peu de ressources. Il supporte les séquençages simples (*simple end*) et pairés (*paired end*).

SSAHA est un algorithme dit de hachage (hashing algorithm) qui permet de « couper » les séquences en fragment avant de procéder à l'alignement. Il est plus gourmand en ressources et plus lent, mais d'une bonne sensibilité. Super-maximal Exact Matching, couplé à BWA (BWA-MEM), permet d'améliorer la vitesse et la précision de l'alignement de grandes séquences (de 70 pb à plusieurs Mb).

TMAP agrège les données des différents algorithmes afin de créer une liste de Candidate Mapping Locations (CMLs), qui sont ensuite alignés en utilisant l'algorithme de Smith Waterman (algorithme dynamique recherchant les alignements globaux et locaux)¹⁷⁹. TMAP travail en deux étapes afin de réduire le temps de calcul : les *reads* qui n'ont pas pu être alignés durant la première phase doivent passer un second round d'alignement paramétré différemment. Par défaut, TMAP travail avec le module map4, qui peut être modifié (map1 : BWA-short ; map2 : BWA-long ; map3 : SSAHA) lorsque un ré-alignement est demandé.

5.5.2. ÉTAPE DE L'APPEL DE VARIANT (VARIANT CALLING)

Cette étape permet de détecter les variants après l'alignement des séquences sur un génome de référence et de donner une information « génétique » aux lectures. Il s'accompagne d'une estimation de la fréquence du variant par rapport à un nucléotide de référence. D'autres informations peuvent être utiles afin de mesurer la confiance à donner à un appel de variant :

- Base qualities : une qualité faible indique qu'une erreur de séquençage a pu se produire
- Read positions : des biais peuvent indiquer un problème de mapping
- Genomic strand : un biais de brin peut indiquer un problème de mapping
- Genomic position : permet de repérer les duplicats de PCR
- Mapping info : score qualité donné à l'alignement

Si par définition, l'appel de variant doit être assez peu stringent afin de ne manquer aucun variant, cela peut entraîner la présence de nombreux faux positifs dans le fichier VCF final. Pour empêcher un trop gros nombre de faux positifs tout en gardant la meilleure sensibilité possible, des filtres doivent être appliqués afin de trier les variants faux positifs des vrais mutations.

Ces filtres sont dépendants de chaque analyse (échantillons passés, plateforme utilisé, utilité finale), et doivent être paramétrés très finement et individuellement. Les performances du variant caller sont dépendantes de la profondeur de couverture pour une base donnée : plus elle est profonde, meilleure en seront la sensibilité et la spécificité. Il est important afin d'améliorer

les performances, d'analyser toutes les patients dans leur ensemble (comme préciser dans les bonnes pratiques de GATK), et non d'individualiser l'analyse,

| | | sequencing error filter level | | | | | | |
|-----------|------|-------------------------------|-----|-----|-----|-----|-----|-----|
| Power (Q) | | 5% | 10% | 15% | 20% | 25% | 30% | 35% |
| 90.00% | (10) | 4 | 4 | 4 | 7 | 7 | 12 | 24 |
| 95.00% | (13) | 5 | 5 | 8 | 8 | 11 | 18 | 30 |
| 99.00% | (20) | 7 | 7 | 11 | 17 | 19 | 35 | 54 |
| 99.50% | (23) | 8 | 12 | 12 | 18 | 26 | 42 | 71 |
| 99.90% | (30) | 10 | 14 | 18 | 27 | 38 | 61 | 110 |
| 99.95% | (33) | 11 | 15 | 19 | 28 | 42 | 68 | 117 |
| 99.99% | (40) | 14 | 18 | 28 | 34 | 54 | 83 | 148 |
| 99.995% | (43) | 15 | 19 | 30 | 42 | 61 | 92 | 165 |
| 100.00% | (50) | 17 | 25 | 36 | 51 | 70 | 109 | 194 |

Figure 44. Profondeur minimale théorique pour détecter avec une puissance suffisante (Q), les variants hétérozygotes ayant une fréquence allélique supérieure au minimum fixé qui est le *sequencing error filter level*. (De Leener K. et al, 2011¹⁸⁰).

La figure 44 nous indique qu'à Q30 (Q est similaire au score PHRED), le standard de confiance en biologie-moléculaire, un minimum de profondeur théorique de 194 est requis.

Comme pour les outils d'alignement, de nombreux logiciels (académiques ou commerciaux) ont été développés ces dernières années.

Pour notre étude, un **minimum** de **100 X** de profondeur nous parait un minimum pour offrir une confiance suffisante dans l'appel de variant.

Plateforme Ion Torrent :

L'algorithme de détection des SNV inclut en amont plusieurs filtres pour une position donnée sur :

- **Qualité des lectures** : les *reads* avec un score de d'alignement (*mapping QV*) trop faible sont retirés.
- **Nombre de reads** : un nombre suffisant de reads doivent couvrir la base qui porte un variant (en relation avec la fréquence allélique minimum du variant définie par l'utilisateur).
- **Biais de brins** : la position qui porte un potentiel variant ne doit pas être couverte par des *reads* d'un même sens (présentant un biais de brin par rapport aux autres *reads* couvrant la même base et non porteur du variant).
- **Qualité de la base** : les *reads* qui contiennent une base avec un score qualité trop faible sont éliminés.

Si la position passe les filtres, le SNV est appelé et un génotype ainsi qu'une p-value lui est assigné.

Deux méthodes basées sur des modèles statistiques sont utilisées conjointement pour le calcul de prédiction d'une mutation ponctuelle : une méthode Bayésienne et une fréquentiste.

La méthode Bayésienne est utilisée quand la profondeur de couverture est faible (10-60X). L'approche fréquentiste est utilisée pour les positions avec une grande profondeur, et peut être utilisée pour détecter des variants de faible fréquence (utilisé par défaut si la profondeur est supérieure à 60X).

La détection des insertions ou délétions se fait en plusieurs étapes en utilisant une version améliorée de GATK unifiedGenotyper. Une première étape liste toutes les possibilités de variants ayant une fréquence allélique supérieure à 10%. La deuxième étape utilise une modélisation Bayésienne en prenant en compte le « flow intensity » et le « flow-space alignment » des *reads* qui contiennent le potentiel variant. Les variants candidats sont triés en 3 catégories : Variant dans une région homopolymérique ('TTTT' -> 'TTT'), dans une région répétée multi-nucléotidique ('ATATAT' -> 'ATAT') et les autres. Le traitement des trois classes est différent, spécialement les régions homopolymériques qui ont une forte proportion de faux positifs, et pour lesquels des corrections doivent être effectuées (estimation du biais de brin, estimation du *Flow Peak*...). La troisième étape permet d'appeler les longues insertions ou délétions. La quatrième étape compile les données des étapes précédentes et applique les filtres de score qualité, fréquence, biais de brin, et enfin produit la liste finale des variants.

La figure 45 schématise le *workflow* du Variant Caller et montre que des fichiers *input* peuvent être nécessaires. C'est le cas pour le fichier BED (*Browser Extensible Data*) qui permet de donner des coordonnées génomiques dans lesquels les variants seront recherchés. De plus, on peut forcer le variant caller à écrire dans son fichier de sortie VCF, les positions que l'on veut étudier à travers un fichier « hotspot » qui est lui-même un fichier BED ou VCF (voir figure 41 pour le format des fichiers).

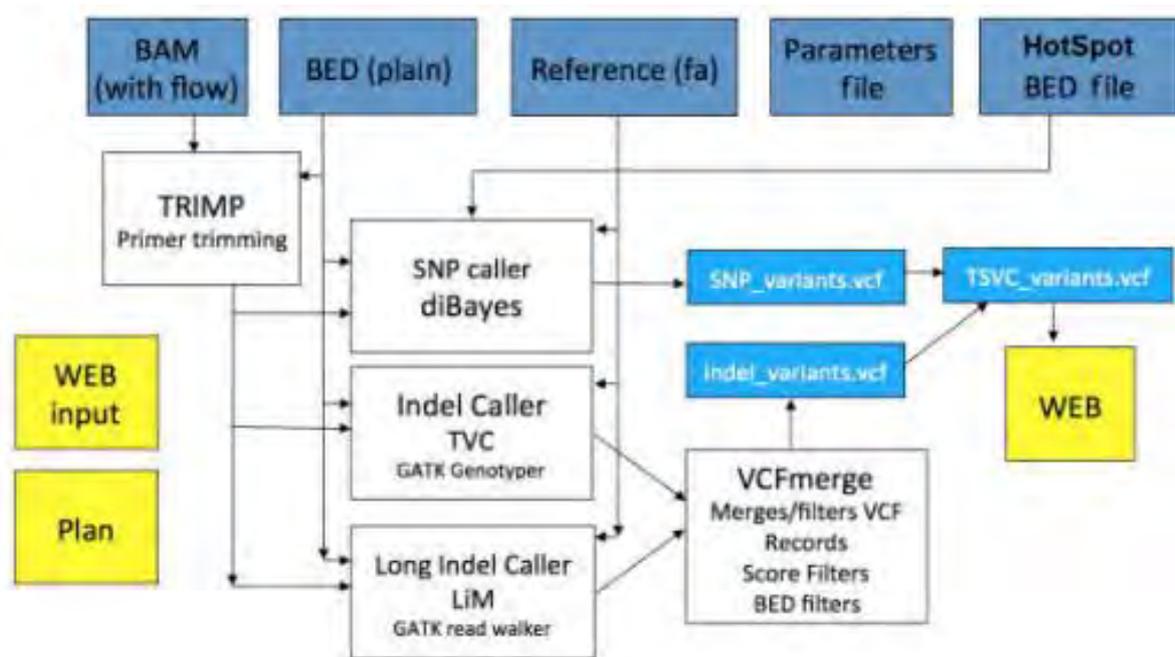


Figure 45. Workflow du Variant Caller (plateforme Ion torrent).

5.5.3. LES ERREURS ET BIAIS DU NGS

Quelques erreurs et biais sont inhérents au séquençage de 2^{ème} génération. Ils peuvent arriver à différentes étapes et doivent être contrôlés puis éliminés avant de passer à l'étape suivante. Voici quelques erreurs et biais classiques :

- *Artéfacts de séquençage/PCR : les duplications*

Les duplicats de PCR surviennent lors de la première phase de préparation des échantillons. Ils correspondent à des copies de la même séquence (même alignement, même CIGAR etc). Il est nécessaire de les éliminer afin d'éviter les problèmes de biais qui peuvent arriver à l'appel de variant. Les duplicats entraînent une sous-estimation d'un potentiel variant (*downstream analysis*), ce qui peut l'empêcher de passer le filtre de fréquence allélique. Ils peuvent aussi amplifier les erreurs de PCR et donner de faux positifs.

Picard, un outil avec ligne de commande codé en java, permet d'éliminer ces duplicats (*picard mark and remove duplicate*).

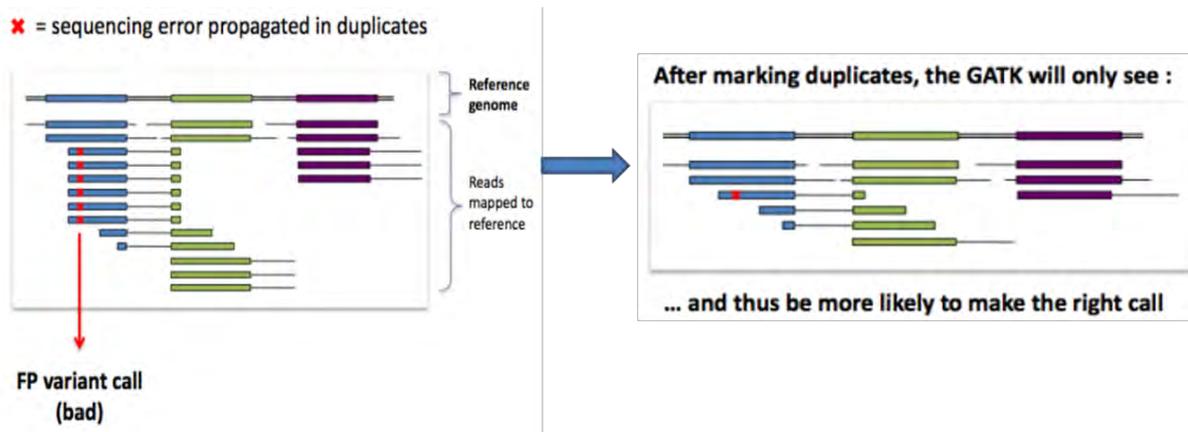


Figure 46. Erreur de PCR propagée dans les duplicats et résultat après leur élimination (source : GATK).

-
- *Réalignement des indels*

Des problèmes d'alignement peuvent survenir dans les régions de faible complexité autour des indels. Des mésappariements dues aux indels sont détectés comme des variants et peuvent biaiser les modèles statistiques utilisés lors de la détection des variants. Une étape de réalignement autour des indels permet de réduire ces erreurs en prenant en compte :

- Les indels connus (base de données : dbSNP, 1000 génomes...)
- Les indels détectés dans les échantillons analysés (en utilisant le CIGAR)
- Les sites avec des indels supposés

Un score de consensus (égal à la somme des qualités des bases avec mésappariement) est calculé et conservé s'il est supérieur au score de l'alignement original.

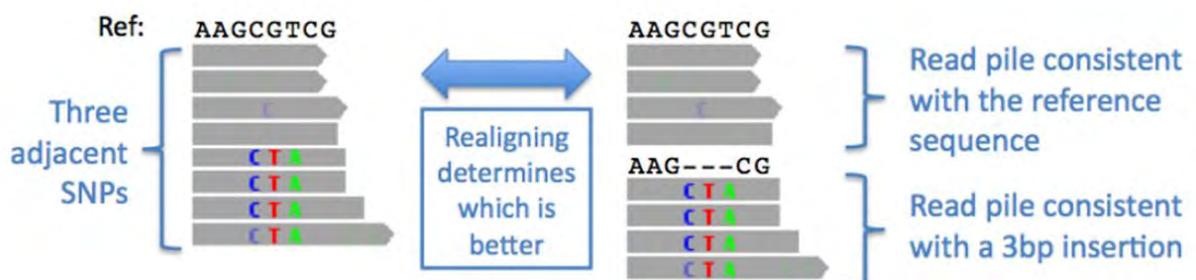


Figure 47. Exemple de réalignement d'une indel.

L'outil GATK permet de chercher les régions à problème (« *RealignerTargetCreator* »), et il est réalisé par une seconde commande : « *IndelRealigner* ».

5.5.4. NOTIONS DE BASE ET DIFFERENTS FORMATS DE L'ANALYSE BIO-INFORMATIQUE

5.5.4.1. Les formats

A la fin de l'étape primaire, secondaire et tertiaire (figure 41), un fichier avec un format spécifique est produit. Quelques notions sont à connaître afin de les manipuler au mieux.

Format FASTQ (étape primaire):

C'est un format de fichier texte permettant de stocker des séquences biologiques (ici nucléiques) en leur associant un score de qualité. Une base d'une séquence et son score sont codés en caractères ASCII (American Standard Code for Information Interchange). Il a été inventé au Wellcome Trust Sanger Institute et fusionne deux fichiers : FASTA et QUAL. Il peut être ouvert par un éditeur de texte (attention à la taille), et peut atteindre plusieurs giga-octets.

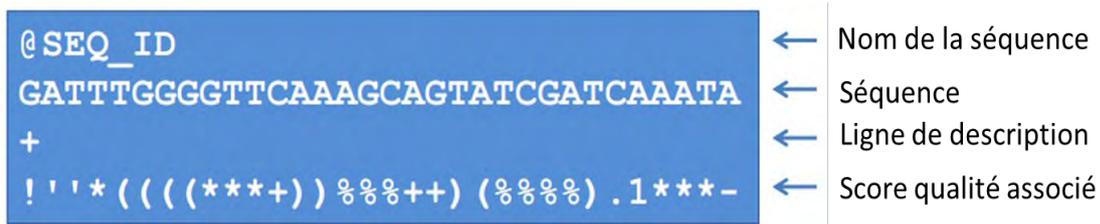


Figure 48. Description du format FASTQ.

Le score de qualité est un score Phred, développé originellement pour le programme Phred (Projet Génome Humain). Les scores phred Q ont pour propriété d'être reliés de façon logarithmique à la probabilité d'erreur d'identification d'une base P .

Le score phred est devenu un standard, et permet de comparer la qualité des séquences des différentes plateformes.

Voici la formule qui relie les deux critères, ainsi que le tableau de correspondance :

$$Q = -10 \log_{10} P \Leftrightarrow P = 10^{\frac{-Q}{10}}$$



| Score de qualité phred | Probabilité d'une identification incorrecte | Précision de l'identification d'une base |
|------------------------|---|--|
| 10 | 1 pour 10 | 90 % |
| 20 | 1 pour 100 | 99 % |
| 30 | 1 pour 1000 | 99.9 % |
| 40 | 1 pour 10000 | 99.99 % |
| 50 | 1 pour 100000 | 99.999 % |

Figure 49. Correspondances du score Phred.

Un score de Q30 ou plus est généralement demandé pour les analyses de génétique. A partir d'un certain nombre de nucléotides, la qualité des bases séquencées diminue.

Une analyse de la qualité en fonction de la position dans le *read* permet d'estimer cette baisse de qualité. Les bases situées aux extrémités des *reads* peuvent être coupées (*triming*). Ci-dessous, on peut voir un exemple d'un graphe issu de l'outil FASTQC, avec perte de qualité.

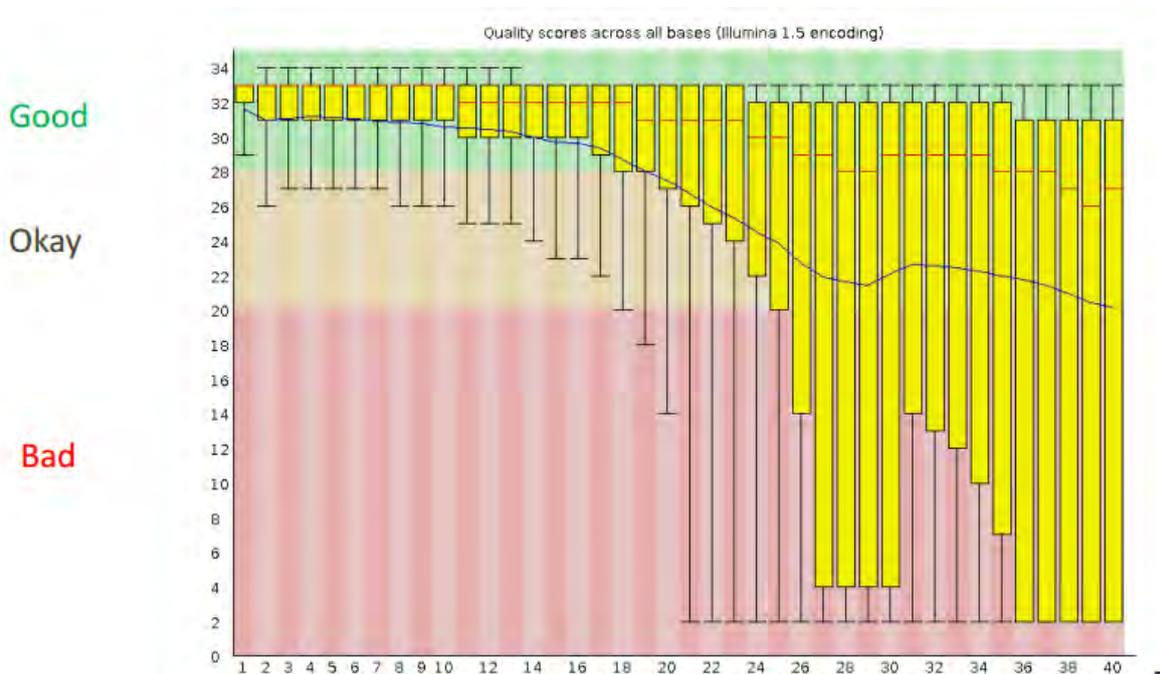


Figure 50. Score qualité (Phred) par base (outil FASTQC).

Format SAM/BAM (étape secondaire) :

Les fichiers SAM (Sequence Alignment/Map) sont des fichiers d'alignement générique, utilisés comme fichiers de sortie par le projet 1000 génomes. C'est un fichier texte tabulé (SAM), pouvant s'ouvrir avec un éditeur de texte (attention à la taille) qui est créé par tous les logiciels d'alignement récents. Ils comprennent deux sections : En tête optionnel (*headers*) et les informations d'alignement (*reads*).

L'alignement a plusieurs critères (voir figure 51), dont le score qualité en Phred ou la localisation sur la séquence de référence. De nombreux « *flags* » permettent de coder des informations importantes, comme celle contenue dans le « CIGAR ». Elle définit plusieurs opérations possibles pouvant décrire l'alignement d'un *read* (les numéros codant pour les flags sont une addition de plusieurs informations).

Le format BAM est la version compressée du fichier SAM (compatible avec GZIP). Une fois indexé (recherche d'alignements plus rapide sur le fichier), un fichier « *.bam.bai » est créé. Il porte les coordonnées chromosomiques qui forment l'index.

Le fichier BAM est au final le fichier stocké et utilisé en pratique à cause de sa compacité par rapport au fichier SAM. Son poids peut atteindre plusieurs giga-octets.

Format VCF (étape tertiaire) :

C'est un format générique qui contient les informations de sortie de l'appel de variants¹⁸¹. Le « Variant Calling format » est un fichier texte tabulé qui permet de stocker les informations de plusieurs échantillons (à l'inverse du format GFF non adapté). Il contient une partie « en-têtes » (optionnelle) et une partie « body » qui contient l'information sur le variant, la séquence de référence, la profondeur de couverture, et le score qualité du variant (Phred). Son poids n'excède habituellement pas les méga-octets.

Il faut être prudent avec la manipulation informatique des VCF, qui peuvent différer dans leur structure selon le logiciel utilisé.

Example

```

##fileformat=VCFv4.0
##fileDate=20100707
##source=VCFtools
##reference=NCBI36
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality (phred score)">
##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods for RR,RA,AA genotypes (R=ref,A=alt)">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1 SAMPLE2
1 1 . ACG A,AT . PASS . H2;AA=T GT:DP 1/2:13 0/0:29
1 2 rs1 C T,CT . PASS . H2;AA=T GT:GQ 0|1:100 2/2:70
1 5 . A G . PASS . H2;AA=T GT:GQ 1|0:77 1/1:95
1 100 T <DEL> . PASS SVTYPE=DEL;END=300 GT:GQ:DP 1/1:12:3 0/0:20
  
```

Mandatory header lines (lines starting with ##)

Optional header lines (meta-data about the annotations in the VCF body) (lines starting with #)

Reference alleles (GT=0) (A, C, G, T)

Alternate alleles (GT>0 is an index to the ALT column) (A, AT, T, CT,)

Types of variants

| Variant Type | Alignment | VCF representation |
|---------------------------|-----------|--------------------------|
| SNPs | ACGT | POS REF ALT |
| | ATGT | 2 C T |
| | | |
| Deletions | ACGT | POS REF ALT |
| | A--T | 1 ACG A |
| | | |
| Insertions | AC-GT | POS REF ALT |
| | ACTGT | 2 C CT |
| | | |
| Complex events | ACGT | POS REF ALT |
| | A-TT | 1 ACG AT |
| | | |
| Large structural variants | ACGT | POS REF ALT INFO |
| | 100 T | SVTYPE=DEL;END=300 |
| | | |

Other annotations: Deletion, SNP, Large SV, Insertion, Other event, Phased data (G and C above are on the same chromosome).

Figure 52. Format VCF, et nomenclature des différents types de variants

(source : <http://vcftools.sourceforge.net/VCF-poster.pdf>)

Format BED (utilisable à toutes les étapes) :

Le format BED (*Browser Extensible Data*) est un format de fichier texte utilisé pour stocker des régions génomiques sous forme de coordonnées (début et fin) ainsi que des annotations associées.

C'est un fichier tabulé qui a été développé au cours du Projet Génome Humain¹⁸² et largement utilisé au sein du *Genome Browser*. Le stockage des coordonnées génomiques à la place de séquences permet une optimisation du temps de calcul lors de la comparaison à un génome. Les fichiers BED ne dépassent habituellement pas le méga-octet.

La première ligne peut contenir des en-têtes facultatifs pour l'inscription de métadonnées. Les trois premières colonnes sont obligatoires, neuf autres colonnes sont facultatives. L'ordre des colonnes doit être respecté.

| Numéro de colonne | Titre | Définition |
|-------------------|-------------|--|
| 1 | chrom | Nom des <i>chromosomes</i> (ex. : chr3, chrY, chr2_random) ou des <i>scaffolds</i> (ex. : scaffold10671) |
| 2 | chromStart | Coordonnée de départ sur le chromosome ou le <i>scaffold</i> pour la séquence considérée (la première base sur le chromosome est numérotée 0) |
| 3 | chromEnd | Coordonnée de fin sur le chromosome ou le <i>scaffold</i> pour la séquence considérée |
| 4 | name | Nom de la ligne du fichier BED |
| 5 | score | Score compris entre 0 et 1000 |
| 6 | strand | Orientation du brin d'ADN (positif ("+") ou négatif ("-")) |
| 7 | thickStart | Coordonnée de départ à partir de laquelle l'annotation est affichée de façon plus épaisse sur une représentation graphique (ex. : le <i>codon start</i> d'un gène) |
| 8 | thickEnd | Coordonnées de fin à partir de laquelle l'annotation n'est plus affichée de façon plus épaisse sur une représentation graphique (ex. : le <i>codon stop</i> d'un gène) |
| 9 | itemRgb | Valeur RGB sous forme R,G,B (ex. : 255,0,0) déterminant la couleur d'affichage de l'annotation contenue dans le fichier BED |
| 10 | blockCount | Nombre de blocs (ex. : <i>exons</i>) sur la ligne du fichier BED |
| 11 | blockSizes | Liste de valeurs séparées par des virgules correspondant à la taille des blocs (le nombre de valeurs doit correspondre à celui du <i>blockCount</i>) |
| 12 | blockStarts | Liste de valeurs séparées par des virgules correspondant aux coordonnées de départ des blocs, coordonnées calculées relativement à celles présentes dans la colonne <i>chromStart</i> (le nombre de valeurs doit correspondre à celui du <i>blockCount</i>) |

```
browser position chr7:127471196-127495720
browser hide all
track name="ItemRGBDemo" description="Item RGB demonstration" visibility=2 itemRgb="On"
chr7 127471196 127472363 Pos1 0 + 127471196 127472363 255,0,0
chr7 127472363 127473530 Pos2 0 + 127472363 127473530 255,0,0
chr7 127473530 127474697 Pos3 0 + 127473530 127474697 255,0,0
chr7 127474697 127475864 Pos4 0 + 127474697 127475864 255,0,0
chr7 127475864 127477031 Neg1 0 - 127475864 127477031 0,0,255
chr7 127477031 127478198 Neg2 0 - 127477031 127478198 0,0,255
chr7 127478198 127479365 Neg3 0 - 127478198 127479365 0,0,255
chr7 127479365 127480532 Pos5 0 + 127479365 127480532 255,0,0
chr7 127480532 127481699 Neg4 0 - 127480532 127481699 0,0,255
```

Figure 53. Format BED (en rouge les colonnes obligatoire) et un exemple tiré du genome browser (à 9 colonnes)

(source : [https://fr.wikipedia.org/wiki/BED_\(format_de_fichier\)#cite_note-3](https://fr.wikipedia.org/wiki/BED_(format_de_fichier)#cite_note-3))

5.5.4.2. Les notions

Système de coordonnées :

Un des points important du format BED est l'emploi d'un système de coordonnées basé sur l'utilisation du 0. Ainsi, la longueur des régions considérées peut s'obtenir par soustraction de la colonne 3 (fin) par la colonne 2 (origine des coordonnées): $x_{fin} - x_{origine}$. Lorsque le système est basé sur l'utilisation du 1, le calcul demande une addition de plus : $x_{fin} - x_{origine} + 1$.

Cette différence permet un gain important en terme de temps de calcul lorsque le jeu de données est important.

Cette optimisation n'est pas retrouvée dans tous les formats de fichiers contenant des coordonnées génomiques. Il faut donc être vigilant quant à la manipulation et la conversion d'un fichier BED en un autre format.

Couverture et profondeur

Deux notions importantes et relatives au séquençage haut-débit sont : la couverture (*coverage*) et la profondeur de la couverture (*depth*) :

- Couverture : proportion d'une région étudiée (souvent donnée par un BED) qui est couverte avec une profondeur **X** donnée.
C'est-à-dire, le pourcentage de bases qui sont séquencées à une profondeur minimum X.
- Profondeur : nombre de fois qu'une base donnée Y est séquencée. C'est-à-dire, nombre de lecture (*reads*) qui couvrent cette base Y après l'étape d'alignement.

Sur la figure 54, on peut voir que la base A est couverte par 18 *reads* différents, la profondeur de couverture pour cette base A est donc de 18. Cette figure montre aussi les différents types d'altération moléculaire et chromosomique qui peuvent être analysés grâce au séquençage haut-débit.

Par abus de langage, la couverture est souvent confondue à l'oral en lieu et place de la profondeur de couverture.

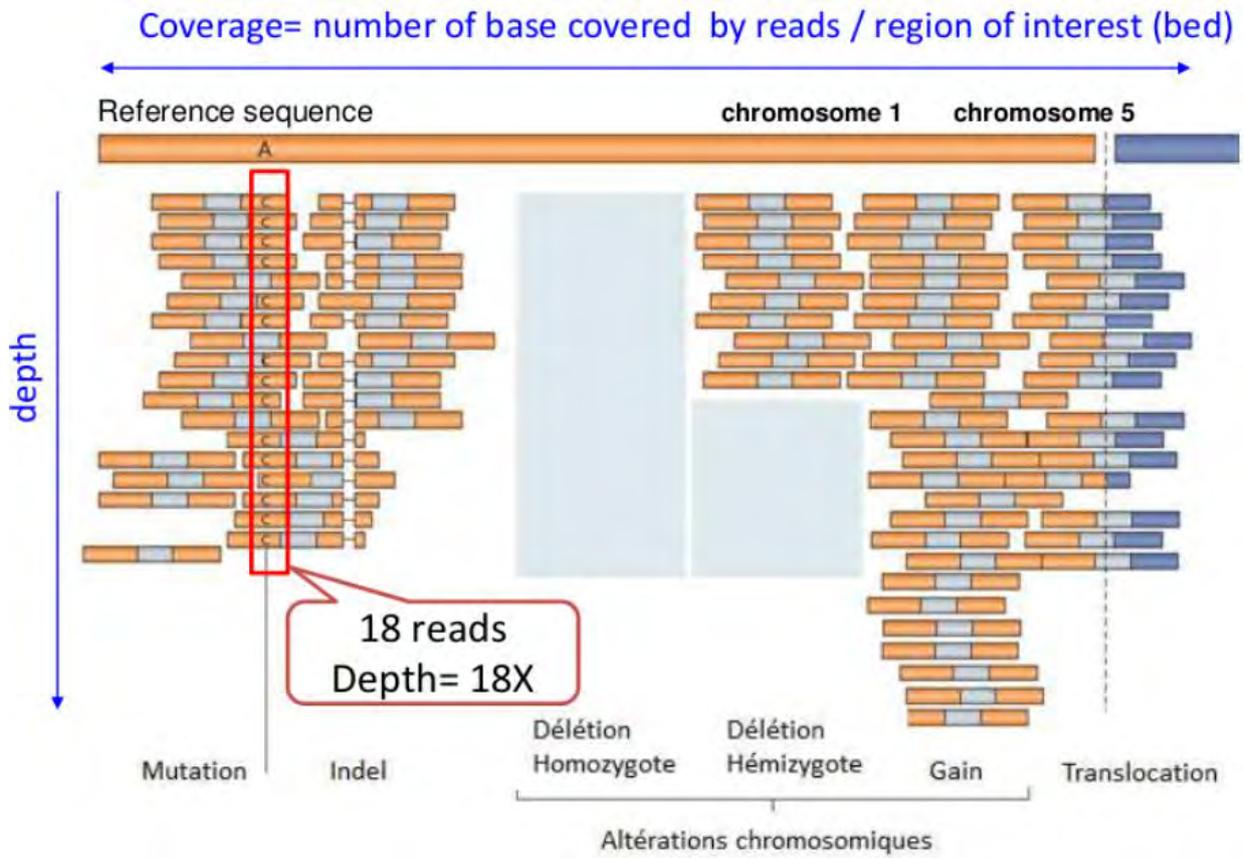


Figure 54. Notions de couverture et profondeur. Le variant C est porté par 18 lectures (18X).

PARTIE II : Expériences et Résultats

I. MATERIELS ET METHODES

L'objectif de ce travail est de mettre au point une nouvelle technique de séquençage haut-débit pour l'analyse moléculaire du gène *NF1* et *SPRED1*. L'intérêt principal réside dans l'augmentation du débit d'analyse proposé par les techniques de séquençage de 2^{ème} génération. Devant l'augmentation du nombre des analyses et la capacité insuffisante des techniques actuelles, une solution globale et performante était nécessaire. Pour cela, nous avons utilisé l'automate de paillasse PGM™ de Life Technologies, en faisant le choix de leur méthode de fabrication de librairies : la technologie Ampliseq™. Elle permet l'analyse simultanée des deux gènes analysés en séquentiel jusqu'ici.

La mise au point a nécessité de nombreux essais, et l'utilisation de deux librairies différentes. Une attention particulière a été portée à la sensibilité et spécificité de la technique, ainsi qu'à la couverture obtenue lors de ces mises aux points. Pour cela, des patients déjà génotypés avec les méthodes standards ont été analysés avec notre panel Ampliseq™. La comparaison des résultats nous a permis de caractériser les performances de notre technique.

Cette étude a aussi permis à notre équipe d'avoir une expérience solide dans la gestion des données, l'analyse bio-informatique, et la validation des variants retrouvés, dans le domaine du NGS.

6. FABRICATION DE LA LIBRAIRIE

6.1. LA TECHNOLOGIE AMPLISEQ™

La technologie Ampliseq™ est basée sur le principe de la PCR multiplex, qui permet de réaliser dans un même tube, l'amplification de plusieurs amplicons différents grâce à de multiples couples d'amorces. L'innovation de cette technique, est de pouvoir incorporer jusqu'à 24 000 couples d'amorces dans une unique réaction (Ion Ampliseq™ exome kit). L'avantage par rapport à d'autres techniques de fabrication de banque est le temps minimum requis pour la création de la banque. Par exemple, aucune étape d'hybridation n'est nécessaire comme pour les méthodes de capture. La quantité d'ADN demandée au démarrage est faible : de 5 à 10 ng.

Plusieurs types d'échantillons, ADN somatique ou ARN par exemple, sont utilisables avec cette technologie. Elle permet le séquençage de régions ciblées (*targeted sequencing*), comme employé dans cette étude, mais qui peut s'étendre jusqu'au séquençage d'exomes.

Life Technologies permet la création d'un design à façon, et propose également des designs déjà créés et validés par la communauté (*community panel*). Un panel, qui reste la propriété de son créateur, peut être designé grâce à l'outil Ion Ampliseq™ Designer à travers sa plateforme internet (www.ampliseq.com).

Son utilisation se compose de plusieurs étapes, il faut :

- Créer un fichier .bed des régions génomiques d'intérêt.
- Charger ce fichier à travers la plateforme internet, choisir les spécificités de son design (taille voulue des amplicons).
- Un algorithme calcule les meilleurs couples d'amorces possibles, et leur répartition en plusieurs tubes (dans une PCR multiplexe, certaines amorces ne peuvent pas être associées dans un même tube pour différentes raisons : hybridation, compétition...).
- Un design est alors proposé, s'il convient, il peut être commandé directement en ligne et livré en quelques semaines.

6.2. PREMIER DESIGN AMPLISEQ™

Notre laboratoire avait un cahier des charges précis pour les designs à venir. Plusieurs points devaient être respectés, parmi eux :

- Une couverture de 100% des régions exoniques des deux gènes et des 20 nucléotides introniques flanquants les exons en 5' et en 3'.
- Des amplicons de tailles homogènes.
- Un nombre de pool (tube avec amorces) limité afin de faciliter les manipulations techniques.
- Un coût raisonnable, non supérieur à celui du séquençage Sanger.

Au commencement, un fichier bed contenant les coordonnées génomiques des régions d'intérêt a été créé avec ces caractéristiques :

- Transcrits de référence : NM_000267.3 pour *NF1*, NM_152594.2 pour *SPRED1*.
- Ajouts des exons alternatifs de NF1 (9br, 10a2, 48a).
- Ajouts des régions 5' et 3' UTR pour les deux gènes.
- Ajouts de 100 nucléotides flanquants les régions exoniques (appelés : *exon padding*) afin de pouvoir détecter les variants qui modifient l'épissage.
- Renseignement du fichier bed avec les noms des exons (colonne 4), qui permet de traiter facilement les données par la suite.

Ce fichier bed a été soumis plusieurs fois au Ion Ampliseq Designer™ après modification du *padding* de 10 en 10 (+/- 90 ; 80 ; 70 etc) afin de trouver le design qui offrait la meilleure couverture possible.

La limitation de cette stratégie Ampliseq™ est, comme pour toute PCR multiplexe, que le choix des amorces est rendu difficile en présence de pseudogènes ou de régions contenant des répétitions. La spécificité des amorces étant essentielle, le logiciel apporte par défaut une « *high specificity* » aux designs proposés. Une option permet de relâcher ce paramètre et d'améliorer globalement la couverture du design, c'est la solution « *relaxed specificity* ».

Malgré l'emploi de cette solution et plusieurs tentatives de designs, la couverture proposée n'a jamais atteint 100%, et des régions non couvertes subsistent dans les régions exoniques. Le design qui avait la plus grande couverture fut choisi (en bleu sur la figure 55).

| Solution ID | Solution Type | DNA Type | Amplicon Range | Instrument & Application | Parameters | Pools (Input DNA) | Amplicons | Missed (bp) | Coverage (%) | |
|----------------|---------------------|--------------|----------------|---|--|-------------------|-----------|-------------|--------------|-------|
| IAD56660_170 ⚠ | Medium Specificity | FFPE DNA | 125 - 175 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline and somatic germline and somatic | 3.4 | 2 (20 ng) | 328 | 3 029 | 90,11 |
| IAD56660_173 ⚠ | Relaxed Specificity | FFPE DNA | 125 - 175 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline and somatic germline and somatic | 3.4 | 2 (20 ng) | 332 | 2 685 | 91,23 |
| IAD56660_182 | High Specificity | Standard DNA | 125 - 275 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline germline | 3.4 | 2 (20 ng) | 200 | 1 513 | 95,06 |
| IAD56660_185 ⚠ | Medium Specificity | Standard DNA | 125 - 275 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline germline | 3.4 | 2 (20 ng) | 201 | 1 345 | 95,61 |
| IAD56660_188 ⚠ | Relaxed Specificity | Standard DNA | 125 - 275 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline germline | 3.4 | 2 (20 ng) | 204 | 1 134 | 96,3 |
| IAD56660_189 ⚠ | Relaxed Specificity | Standard DNA | 125 - 275 bp | <ul style="list-style-type: none"> • S5/SS XL (520,530,540) • PGM • Proton | germline germline germline | 3.4 | 3 (30 ng) | 215 | 235 | 99,23 |

Figure 55. Premier design choisi (en bleu) avec l'outil *Ampliseq™ Designer*.

6.3. CARACTERISTIQUES DU 1^{ER} DESIGN

Le tableau 12 Voici regroupe les principales caractéristiques théoriques du premier design. La couverture est proche des 100%, avec seulement 235 bases non couvertes. Trois pools d'amorces sont nécessaires, ce qui entraine un coût supplémentaire à la fabrication.

| Critère | Valeur |
|---------------------------------------|--|
| Taille du panel | 36.52 kb |
| Taille de l'insert (moyenne) | 189 bp |
| Nombre d'amplicons (couple d'amorces) | 215 |
| Nombre de Pools d'amorces | 3 |
| Couverture | 99.23 % |
| Régions du bed non couvertes | 2 |
| Bases non couvertes | 218b (SPRED1_exon7) 17b (NF1_exon4b) Total: 235 bases |

Tableau 12. Caractéristiques du design 1

L'utilisation d'un logiciel de visualisation (Genome browser, IGV, Alamut™) permet de caractériser si les régions non couvertes sont exoniques ou introniques (figure 56).

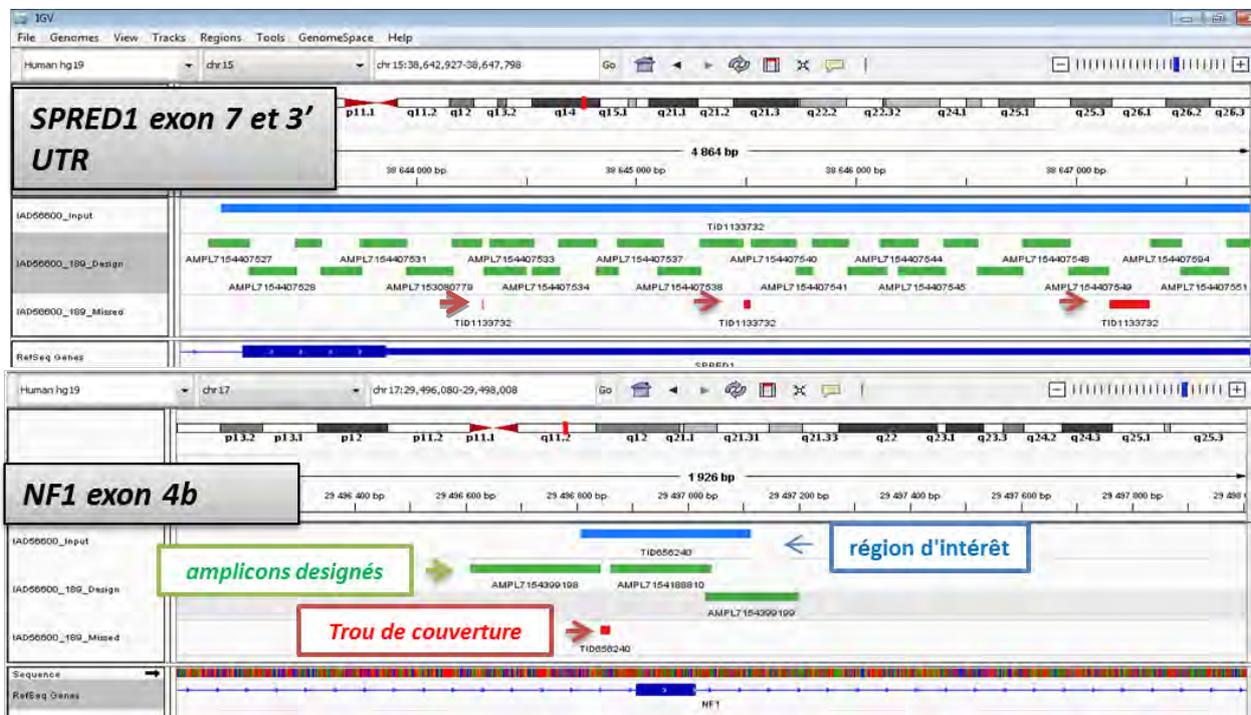


Figure 56. Visualisation des défauts de couverture (gaps) du design 1 sur IGV.

La figure 56 montre que 3 régions sont non couvertes (en rouge) dans la région 3'UTR de *SPRED1*, et 1 région dans l'intron 4 de *NF1*. Cependant, aucune base exonique ne se trouve être non couverte.

Les nucléotides qui se situent entre les amorces d'un amplicon forment les « inserts ». Ils sont informatifs, et une étude de leur homogénéité de taille a été faite (figure 57).

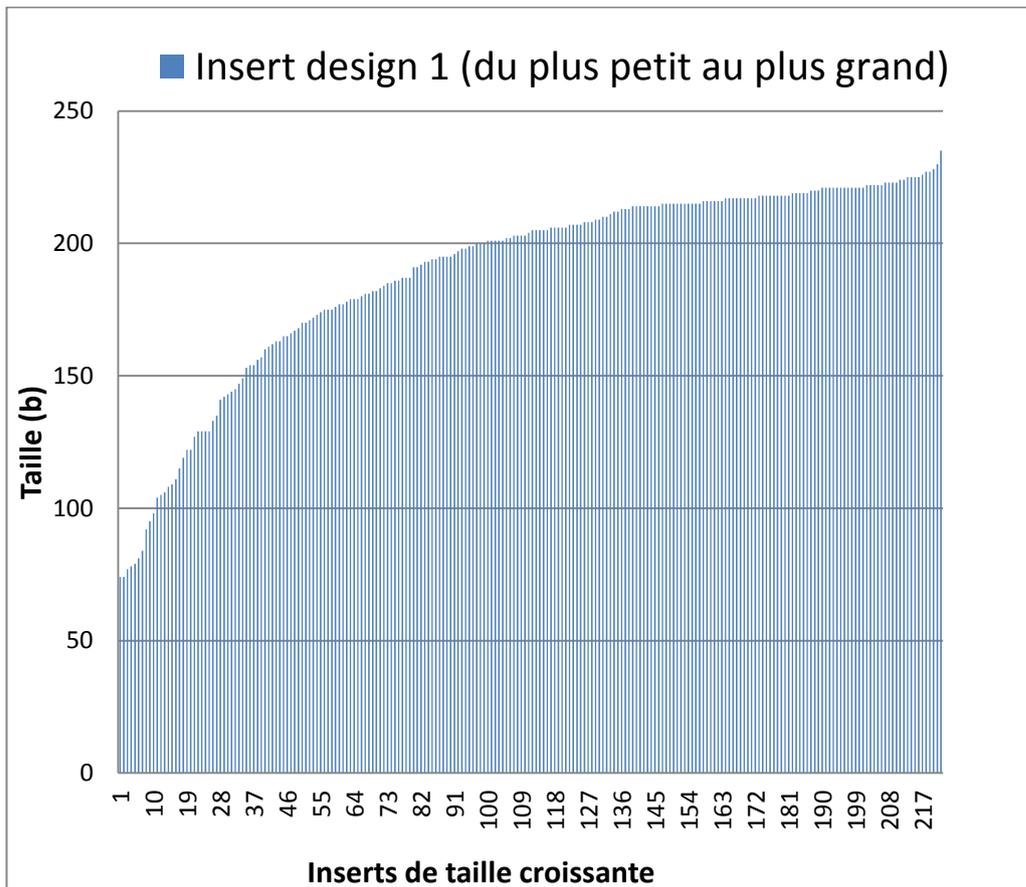


Figure 57. Répartition de la taille des inserts du design 1.

Les inserts ont une taille allant de 74 à 235 pb (les amorces sont environ de 20 nt). Environ 85% des inserts ont une longueur supérieure à 150 pb. La disparité de taille semble assez importante, pourtant, elle ne semble pas porter à conséquence en pratique.

Au final, le cahier des charges est respecté car toutes les bases exoniques sont couvertes par le design.

Le 1^{er} design a été commandé au printemps 2014 et fut livré rapidement.

6.4. DEUXIEME DESIGN AMPLISEQ™

Les différents essais faits avec le premier design ont donnés des résultats encourageants mais devant être améliorés pour être utilisable en diagnostic clinique. Un deuxième design a été entrepris avec les mêmes demandes de couverture. Cependant, notre cahier des charges stipulait cette fois la nécessité d'un recouvrement moyen de 50 bases (minimum de 20) pour les amplicons chevauchants. Cette demande était formulée après avoir rencontré des défauts de détection de variants localisés dans des régions du design où le chevauchement des amplicons était insuffisant. De plus, quelques régions étaient mal amplifiées, entraînant un manque de profondeur de couvertures, et donc de sensibilité de détection.

Pour cette nouvelle demande, nous avons utilisé le service « White Glove » de Life Technologies, spécialisé dans les panels à façon. Les bioinformaticiens retravaillent l'optimisation des amorces manuellement, après l'emploi de l'algorithme de l'Ampliseq™ Designer, de façon à combler les défauts de couverture (amplicons avec une amplification peu efficace) et le chevauchement des amplicons. Afin de faciliter le travail sur les chevauchements et de profiter d'un nouveau kit de séquençage pour longs fragments, des amplicons de plus grande taille ont été désignés (design 375 pb).

Un travail collaboratif de plusieurs semaines fut nécessaire entre les équipes de Life Technologies et notre laboratoire afin de garantir un résultat optimal.

6.5. CARACTERISTIQUES DU 2EME DESIGN

La taille et le nombre d'amplicons sont légèrement supérieurs pour le deuxième design. La couverture est totale (100%), le nombre de pools est toujours de 3.

| Caractéristiques | Valeur |
|---------------------------------------|----------|
| Taille du panel | 43.04 kb |
| Taille de l'insert (moyenne) | 219 bp |
| Nombre d'amplicons (couple d'amorces) | 230 |
| Nombre de Pools | 3 |
| Couverture | 100 % |
| Régions du bed non couvertes | / |
| Bases non couvertes | / |

Tableau 13. Caractéristiques du design 2.

La taille moyenne des inserts est supérieure avec le deuxième design (219 vs 189 bases). La taille maximale est cette fois de 297 bases.

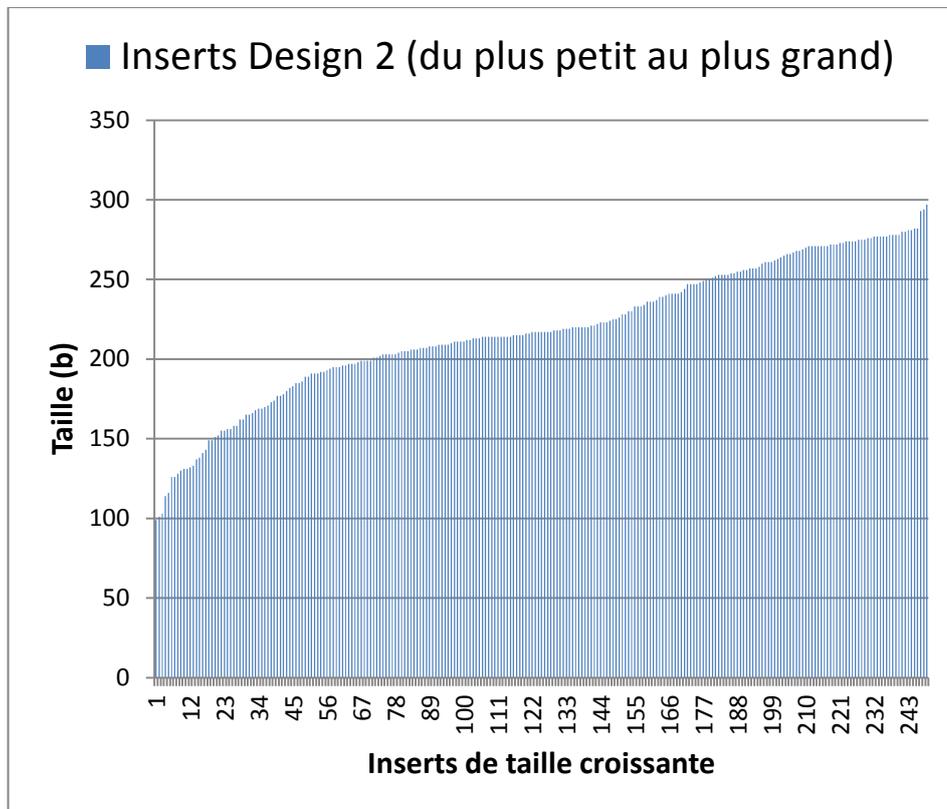


Figure 58. Répartition de la taille des inserts du design 2.

Grâce à une attention particulière, une grande partie des chevauchements (tableau 14) a une taille supérieure à 50b.

| Critère (chevauchement) | Valeur |
|-------------------------|-----------|
| Moyenne | 59,24 |
| 1er quartile | 59 |
| 2e quartile (médiane) | 59 |
| 3e quartile | 62 |
| 4e quartile (Max) | 192 |

Tableau 14. Caractéristiques des chevauchements du 2ème design

Malheureusement, pour 13 d'entre eux, les chevauchements restent inférieurs à 20 bases. Malgré tous les efforts de l'équipe *White Glove*, deux zones chevauchantes exoniques sont encore très courtes, dans les exons 29 et 34.

| Région d'intérêt | Entre les nucléotides | Taille du chevauchement | Zones d'intérêt (CDS +/- 20) |
|--------------------|--------------------------|-------------------------|------------------------------|
| NF1_Exon_04a | c.385 et c.394 | 10 bases | X |
| NF1_Exon_04b | c.480-36 et c.480-23 | 14 bases | |
| NF1_Exon_10a2 | c.1392+722 et c.1392+739 | 18 bases | |
| NF1_Exon_12b | c.2001+65 et c.2001+76 | 12 bases | |
| NF1_Exon_29 | c.5489 et c.5491 | 3 bases | X |
| NF1_Exon_31 | c.5832 et c.5841 | 10 bases | |
| NF1_Exon_34 | c.6407 et c.6407 | 1 base | X |
| NF1_Exon_41 | c.7259-86 et c.7259-86 | 1 base | |
| NF1_Exon_47 | c.8097+53 et c.8097+54 | 2 bases | |
| NF1_Exon_48 | c.8114 et c.8129 | 16 bases | X |
| NF1_Exon_49 | c.*702 et c.*707 | 6 bases | |
| SPRED1_Exon_06 | c.625 et c.634 | 10 bases | X |
| SPRED1_Exon_07_UTR | c.*5043 et c.*5044 | 2 bases | |

Tableau 15. Chevauchement inférieurs à 20 bases.
En orange les plus critiques (<10 bases dans un exon).

Un quatrième pool d'amorces Ampliseq™ aurait pu améliorer ces chevauchements critiques. Cependant, nous ne voulions pas changer notre mise au point de préparation de librairie, déjà optimisée pour un design à 3 pools.

Une autre possibilité aurait été de changer de méthode et de choisir une autre façon de fabrication de librairie.

Malgré tout, nous avons opté pour ce design qui semblait améliorer de manière significative les points négatifs du premier.

6.6. PREPARATION DE LA LIBRAIRIE

6.6.1. PROTOCOL GENERAL

La préparation des librairies par la méthode Ampliseq™ se fait en plusieurs étapes de PCR, de ligation et de purification, avant une quantification et un pool final dans un seul tube (figure 59). Hors pool d'amorces, les principaux réactifs nécessaires, sont contenus dans un kit commercial vendu par Life Technologies : le « *Ion Ampliseq™ library kit* ».

L'étape 4 (sur 5 au total) est composée d'une amplification et d'une purification double optionnelle. Elles ne sont plus nécessaires après le changement de méthode de quantification. Après quelques essais, le dosage fluorimétrique par QuBit™ a été remplacé par une méthode qPCR plus précise.

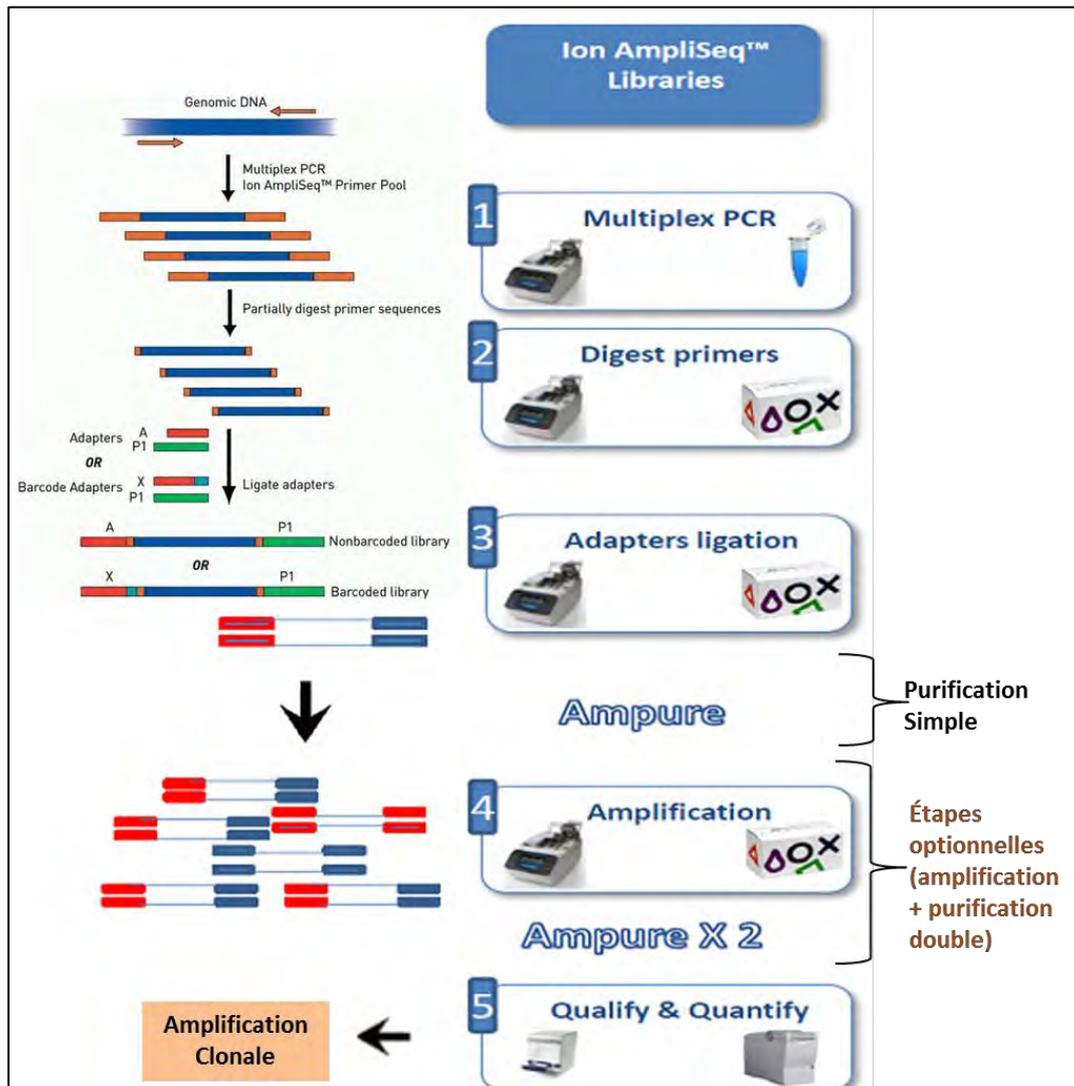


Figure 59. Protocole de préparation de librairie par la technologie Ampliseq™.

6.6.2. EXTRACTION ET DOSAGE DE L'ADN

Il est admis que la qualité de l'extraction d'ADN influe sur les résultats obtenus par séquençage haut-débit. Cependant, une étude NGS réalisée au laboratoire d'hématologie de l'hôpital Edouard Herriot par A. Janin testa 4 méthodes d'extraction différentes sans conclure à une supériorité de l'une d'entre elles¹⁸³.

6.6.2.1. Méthode d'extraction du laboratoire

Au sein de notre laboratoire, les ADN sont extraits à partir de sang prélevé sur tube EDTA par le kit NucleoSpin® 96 Blood Core (MACHEREY-NAGEL GmbH & Co. KG). La méthode a été mise au point selon le protocole du fournisseur sur un automate de pipetage Hamilton® qui permet l'automatisation et une standardisation de l'extraction d'ADN.

6.6.2.2. Méthode de dosage d'ADN

Dosage fluorimétrique

Le protocole du *Ion Ampliseq™ library kit* stipule que le dosage de l'ADN d'une solution mère, afin d'établir une dilution à 5 ng/μl doit se faire par méthode fluorimétrique avec un petit automate de paillasse QuBit™ (Life technologies).

Le kit HS (*High sensitivity*) qui permet de doser des concentrations d'ADN de 0.1 à 100 ng/μL est utilisé. L'utilisation du Qubit™ permet un dosage sensible et spécifique par l'utilisation d'un fluorochrome n'émettant qu'après liaison à l'ADN double brin. La prise d'essai est minime (2 μl minimum) et adaptée aux faibles volumes utilisés dans les différentes étapes de création de librairie.

Dosage spectrophotométrique

Le Qubit™ ne permet pas la qualification de la qualité de l'extraction. Le contrôle de pureté de l'ADN extrait, par mesure des densités optiques (DO), est important.

Il a été réalisé par méthode spectrophotométrique grâce au NanoDrop™ 1000 (TermoFisher Scientific). Les rapports DO_{260}/DO_{280} et DO_{260}/DO_{230} sont mesurés et seuls les ADN avec des valeurs supérieures à 1.7 pour les deux rapports ont été considérés comme de qualité suffisante.

En pratique : après l'extraction, un premier dosage par le NanoDrop™ permet de quantifier la concentration d'ADN, qui est ramenée par une première dilution à 20 ng/μL. Il s'en suit un dosage précis par Qubit™ qui permet alors d'obtenir une solution d'ADN concentrée à 5 ng/μL. C'est cette concentration finale qui sera alors utilisée pour initier la fabrication de la librairie.

6.7.1. AMPLIFICATION DES REGIONS CIBLES

L'*Ion Ampliseq™ Library kit*, utilisé pour l'amplification des régions cibles, existe en trois conditionnements: 8, 96 ou 384 réactions. Il est composé de tous les réactifs nécessaires à la création d'une banque :

- Le « **5x Ion AmpliseqHiFi Master Mix** »; il permet la réalisation de la PCR multiplexe de la première étape.
- L'enzyme « **Fupa** », qui digère partiellement les amorces utilisées pour la PCR multiplexe.
- La « **Switch solution** », utilisée lors de l'étape de ligation.
- L'enzyme « **ADN ligase** », qui permet la ligation des adaptateurs et des barres-codes.
- Le « **Platinum PCR SuperMix HiFi** », mélange nécessaire à l'étape d'amplification optionnelle des amplicons barre-codés.
- Le « **Library Amplification Primer Mix** », utilisé pour l'étape d'amplification optionnelle, ce mix comprend les amorces complémentaires des séquences A et P1 composant les adaptateurs.

Pour les deux designs créés avec l'*Ion Ampliseq™ Designer*, 3 pools d'amorces PCR sont nécessaires. Ces 3 réactions PCR demandent une grande quantité du « 5x Ion AmpliseqHiFi Master Mix », qui devient le réactif limitant du kit. Afin de diminuer le coût de la fabrication de banques (60% du coût total par patient¹⁸⁶), et devant les bons résultats retrouvés par l'équipe d'hématologie, nous avons testé une modification du protocole fourni par Life Technologies. Elle consiste à mettre moitié moins de *Hifi Master Mix* : 2 µL vs 4 µL normalement demandés. Les résultats de couverture pour certains exons étant décevants (pour 2 essais), nous avons conservé le protocole initial par la suite. Les autres conditions indiquées : mix et conditions de la PCR, ont été utilisés comme le protocole du fabricant le stipulait (tableau 16).

| | Vol pour 1 patient (µl) (Protocole Standard) | Vol pour 1 patient (µl) (Protocole modifié) |
|---------------------------------|---|--|
| 5X Ion AmpliSeq HiFi Master Mix | 4 | 2 |
| 2X Ion AmpliSeq Primer Pool | 10 | 10 |
| Eau (nucléase-free) | 4 | 6 |
| ADN (5ng/µl) | 2 | 2 |
| TOTAL | 20 | 20 |

Tableau 16. Mix PCR de l'étape d'amplification des cibles (protocole standard et modifié).

Pour toutes les étapes nécessitant un thermocycleur, nous avons utilisé le thermocycleur Veriti® (Life Technologies).

| Etapes | Températures | Temps |
|---|--------------------------|-------|
| Activation de l'enzyme | 99°C | 2min |
| 19 cycles (nombre de primer entre 49-96) | Dénaturation | 99°C |
| | Hybridation - Elongation | 60°C |
| Hold | 10°C | Pause |

Tableau 17. Programme d'amplification PCR de l'étape d'amplification des cibles.

Un dosage au Qubit™ est ensuite effectué pour contrôler l'efficacité des réactions PCR. Une prise d'essai de 2 à 5 µL suffit pour doser une concentration attendue entre 7 et 15 ng /µL.

Après cette première PCR, les amplicons peuvent se conserver plusieurs semaines (congélation à -20°). Ces pauses dans la fabrication des bibliothèques sont au nombre de 3.

6.7.2. DIGESTION DES PRIMERS

L'ajout de 2 µL de l'enzyme de restriction « *Fupa* » permet la digestion partielle et la phosphorylation des amorces utilisées. Cette étape de coupure en bout franc des amorces nécessite l'activation préalable de la *Fupa* avec un passage sur le thermocycleur (tableau 18).

| Température | Temps |
|-------------|--------|
| 50°C | 10 min |
| 55°C | 10min |
| 60°C | 20min |
| 10°C | Pause |

Tableau 18. Programme de l'étape de digestion des amorces.

D'après le protocole de Life Technologies, cette étape comme les suivantes doivent être réalisées en parallèle pour chaque réaction PCR (3 pour notre panel).

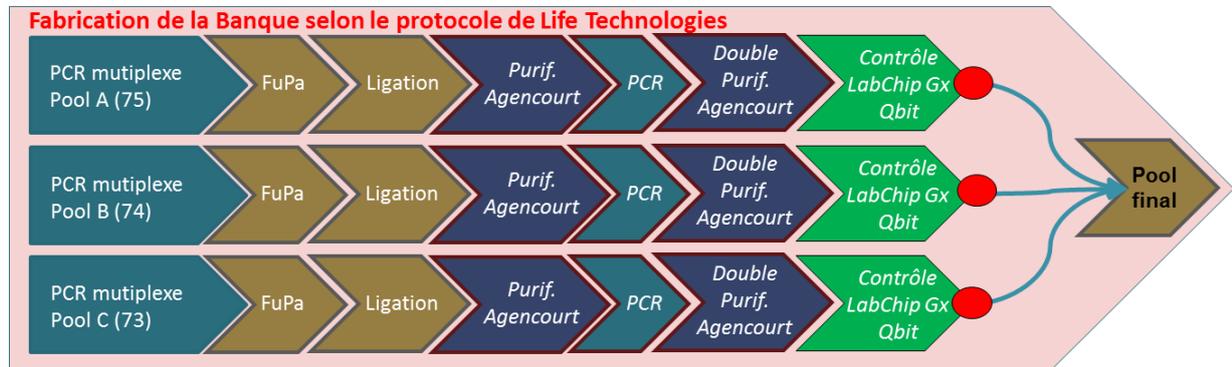


Figure 60. Protocole standard de fabrication de banque (pool final).
En rouge les étapes de pooling ou de dilution.

Aux fins de simplification et après plusieurs tests, les 3 réactions PCR pour chacun des patients sont regroupées dans un seul tube en prenant 7 μ L de chacune.

L'ajout de la *Fupa* se fait donc sur 21 μ L et non sur 20 μ L comme décrit dans le protocole initial (figure 61).

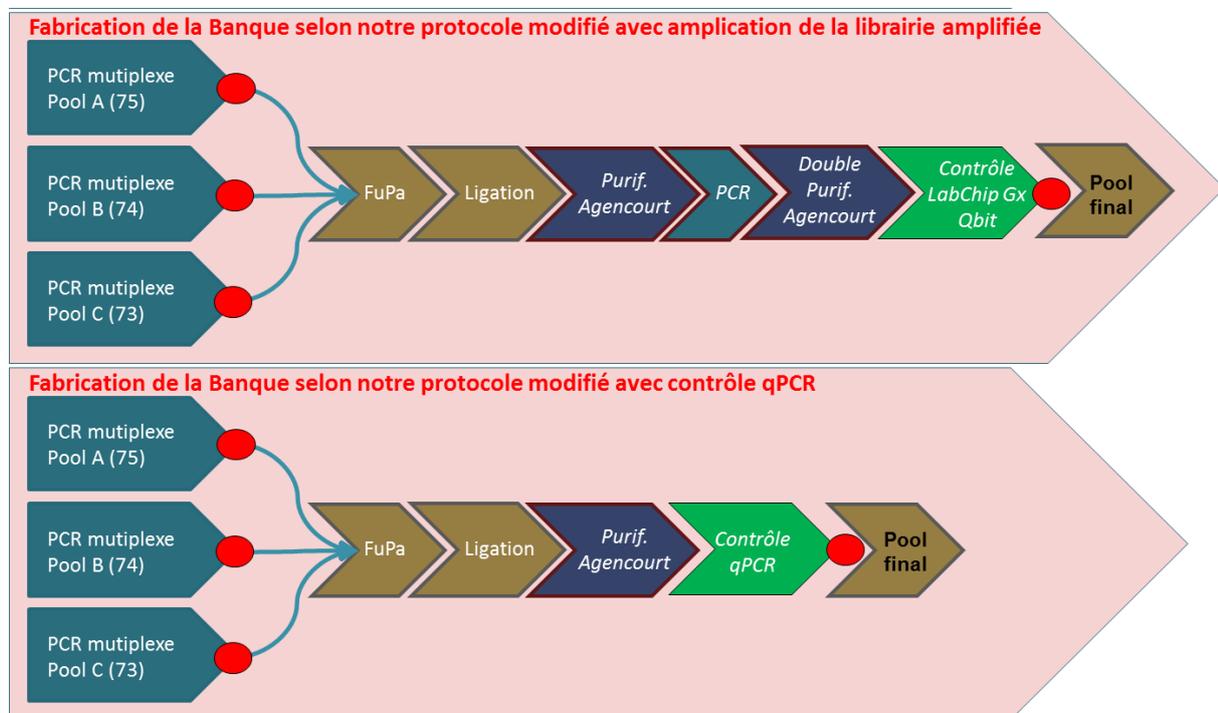


Figure 61. Protocole modifié de fabrication de banque (pool après amplification des cibles) selon la méthode de quantification.

6.7.3. LIGATION DES CODES-BARRES ET DES ADAPTATEURS

Cette étape permet de lier les codes-barres qui serviront à différencier chaque patient dans le pool final, grâce à une séquence d'une dizaine de nucléotides unique (1 code-barre = 1 patient).

Les kits spécifiques « Ion Xpress™ Barcode Adapters » contiennent les codes-barres ainsi que l'adaptateur P1. Il faut d'abord diluer au 1/4 chaque barre-code avec l'adaptateur P1 dans de l'eau *nuclease-free*. On rajoute ensuite dans chaque tube post-digestion 2 µL de cette dilution, 4 µL de *Switch Solution* et 2 µL de *DNA Ligase*. La ligation nécessite aussi l'activation préalable de la ligase avec un passage sur le thermocycleur (tableau 19).

| Température | Temps |
|-------------|-------|
| 22°C | 30min |
| 72°C | 10min |
| 10°C | pause |

Tableau 19. Programme de l'étape de ligation

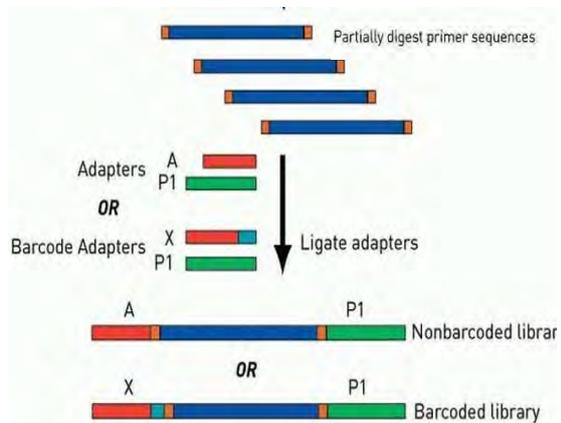


Figure 62. L'étape de ligation.

6.7.4. PURIFICATION SIMPLE

Cette étape consiste à éliminer les réactifs en excès qui n'ont pas été consommés dans les premières étapes de PCR, ligation, et barcodage. Elle nécessite l'emploi de billes magnétiques. Deux différents produits ont été testés : les billes Agencourt AMPure XP® (Beckman Coulter) et NucleoMag® NGS Clean-up and Size Select (Macherey-Nagel).

Ces billes sont équivalentes, elles permettent la purification d'amplicons de 150 à 800pb à leur surface. Selon le rapport : volume billes / volume total de la solution, l'hybridation des amplicons sur les billes sera préférentiellement faite.

Les billes sont retenues par un aimant, et lavées par de l'éthanol absolu à 70% préparé extemporanément. Aux 31 µL des étapes précédentes sont rajoutés 46.5 µL de billes (ratio : 1.5). L'ADN génomique et les différents résidus des premières étapes sont éliminés.

L'élution des amplicons, sélectionnés par les billes et retenues par un aimant, est faite de deux façons différentes :

- **Elution 1 (optionnelle)** : Si l'étape d'amplification des amplicons barre-codés est réalisée, il faut rajouter 52 µL d'un mélange de *Platinum PCR SuperMix High Fidelity* (50 µL) et de *Library Amplification Primer Mix* (2 µL). L'amplification par quelques cycles PCR nécessite l'emploi d'un thermocycleur (tableau 20).

| Température | Temps | Température |
|-----------------|-------|-------------|
| Hold | 98°C | 2 min |
| 5 cycles | 98°C | 15sec |
| | 60°C | 1min |
| Hold | 10°C | Hold |

Tableau 20. Programme de l'étape d'amplification de la librairie barre-codée.



Suivi d'une purification double (optionnelle)

Une nouvelle étape de purification, cette fois-ci en deux temps, est réalisée après l'étape d'amplification de la librairie barre-codée. Un ratio de billes de 0,3 puis de 1,2 permet d'éliminer par de l'éthanol absolu à 70%, les produits PCR restants puis de sélectionner les amplicons de taille attendue.

L'élution finale est réalisée par l'ajout de 50 µL de *low TE* aux billes.

A partir de l'utilisation d'un dosage final des librairies par une technique qPCR cette étape n'a plus été réalisée.

- **Elution 2** : L'élution a depuis été faite par l'ajout de 25 µL de tampon *low TE*.

6.7.5. QUANTIFICATION PAR QUBIT™ ET LABCHIP®

Le LabChip® GX (PerkinElmer) est un automate utilisant la technologie d'électrophorèse capillaire par micro fluide et permet la séparation par la taille des amplicons de librairie.

Il lui est possible de calculer la taille moyenne des amplicons et leur pureté. Les concentrations de librairies étant très faibles, il est utilisé avec le kit *DNA High Sensitivity* qui a une sensibilité de 10 à 500 pg/μL.

La quantification du LabChip® n'étant malgré tout pas assez précise à ces faibles concentrations, les librairies sont aussi dosées par QuBit™.

La conversion en molarité (pmol/L) se fait par la formule suivante :

$$\text{Concentration molaire (pmol/L)} = \frac{[\text{concentration QuBit (ng /}\mu\text{L)} * 10^3]}{\text{Taille moyenne (pb)} \times 6.6 \times 10^{-4}}$$

Après cette quantification, toutes les librairies sont mélangées dans un seul tube de façon équimolaire afin d'obtenir une librairie finale à une concentration de 100 pmol/L. Ce pool final est très important et doit être précis, car il conditionne la qualité de l'amplification clonale suivante (se fait avec un rapport d'amplicon / billes de 1) et le pourcentage d'ISP multiclonales.

Les concentrations finales sont hétérogènes, et peuvent aller d'une centaine à plusieurs milliers de picomoles.

Cette méthode n'est plus utilisée.

6.7.6. QUANTIFICATION PAR QPCR.

Dans le protocole de Life Technologies pour la fabrication de librairie, il est indiqué que l'étape de quantification doit se faire de manière précise par une méthode qPCR. Le laboratoire a pu utiliser un LightCycler® 480 (Roche Applied Science) avec le kit « Ion library taqMan™ Quantification » (Thermo Fisher Scientific) selon le protocole du fabricant. A l'aide d'une gamme étalon, cette méthode TaqMan® rapide et robuste permet de mesurer la concentration des librairies diluées au centième.

Les librairies sont par la suite mélangées de la même façon que citée précédemment.

6.8. AMPLIFICATION CLONALE ET SEQUENÇAGE

L'amplification clonale des banques se fait par PCR en émulsion, sur l'appareil Ion OneTouch2™ (durée : 5h environ 200pb, 8h pour 400pb) selon le protocole fourni par Life Technologies (annexe 6). Il nécessite l'emploi d'un *Ion PGM™ Template OT2 kit*. Le succès de cette étape est primordial, car seule la configuration d'un amplicon avec une ISP est valable et génératrice de séquences analysables. Les billes polyclonales généreront des données discordantes (2 signaux différents) qui seront filtrées lors de l'analyse primaire.

L'enrichissement des ISPs positives (ayant incorporé au minimum un fragment) est faite à l'aide de l'Ion OneTouch ES™ et permet le lavage des billes et leur capture par liaison à la streptavidine (durée : 35 minutes environ).

Le contrôle qualité par QuBit™ comme décrit précédemment est ensuite réalisé. Les conditions optimales pour le run de séquençage et un bon remplissage de la puce sont :

- Pourcentage de *template ISP* attendu: 10% à 25%. En dessous, trop d'ISP sont vides, au-dessus trop d'ISP sont polyclonales.
- Valeurs absolues attendues : AF488 > 3000 (ISP totales) et AF647 > 1000 (ISP chargées), en dessous il n'y a pas assez d'ISP pour remplir la puce de façon optimale.

En amont de l'étape de séquençage, une initialisation du PGM™ est requise et permet une homogénéisation des réactifs de l'appareil à une valeur de pH de 7,8.

Une fraction de billes « tests fragments » est également ajoutée aux ISPs purifiées. Elles permettent un contrôle qualité interne au séquençage grâce à leur séquence connue.

Le séquençage sur l'automate PGM™ nécessite l'utilisation d'un *Ion PGM™ Sequencing kit*. Ce kit, tout comme le *Ion PGM™ Template OT2 kit*, a été d'abord commercialisé pour produire des amplicons de 200 pb, puis disponible spécifiquement pour les amplicons de 400 pb. Une chimie dite « Hi-Q » est aussi apparue.

L'initialisation du PGM™, et le chargement de la puce ont été réalisés selon les conditions données par Life Technologies (annexe 6).

6.9. CHOIX DE LA PUCE

Trois puces de différentes capacités sont disponibles (Ion 314™, 316™ et 318™). Elles permettent toutes le séquençage d'amplicons de 200 ou 400 pb.

Le choix de la puce à utiliser est dépendant de la taille de la librairie, du nombre de patients que l'on veut séquençer et de la profondeur de couverture souhaitée. Dans notre cas, une profondeur minimum de 100X est requise afin d'avoir une confiance suffisante dans l'étape d'appel de variant.

Pour la première librairie, de 36.52 kb, une profondeur de 100X pour un patient correspond au séquençage de 3.6 Mb. Cette estimation minimale, nous permettait une exploitation confortable de 16 patients sur une puce Ion 316™ avec des *reads* de 200 pb (57,6 vs 300 Mb générés dans la configuration basse de la puce 316™).

Il semble cependant que l'efficacité de la PCR lors de la première étape de fabrication de librairie soit très hétérogène. Après plusieurs essais avec le 1^{er} design, un certain nombre de « trous » de couverture ont été détectés, à cause d'amplicons sous-représentés (PCR peu efficace). Pour la deuxième librairie de 43 kb, les puces Ion 318™ de plus grande capacité ont été adoptées afin d'améliorer les défauts de couverture et rattraper l'efficacité modeste de certaines réactions PCR. Pour le passage de 24 patients dans cette configuration avec une chimie 400pb cette fois, la capacité relative de la puce est doublée par rapport à une 316™.

| Ion PGM™ Chip | Durée d'un run | | Capacité de sortie | | |
|------------------|----------------|-------------|--------------------|-------------|---|
| | 200 bp read | 400 bp read | 200 bp read | 400 bp read | |
| Ion 314™ Chip v2 | 2.3 hr | 3.7 hr | 30–50 Mb | 60–100 Mb |  |
| Ion 316™ Chip v2 | 3.0 hr | 4.9 hr | 300–500 Mb | 600 Mb–1 Gb |  |
| Ion 318™ Chip v2 | 4.4 hr | 7.3 hr | 600 Mb–1 Gb | 1.2–2 Gb |  |

Tableau 21. Les puces disponibles pour le PGM™.

7. TRAITEMENT BIO-INFORMATIQUE

La plateforme Life Technologies possède un système de server (Ion Torrent Server) qui est contrôlable via une interface graphique: l'**Ion Torrent Browser**. Par l'usage d'un explorateur internet, il est possible de préparer sa feuille de route, et après le séquençage, de lancer des analyses (ou *plug in*) puis de télécharger les fichiers générés.

7.1. LE RAPPORT DE RUN

Un rapport de run est généré après le séquençage. Il apporte plusieurs éléments : les indicateurs qualité, la configuration du run et les versions des logiciels utilisés.

7.1.1. INDICATEURS DE QUALITE DU SEQUENÇAGE

Un certain nombre d'éléments du rapport de run reflète la **quantité** de données produites :

- La densité de la puce (*ISP loading* ou *Density*) : c'est le pourcentage de puits qui contiennent une ISP.
- Le nombre total de bases générées (*Total Bases*) et de lectures (*Total Reads*).
- Le pourcentage de lectures utilisables (*Usable Reads*): correspond au pourcentage de lecture après soustraction des lectures d'ISPs polyclonales, des tests fragment, des dimères d'adaptateurs, et les lectures de mauvaise qualité.

Et de la **qualité** des données produites :

- Le signal « clé » (*Key signal*) : contrôle interne au séquençage.
- Une moyenne, une médiane et le mode de la longueur des lectures.

Run Summary

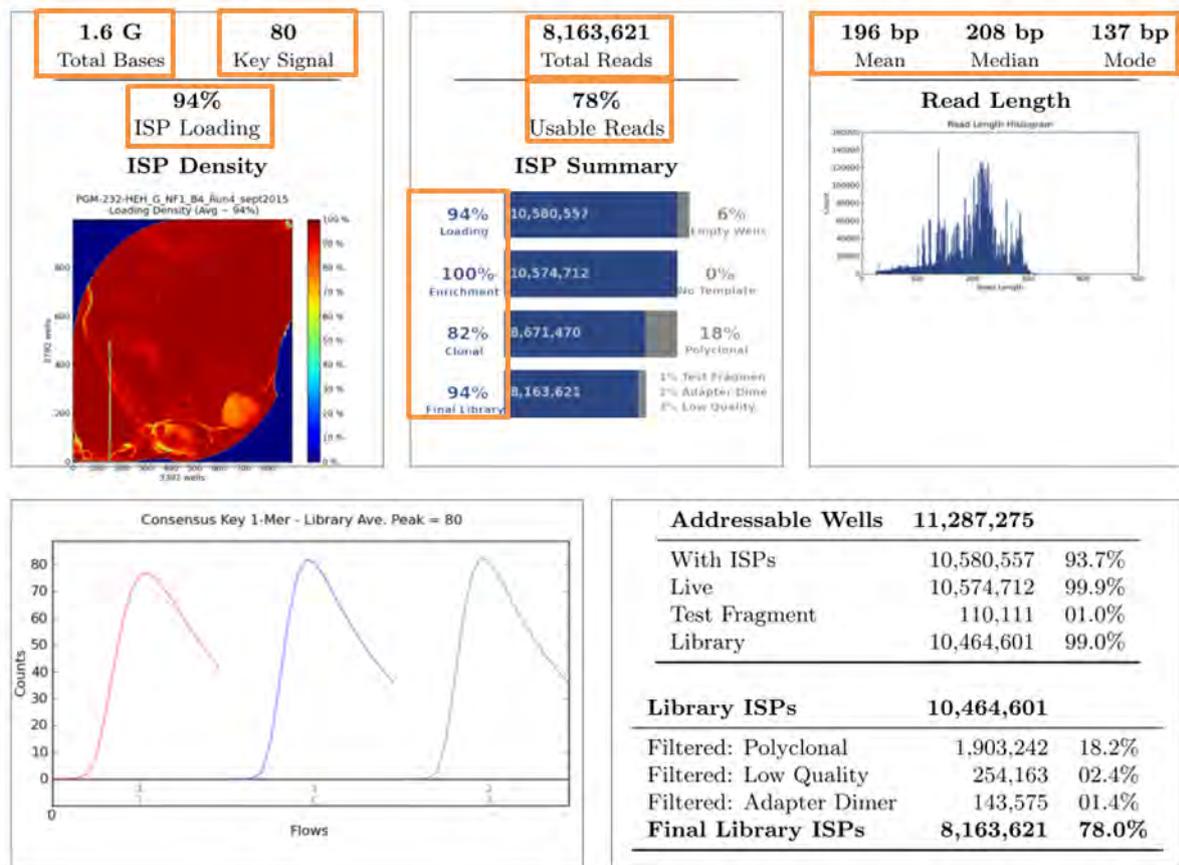


Figure 63. Exemple de rapport de run d'un séquençage (indicateur qualité de séquençage).

Pour chaque échantillon (ou barre-code), on retrouve :

- Le nombre total de bases séquencées.
- Le nombre total de bases séquencées appelées avec une qualité supérieur à Q20 (taux d'erreur de 1%).
- Le nombre de lectures.
- La moyenne de taille des lectures.

On peut d'ores et déjà s'assurer à cette étape que tous les codes-barres ont été bien séquencés et démultiplexés.

7.1.2. INDICATEURS QUALITE DE L'ALIGNEMENT

L'alignement est fait directement à la suite de l'appel de base. Il est réalisé sur le génome humain dans son ensemble (Hg19).

Plusieurs critères qualité de l'alignement sont utiles et inscrits dans le rapport de run:

- Le pourcentage et le nombre de bases totales alignées (*Total Alignment Bases*).
- Le nombre de bases avec un score qualité d'alignement (AQ, basé sur le score Phred) supérieur à AQ17 (taux d'erreur 2%), AQ20 et parfait (AQ47).
- La longueur moyenne et maximale des lectures avec un score AQ17 et AQ20.
- La précision moyenne de l'alignement par base (*Mean raw accuracy 1x*): correspond au nombre de *mismatch* moyen par base.

Enfin plusieurs graphiques permettent de contrôler la répartition de la taille des séquences.

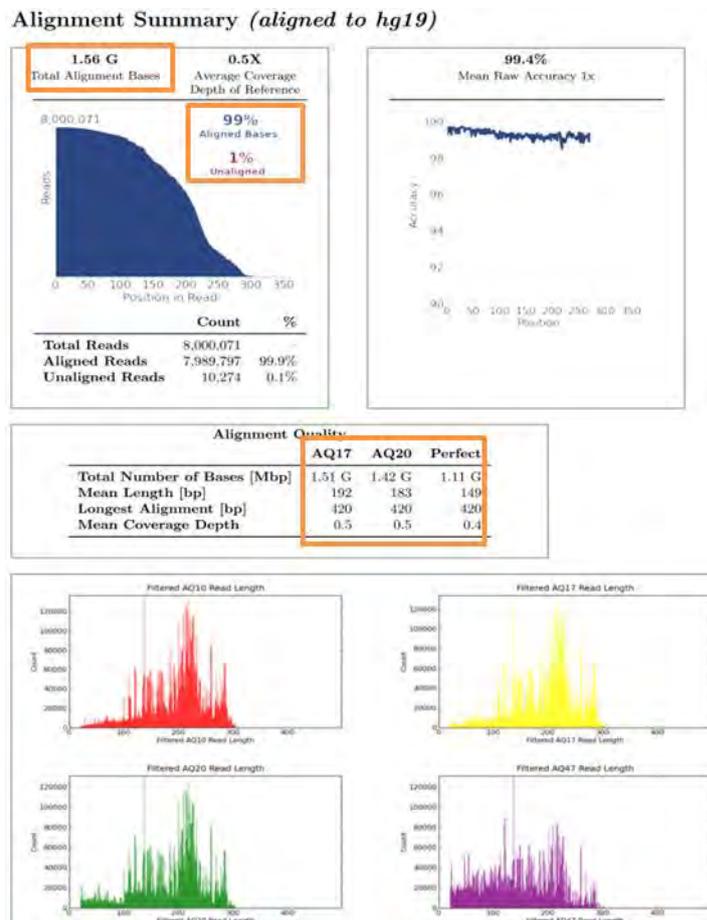


Figure 64. Exemple de rapport de run d'un séquençage (indicateur qualité de l'alignement).

7.1.3. INDICATEURS QUALITE DES SEQUENCES

L'outil FastQC (Brabraham Insitute) permet le contrôle qualité des séquences générées par millions lors d'un run. Il produit plusieurs graphiques donnant des informations diverses qui peuvent mettre en lumière des erreurs survenues lors de la création de librairie ou au séquençage. Il est possible d'obtenir ces graphiques pour l'ensemble des patients ou de manière séparée. Pour chaque échantillon, les résultats du FastQC ont été analysés, parmi eux :

- **Per base sequence quality (qualité par base):**

Comme décrit précédemment, ce graphique représente les scores qualités des bases en fonction de leur position dans les lectures. Il permet de contrôler l'épuisement de la Taq polymérase. La qualité des bases à l'extrémité des séquences est naturellement moindre que pour le corps central.

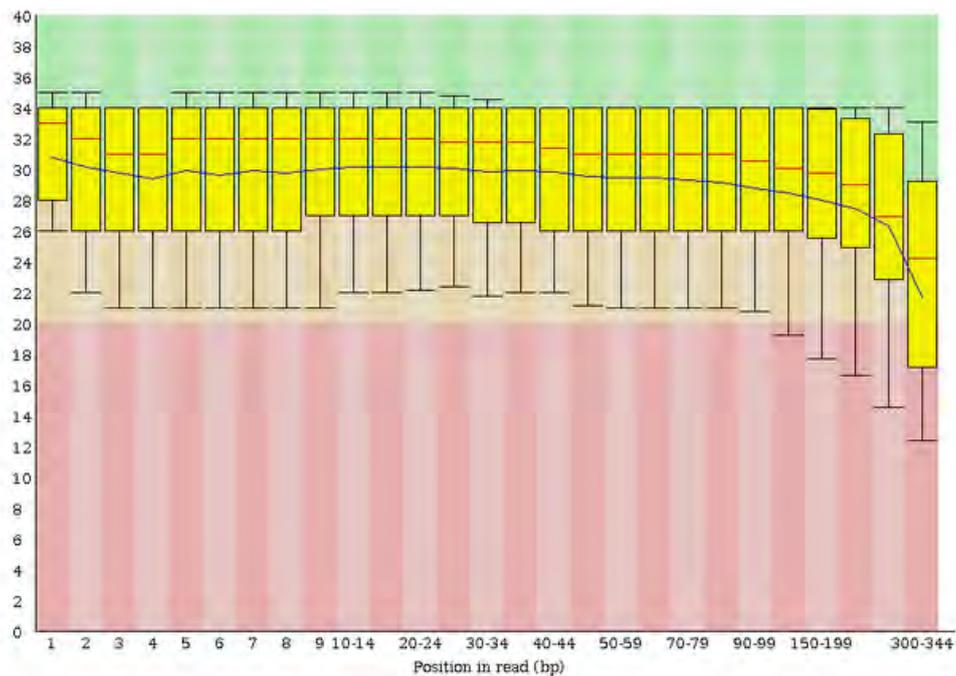


Figure 65. *Per base sequence quality* de bonne qualité d'un séquençage 400pb.

En fond vert les valeurs supérieures à Q30, en orange entre Q20 et Q30, en rouge inférieur à Q20 (non souhaitable).

- **Per sequence quality scores (qualité par séquence) :**

Le graphique correspond à la distribution de la qualité moyenne des lectures. Il est préférable que cette courbe soit centrée sur un score Phred le plus grand possible.

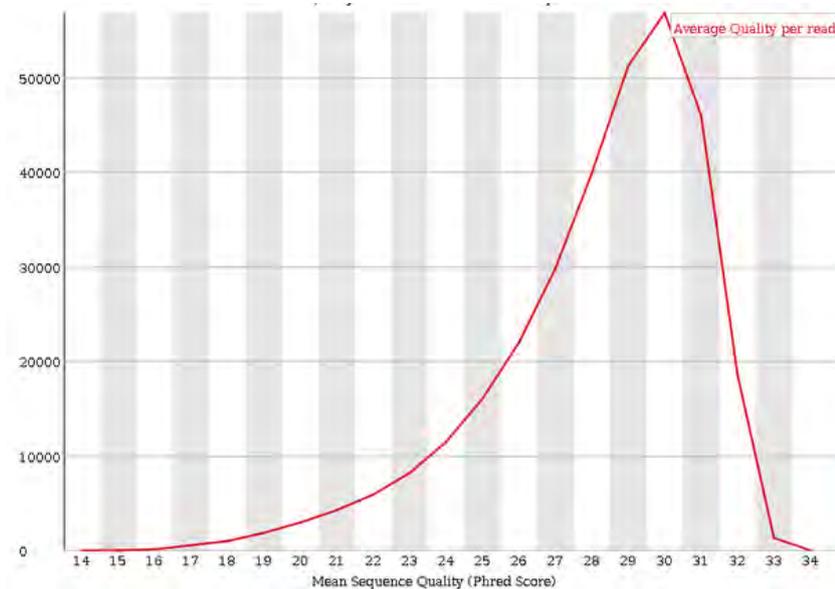


Figure 66. Distribution de la qualité moyenne des lectures.

- **Per base sequence content (distribution des bases):**

Le graphique représente la distribution moyenne des nucléotides selon leur position dans la lecture. Dépendante des séquences analysées, cette distribution doit être fixe entre les différentes séries d'un même panel.

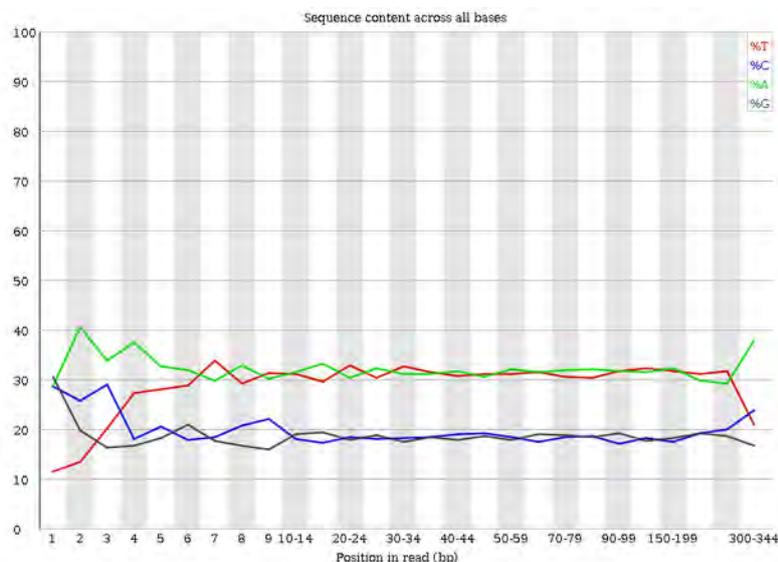


Figure 67. Distribution moyenne des bases à travers les lectures.

- **Per Base / Per sequence GC content**

Le pourcentage de GC dans un amplicon influence la qualité et la quantité des lectures produites lors de l'étape de séquençage. Les deux graphiques ci-dessous représentent la distribution moyenne en GC par base et par séquence. Les taux doivent être constants intra-patients et inter-runs.

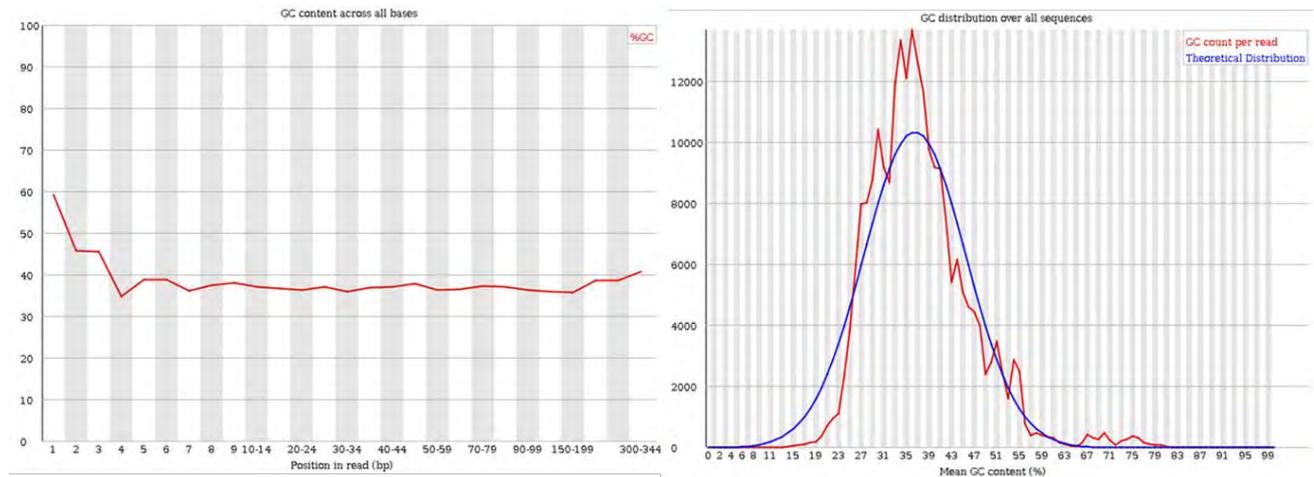


Figure 68. Distribution taux de GC par base (gauche) et par lectures (droite).

- **Sequence length distribution :**

Similaire à ce qui se trouve dans le rapport produit à la fin du séquençage, ce graphique représente la distribution des longueurs de lecture. Spécifique d'un design, il doit être constant intra-patients et inter-runs. Si la présence de dimères d'amorces est importante, on apercevrait un pic de faible taille (signifiant une mauvaise purification).

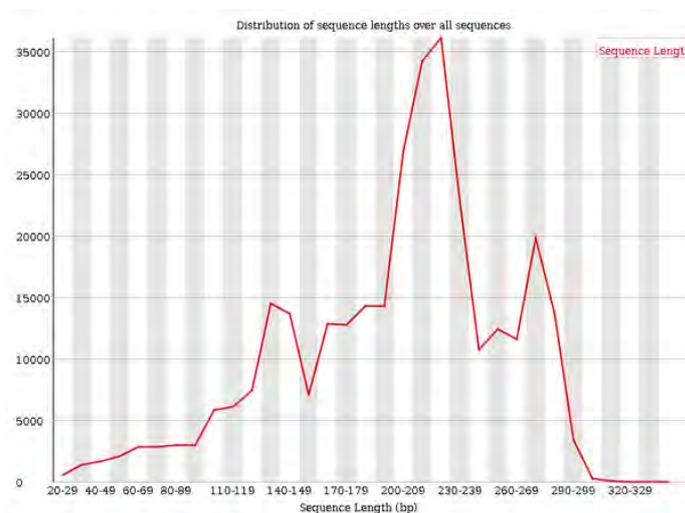


Figure 69. Distribution des longueurs des lectures.

- **Overrepresented sequences**

Pour finir, un tableau rapporte le compte brut et le pourcentage associé de chaque séquence par rapport aux séquences totales. Il permet de repérer les dimères d'amorces, adaptateurs, ou toutes autres séquences aspécifiques de notre librairie qui seraient surreprésentées.

7.1.4. CRITERES D'ACCEPTATION

Après un certain nombre d'essais, des seuils d'acceptabilité ont été inscrits dans notre protocole afin de pouvoir valider techniquement le run de séquençage. Life Technologies ne fixe pas de limite, mais ces paramètres sont pour nous les plus représentatifs du succès technique d'un run de séquençage.

- **Limites pour les critères qualité du séquençage et de l'alignement**

| critère | Accepté si: |
|-----------------|---|
| Total bases | > 1G pour 318 >500M pour 316 |
| Key signal | > 70 |
| Density | > 50 % |
| Enrichment | proche 100% |
| Total Reads | >6 000 000 pour 318 > 3 000 000 pour 316 |
| Usable reads | > 60 % |
| Polyclonal | < 40 % |
| Test Fragment | < 4% |
| Mean lenght | [150-180] pour le 1er design [180-205] pour le 2ème design |
| unaligned bases | < 3 % |
| Raw accuracy 1x | > 99 % |

- **Limites pour les critères qualités des séquences**

| Critère | doit être: |
|--------------------------------|--|
| Per base sequence quality | > 30 sauf à l'extrémité (baisse normale) |
| Per sequence quality scores | gaussienne avec épaulement à gauche (moyenne qualité à 30) |
| Per base sequence / GC content | A/T à 30% - C/G à 20% sauf à l'extrémité |
| Sequence Length Distribution | Pics similaires pour tous les patients du même run |

Tableau 22. Critères qualité du séquençage et de l'alignement et leurs limites correspondantes.

7.2. ANALYSES DU LABORATOIRE

Le *Torrent Browser* permet de télécharger le rapport de run, mais aussi de lancer des analyses à travers des plug-in directement intégrés. Ces plug-in sont des scripts libres de droit qui sont ajoutés par défaut par Life Technologies mais peuvent aussi être des scripts personnels importés.

Le *Server* met aussi à disposition différents fichiers qui permettent à l'utilisateur de les stocker, ou de lancer des analyses externes de son propre chef.

Trois plug-ins différents sont utilisés par notre équipe.

7.2.1. PLUG-IN : FILE EXPORTER

Ce plug-in n'est lié à aucune analyse mais permet de récupérer deux types de fichiers :

- Les fichiers **BAM** de chaque patient après l'alignement (*.bam et *.bai). Ils peuvent être ouverts par un logiciel de visualisation (IGV, Alamut™) et donner un aperçu de l'alignement des lectures.

Cependant, en **aucun cas** ces logiciels de visualisation ne **remplacent** l'étape de **variant calling ou de l'analyse des profondeurs**.

L'analyse des variants ne doit pas se faire directement par leur biais.

- Les fichiers **FASTQ** de chaque patient. ils sont composés des lectures non alignées mais filtrées par défaut (par la qualité, la taille) et après le *triming* des adaptateurs. Ils permettent par exemple, de réaliser un nouvel alignement avec un logiciel externe au *Torrent browser*.

7.2.2. PLUG IN : COVERAGE ANALYSIS

Ce script permet de générer des statistiques et des graphiques informatifs sur la profondeur de couverture des régions d'intérêts.

Il demande de renseigner les régions cibles par l'utilisation d'un fichier BED. Après analyse, de nombreux paramètres sont renseignés dans des fichiers Excel.

L'utilisation de ce plug-in est double :

- D'une part, il permet de calculer le nombre de lectures qui sont alignées sur les régions d'intérêt modifiées. Elles correspondent aux régions de notre design auxquelles nous avons supprimé les zones chevauchantes entre deux ou plusieurs amplicons.

Le fichier Excel **NC** (Non Chevauchant) qui est téléchargé sert à une analyse des CNV des deux gènes par méthode de « *Count Reads* ».

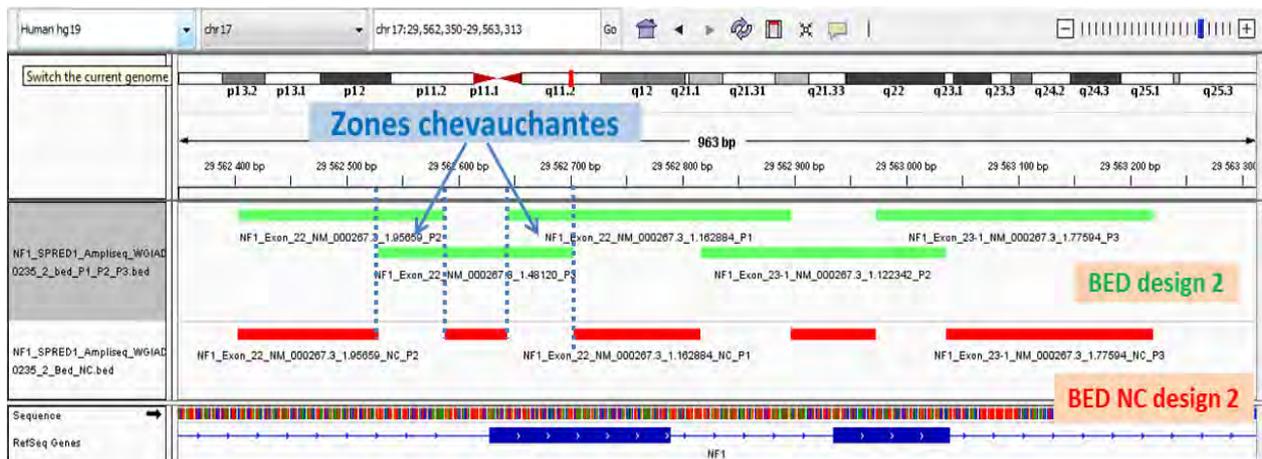


Figure 70. Représentation d'amplicons non chevauchants sur IGV (en rouge).

- D'autre part, il permet de contrôler le pourcentage de lectures qui s'alignent sur les pseudogènes de NF1 (par l'emploi d'un *BED pseudogènes*) et de quantifier aussi l'aspécificité de l'alignement.

7.2.3. PLUG IN : VARIANT CALLER

Comme son nom l'indique, le *Variant Caller* permet l'appel des variants. Pour cela, il utilise des paramètres de filtre par défaut (*low-stringency*) qui peuvent être modifiés manuellement au lancement de l'analyse, ou par l'intermédiaire d'un fichier portant les nouveaux paramètres : un fichier *.json*.

Pour éviter un maximum de faux négatif, ce fichier *.json* a été modifié plusieurs fois à mesure des problèmes rencontrés lors cette dernière étape.

Une première version (V1) a été adoptée par l'intermédiaire de l'équipe de notre laboratoire travaillant sur les gènes *BRCA*. Ces gènes se distinguent comme *NF1* par un fort taux d'homopolymères, véritable talon d'Achille du séquençage Ion Torrent.

Une autre modification sur un paramètre de l'algorithme Freebayes a donné une troisième et dernière version des paramètres du variant caller (V2). Elle a permis de détecter dans un autre panel utilisant le PGM™, un variant qui n'était jusqu'à présent non appelé.

Comme on peut le voir dans le tableau 23, deux paramètres principaux ont été modifiés :

- Le paramètre ***strand bias***, qui n'est plus pris en compte (en le fixant à 1). Ceci permet d'éviter les faux négatifs qui ne seraient portés que par des *reads* d'un seul et même sens.
- Le paramètre ***allow_MNPs*** (*Multiple nucleotide polymorphisms*): il permet de détecter (si fixé à 0) ou non (si 1) plusieurs *mismatches* qui se trouvent dans un environnement proche. Ce paramètre est similaire au ***allow_Complex*** (si l'un est activé l'autre doit être désactivé), à la différence près que ***allow_Complex*** permet de générer des hypothèses de « *block* » de différentes tailles alors que ***allow_MNPs*** ne gère que les hypothèses de « *block* » de même taille. Les paramètres de la V2 (*allow_MNPs*) sont toujours utilisés pour générer les fichiers VCF.

Ces changements, moins stringents, augmentent considérablement le nombre de faux-positifs. Cependant, dans le cadre d'une analyse diagnostic, la priorité de la méthode est mise sur la sensibilité.

| Parameter | Germline | | | Parameter | Germline | | |
|-------------------------------|----------------|----|----|-------------------------------|----------------|----|----|
| | Low Stringency | V1 | V2 | | Low Stringency | V1 | V2 |
| torrent_variant_caller | | | | long_indel_assembler | | | |
| data_quality_stringency | 6.5 | | | kmer_len | 19 | | |
| do_mnp_realignment | 1 | | | min_var_count | 5 | | |
| do_snp_realignment | 1 | | | short_suffix_match | 5 | | |
| downsample_to_coverage | 4000 | | | min_indel_size | 4 | | |
| filter_deletion_predictions | 0.2 | | | max_hp_length | 8 | | |
| filter_insertion_predictions | 0.4 | | | min_var_freq | 0.15 | | |
| filter_unusual_predictions | 0.7 | | | relative_strand_bias | 0.8 | | |
| heavy_tailed | 3 | | | freebayes | | | |
| hotspot_min_allele_freq | 0.1 | | | allow_complex | 0 | 0 | 1 |
| hotspot_min_cov_each_strand | 3 | | | allow_indels | 1 | | |
| hotspot_min_coverage | 6 | | | allow_snps | 1 | | |
| hotspot_min_variant_score | 5 | 10 | 10 | allow_mnps | 1 | 0 | 0 |
| hotspot_strand_bias_pval | 1 | | | min_mapping_qv | 4 | | |
| hotspot_strand_bias | 0.98 | 1 | 1 | read_snp_limit | 10 | | |
| hp_max_length | 10 | | | read_max_mismatch_fraction | 1 | | |
| indel_as_hpindel | 0 | | | gen_min_alt_allele_freq | 0.15 | | |
| indel_min_allele_freq | 0.1 | | | gen_min_indel_alt_allele_freq | 0.15 | | |
| indel_min_cov_each_strand | 5 | | | gen_min_coverage | 6 | | |
| indel_min_coverage | 15 | | | | | | |
| indel_min_variant_score | 10 | | | | | | |
| indel_strand_bias | 0.85 | 1 | 1 | | | | |
| indel_strand_bias_pval | 1 | | | | | | |
| mnp_min_allele_freq | 0.1 | | | | | | |
| mnp_min_cov_each_strand | 0 | | | | | | |
| mnp_min_coverage | 6 | | | | | | |
| mnp_min_variant_score | 10 | | | | | | |
| mnp_strand_bias | 0.95 | 1 | 1 | | | | |
| mnp_strand_bias_pval | 1 | | | | | | |
| outlier_probability | 0.01 | | | | | | |
| position_bias | 0.75 | | | | | | |
| position_bias_pval | 0.05 | | | | | | |
| position_bias_ref_fraction | 0.05 | | | | | | |
| prediction_precision | 1 | | | | | | |
| realignment_threshold | 1 | | | | | | |
| snp_min_allele_freq | 0.1 | | | | | | |
| snp_min_cov_each_strand | 0 | | | | | | |
| snp_min_coverage | 6 | | | | | | |
| snp_min_variant_score | 10 | | | | | | |
| snp_strand_bias | 0.95 | 1 | 1 | | | | |
| snp_strand_bias_pval | 1 | | | | | | |
| suppress_recalibration | 0 | | | | | | |
| use_position_bias | 0 | | | | | | |

Tableau 23. Comparaison des différents paramètres utilisés pour le Variant Caller.

Nous avons produit deux autres fichiers VCF spécifique de type « *Hotspot* ».

Ils permettent de forcer un appel de variant à chaque position renseignée par le fichier d'entrée. Les fonctions de ces deux fichiers sont différentes :

- Le fichier *Hotspot* « *Polymorphismes* » :

il contient les positions génomiques des polymorphismes les plus fréquents des gènes *NF1* et *SPRED1*. Ces polymorphismes ont été extraits de la base de données dbSNP¹⁸⁴ et filtrés (les SNPs « validés » sont gardés). Encore non utilisée à ce jour, la compilation de ces fichiers sur plusieurs séries permettra dans un futur proche de faire des analyses de concordance génotype-phénotype et constitue pour l'instant une simple base de données.

- Le fichier Hotspot base à base :
Il contient toutes les bases des régions exoniques, UTR, et les 20 bases introniques flanquantes pour nos deux gènes. Au total, 22 510 lignes sont renseignées.

Ce fichier, issu d'une réflexion interne, permet à lui seul de contenir, après analyse par le *Variant Caller*, deux informations très importantes pour la validation du run :

- D'une part, il sert à obtenir la profondeur de couverture au stade de l'appel de variant, c'est-à-dire la profondeur « efficace ». Cette profondeur peut différer de celles que l'on peut obtenir par les logiciels de visualisation après lecture du BAM aligné (filtres du *Variant Caller* non pris en compte).

Cette astuce informatique nous permet par la suite de filtrer les bases, ou régions, qui n'ont pas une couverture suffisante pour assurer un appel de variant avec une **confiance suffisante** pour le rendu d'une analyse constitutionnelle.

Si la profondeur est en dessous de la limite fixée (100X), l'exon auquel la base appartient sera séquencé par méthode Sanger (pour augmenter la sensibilité encore une fois).

- D'autre part, en forçant l'appel de variants sur toutes les bases, nous pouvons récupérer les variants qui n'ont pas passé les filtres du *Variant Caller*. Ces variants qui sont mis « à la poubelle » sont dits : « **no call** ». Le passage par cet outil est nécessaire du fait que l'information des variants non retenus par le *Variant Caller* n'est pas accessible (facilement) sur le *Torrent Server*. Plusieurs variants peuvent ainsi devenir des faux négatifs sans cette intervention, et donc être manqués par le biologiste sans cette solution.

Logiquement, il faut rechercher la cause (le paramètre) qui filtre les *no call*, afin d'améliorer le paramétrage du *Variant Caller* et gagner en sensibilité et spécificité (et donc modifier *in fine* le fichier json).

Des bases porteuses de variants *no call* peuvent, selon l'analyse du biologiste, déclencher un séquençage Sanger afin de vérifier leur présence ou non.

7.3. TRAITEMENT BIOINFORMATIQUE DES FICHIERS VCF OBTENUS

Les fichiers VCF dit *Base à base*, et le fichier de variants ont deux parcours bioinformatique différents.

7.3.1. PARCOURS DU FICHIER VCF *BASE A BASE*

Le fichier Base à base est obtenu après plusieurs dizaines d'heures d'analyse du *Variant Caller*, nous obligeant à modifier le temps maximum imparti par défaut pour ce plug in (de 24h à 72h). Pour 24 patients, il correspond à un tableau Excel d'environ 540 000 lignes et 48 colonnes et pensant environ 180 Mo. Cet imposant fichier nous a montré les limites d'Excel (Excel (1 million de lignes) et les limites des ordinateurs standards (plus d'une minute pour l'ouvrir avec notre ordinateur conçu pour l'analyse bioinformatique).

Nous avons donc choisi un logiciel de gestion de base de données assez puissant pour visualiser l'analyse (FileMaker Pro, FileMaker Inc.). Le fichier Base à base est ainsi incorporé au logiciel, avec la liste de travail du run contenant le nom des patients et le fichier VCF de variants.

7.3.2. PARCOURS DU FICHIER VCF *VARIANT*

Le fichier VCF *Variant* contient l'ensemble des variants de tous les échantillons d'une série. Un traitement informatique en deux étapes permet de filtrer, et d'annoter les variants avant l'analyse finale par le biologiste.

La première étape de filtre se fait grâce à une macro Excel, elle permet dans des nouvelles colonnes :

- De rapatrier le nom du patient correspondant pour chaque ligne du fichier.
- De rapatrier des informations tirées de la base de données dbSNP.
- De calculer l'occurrence intra-run pour chaque variant (ligne) : les variants avec une occurrence inférieure ou égale à 2 sont regardés avec un intérêt particulier.
- De traduire le variant en nomenclature type HGVS (Human Genome Variation Society).
- De générer un fichier texte tabulé intermédiaire, qui permettra l'import de tous les variants du VCF dans la base de donnée interne au logiciel Alamut™.

La deuxième étape consiste à annoter ce fichier texte en utilisant Alamut™. Ce logiciel intègre directement des informations issues de différentes bases de données et utilise des logiciels de prédiction *in silico* afin d'aider à l'interprétation des variants (Align GVDG, SIFT, MutationTaster, Polyphen-2). Il permet aussi la traduction des variants en nomenclature cdna et protéique, ainsi que l'information du génome de référence et du transcrit d'analyse choisie.

Son emploi se déroule comme suit :

- Le fichier texte intermédiaire est importé dans Alamut™ (onglet mutation -> importer des mutations) pour chaque gène séparément.
Ces mêmes mutations sont exportées dans un fichier texte tabulé (onglet mutation -> exporter les mutations internes). Le logiciel annote chaque ligne, qui porte toujours le nom du patient.
- Les deux fichiers obtenus sont ensuite compilés en un seul fichier qui sera intégré au fichier Excel de la première étape.
- Une nouvelle macro permet d'intégrer les données Alamut™ et de mettre en forme un nouveau fichier comportant :
 - Le nom des patients
 - Les variants
 - Les annotations du variants
- Le fichier est sauvegardé dans un nouveau classeur Excel.
- Ce fichier peut être directement analysé ou comme pour le fichier *Base à base*, être importé dans le logiciel FileMaker Pro.

La procédure a été en partie automatisée avec la création de macro VBA Excel, cependant elle comportent plusieurs étapes manuelles d'import-export avec de nombreux copier-coller.

Il faut environ 30 min pour réaliser le traitement complet du fichier VCF. Ces étapes ne demandent pas de connaissances particulières en bio-informatique à l'emploi. Il faut néanmoins être vigilant en pratique, afin d'éviter de possibles erreurs humaines difficiles à corriger par la suite.

7.4. FINALITES ET ANALYSE DU RUN PAR LE BIOLOGISTE

Le biologiste a désormais dans un même logiciel (FileMaker Pro) :

- La liste des patients du run validés techniquement (fichier *liste de travail*).
- L'information sur les profondeurs de couverture de chacune des bases porteuses d'information génétique probante avec les variants qui n'ont pas passés les filtres du *Variant Caller* et la raison de leur discrédit (fichier *base à base*).
- Des variants appelés par le *Variant Caller* et de leurs annotations Alamut™ correspondantes.

Toutes ces informations sont lisibles dans une même fenêtre de lecture. Le biologiste peut choisir quel paramètre analyser et sur quel patient. L'outil est puissant car il permet aussi de comparer rapidement tous les patients d'un même run sur un seul écran.

Le logiciel est aussi programmé pour pouvoir :

- Attribuer une annotation (ou « Flag ») aux variants : faux positifs (récurrents), polymorphismes, autre. Ce *flag* alimente une base de données incrémentée directement dans FileMaker. Cette base dynamique permet, pour les séries suivantes, d'améliorer la décision du biologiste qui prendra en compte les informations déjà enregistrées.
- Construire des listes de travail de séquençage Sanger pour le technicien:
 - o Des variants sélectionnés par le biologiste qui sont à confirmer.
 - o Des exons insuffisamment couverts et donc à vérifier. Une base suffit pour déclencher le contrôle de tout un exon.

7.5. STOCKAGE

Le stockage des données est un paramètre à prendre en compte très tôt dans le développement et passage en routine de méthodes NGS. Il peut rapidement devenir conséquent et poser des problèmes de logistiques.

Plusieurs modes de stockages doivent être mises en place :

- Le premier système de stockage est le disque dur du PGM™ lui-même. Environ 11To d'espace sont dédiés au stockage des données produites par les équipes des HCL. Si cela peut paraître important à première vue, un programme d'export de données rigoureux à quand même dû être configuré afin de ne pas saturer la machine par les différentes équipes.
- Un deuxième espace de stockage permet d'exporter les données et de désengorger rapidement le PGM™ quand les résultats ont été produits et analysés par les plug-ins Ion Torrent™ Server. Il permet aussi de travailler sur les données avec d'autres logiciels.
- Enfin, un dernier espace est alloué au stockage à long terme, une fois que le run est complètement validé. Ces données peuvent ne pas être atteignables directement par le biologiste mais par l'équipe informatique qui contrôle cet espace.

Tous ces transferts découlent d'une organisation établie à l'avance, avec un accord préalable entre toutes les équipes. De plus, des contrôles de l'intégrité des données sont effectués afin de vérifier les transferts entre disques (une corruption des fichiers peut arriver et être identifiée bien après).

Le stockage des données NGS demande une bonne organisation entre les services utilisant la plateforme de génétique et les services informatiques. De plus, il représente un prix conséquent qu'il ne faut pas négliger dans les coûts globaux.

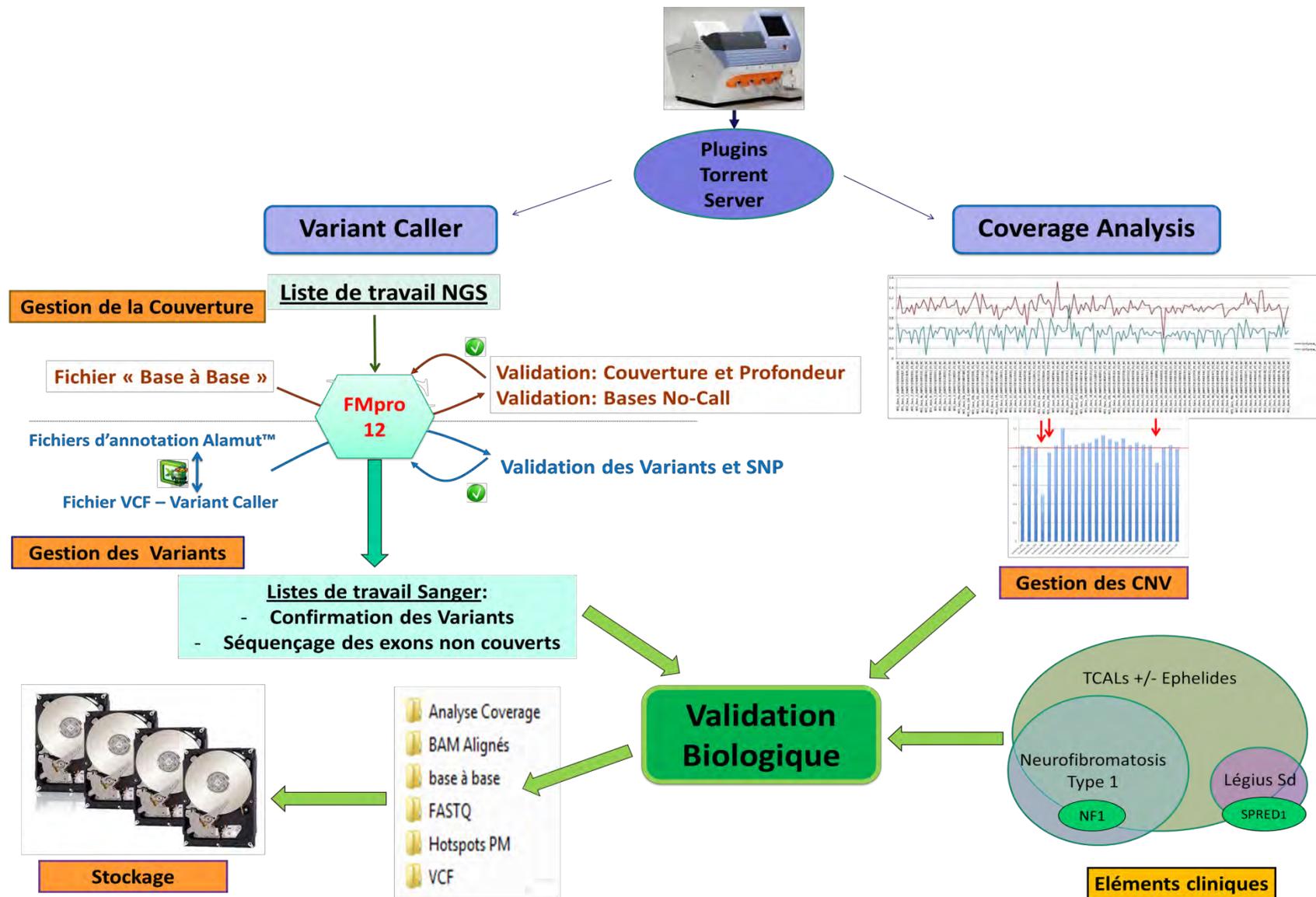


Figure 72. Schéma simplifié du traitement bio-informatique à la validation et stockage d'un run NGS.

8. ECHANTILLONS SEQUENCES

8.1. PREMIER DESIGN

Après plusieurs séries réalisées afin de mettre au point un protocole de séquençage satisfaisant, cinq runs de validation ont été effectués. Au total, 48 patients ont été séquencés sur les quatre premiers essais (par série de 16, dont une série répétée) et 24 patients pour le dernier, pour un total de 72 patients.

Tous les patients ont été génotypés en parallèle par la méthode de référence actuelle dHPLC- Sanger – MLPA. Les variants trouvés seulement par NGS ont été vérifiés en Sanger. Les patients avec une délétion partielle ou totale (MLPA positive) n'ont pas été analysés en dHPLC et Sanger.

Plusieurs paramètres ont été modifiés à la suite des premiers séquençages. Les billes servant à la purification étaient des billes Agencourt AMPure® XP pour le premier, puis des NucleoMag® NGS Clean-up and Size select par la suite (moins cher pour des résultats identiques). Deux quantifications différentes des bibliothèques finales ont été testées : tout d'abord la quantification par QuBit™ et Labchip® pour les deux premières séries, puis par LightCycler® pour les suivantes. Les trois premières séries sont passées sur des puces 316™ quand les suivantes ont été réalisées avec des puces 318™ (figure 73).

En parallèle, Life Technologies fit évoluer au fil du temps plusieurs réactifs de sa plateforme Ion Torrent™. Les kits de réactifs pour l'OT2™ et pour le séquençage ont évolués, pour prendre le nom de kit « Hi-Q » (pour 200pb puis 400pb). La Torrent Suite bénéficia aussi de plusieurs mises à jour (passage de la version 4.x.x à la version 5.0.2 pour la dernière actuelle).

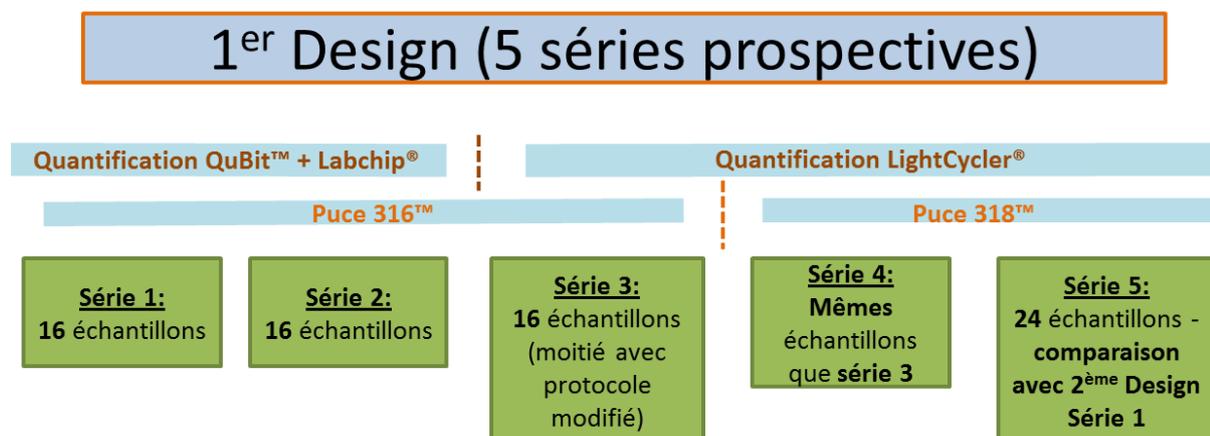


Figure 73. Les séries de validation du design 1

8.2. SECOND DESIGN

Au total, 4 séquençages de banques sur puces Ion 318™ par série de 24 échantillons ont été réalisés. La série 1 est miroir de la série 5 du premier design (prospective). La série 3 a permis de séquencer de nouveaux patients. Enfin, les séries 2 et 4 ont été faites avec des échantillons déjà connus et génotypés.

Pour tous les patients, les méthodes classiques ont été réalisées en parallèle, comme pour les séries du premier design. De la même façon, les résultats des deux méthodes sont analysés et comparés.

Pour ces nouvelles séries, le protocole validé avec quantification LightCycler® a été utilisé. Les kits de réactifs étaient tous « Hi-Q » et la chimie 400 pb a été employée.

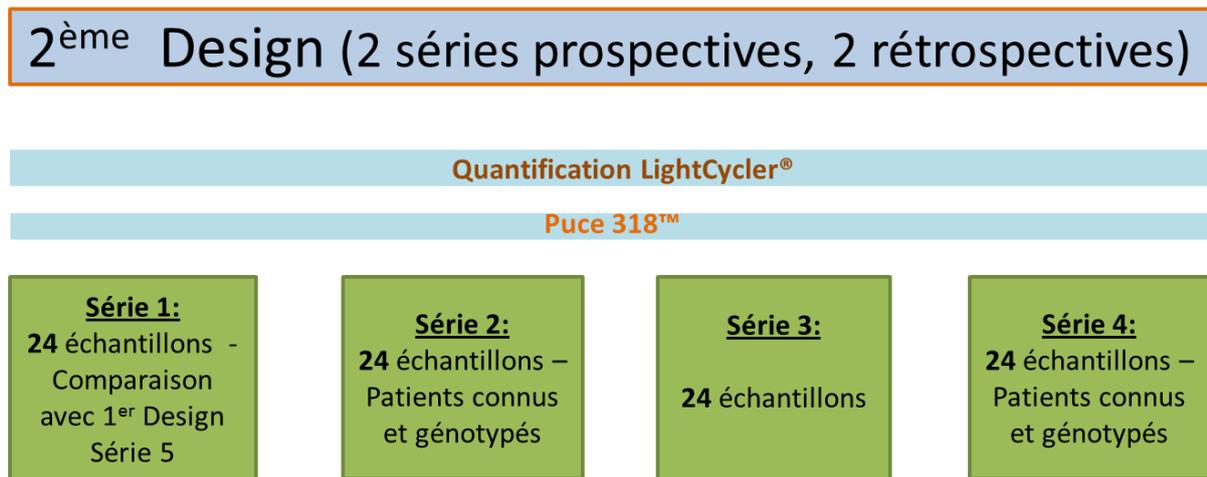


Figure 74. Les séries de validation du design 2.

Certains échantillons, qui ont mis en défaut la détection des variants par le premier design, ont été séquencés à nouveau dans la série 2 : les 3 patients porteurs d'une duplication dans un stretch de C (c.2033dup) et la substitution c.528T>A qui été ratée à cause d'un défaut de couverture.

Comme pour le design 1, après les premières séries, l'analyse des profondeurs a révélé que l'amplification de certains amplicons était peu efficace, voir quasi-nulle. Les amorces de ces amplicons sont toutes mélangées dans le même tube d'amorces (Pool 2 ou P2). Un problème de fabrication ou de design est sûrement à l'origine d'une baisse des performances de ce pool par rapport aux deux autres.

Pour éviter le design d'une troisième librairie, nous avons imaginé rajouter dans les pools d'amorces 1 et 3 du design 2, des couples d'amorces du design 1. Ces couples devaient être sélectionnés pour leur efficacité (pas de défauts de couverture fréquents dans les séries du design 1) et pour leur positionnement similaire aux amplicons défaillants du design 2.

Les couples d'amorces, à la réception du design, sont individualisés dans des plaques à 396 puits à une concentration de 307µM. Ces mêmes couples se retrouvent dans les différents pools à une concentration de 400nM.

Afin d'éviter une dilution ou un déséquilibre trop important des autres couples, il est nécessaire de ne rajouter qu'un faible volume d'amorces du design 1. Pour cela, les couples d'amorces ont été préalablement dilués au 1/38.375^{ème} dans du *low TE* et la prise d'essai pour chaque couple était de 0.5µl.

Les amplicons à remplacer du design 2, et leurs remplaçants du design 1, sont notés dans le tableau 24. Comme on le remarque pour la dernière ligne, le **même** couple d'amorce a été utilisé dans les deux designs pour couvrir l'exon 6 de *SPRED1*. Dans le test, le même couple d'amorce est donc rajouté dans un autre pool : le pool 3.

| Région d'intérêt concernée | Exons à remplacer (design 2) Exons remplaçant (design 1) | Pool Design 2 | Rajout du couple d'amorces design 1 dans plex : |
|----------------------------|---|---------------|---|
| NF1 5' UTR | 1.56026 AMPL7153119858 | P2 | P1 |
| NF1 exon 4b | 1.17809 AMPL7154188810 | P2 | P3 |
| NF1 exon 5 | 1.116315 AMPL7154399192 | P2 | P1 |
| NF1 exon 9 | 1.9948 AMPL7153228397 | P2 | P3 |
| NF1 exon 19b | 1.106516 AMPL7153006304 | P2 | P1 |
| NF1 exon 28 | 1.203727 AMPL7154399203 | P2 | P1 |
| NF1 exon 49 | 1.122138 AMPL71154181793 | P2 | P3 |
| SPRED1 5' UTR | 1.51796 AMPL7154407608 | P2 | P2 |
| SPRED1 exon 6 | AMPL7154395465 AMPL7154395465 | P2 | P3 |

Tableau 24. Les différents rajouts de couples d'amorces testés dans la série 4 du design 2.

Ces essais ont été réalisés sur 2 échantillons duplicats de la série 4. Dans un duplicat des 2 échantillons, nous avons ajouté les nouveaux couples d'amorces comme figurés dans le tableau 24. Un troisième échantillon était aussi séquencé en double, mais sans modification des pools d'amorces.

II. RESULTATS

1. RESULTATS DU PREMIER DESIGN AMPLISEQ™

L'objectif principal de ces premiers essais était d'analyser la couverture et la détection des variants dans un protocole fixe (mélange des 3 PCR multiplexes après l'amplification des cibles).

Les trois premières séries ont permis de quantifier le nombre de régions insuffisamment couvertes, le taux de détection des variants (sensibilité), et de soulever quelques problèmes d'analyse. Le travail s'est porté principalement sur le gène *NF1*, pour lequel le laboratoire a le plus de recul et le plus de patients génotypés.

1.1. COMPARAISON DES TROIS PREMIERES SERIES

1.1.1. PERFORMANCES GENERALES DU SEQUENÇAGE

Les trois premières séries ont été réalisées avec le même nombre d'échantillons sur une puce 316™.

Comme on peut le voir sur la figure 75, le chargement n'est pas optimal pour la première série et s'améliore pour les séries 2 et 3, jusqu'à atteindre un très bon taux de 93%. L'expérience, et une meilleure quantification en sont les raisons principales.

Le taux d'ISP polyclonales influe sur le nombre de lectures utilisables. Il reste raisonnable et est dans la norme des passages sur le PGM™. En moyenne, environ 400 millions de bases avec un score qualité élevé sont séquencées.

La série 3 est la plus performante. La quantification par méthode qPCR peut expliquer l'amélioration du chargement de la puce. En effet, la quantification des librairies finales est bien plus précise par cette méthode, et la dernière étape de mélange équimolaire en est que plus juste. Le changement de réactif avec l'emploi du kit *Hi-Q* de l'OT2 peut aussi avoir amélioré l'amplification clonale (le taux d'ISP polyclonales est tout à fait acceptable étant donné le taux de chargement très élevé).

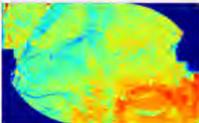
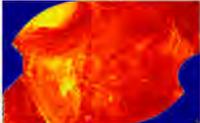
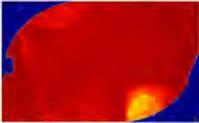
| | Série 1 | Série 2 | Série 3 | | |
|-------------------|---|---|---|---------|------|
| Patients | 16 | 16 | 16 | Moyenne | |
| Chemistry | OT2 200 kit Hi-Q 200 Sequencing kit | OT2 200 kit Hi-Q 200 Sequencing kit | Hi-Q OT2 kit Hi-Q 200 Sequencing kit | | |
| library parameter | Agencourt AMPure XP LabChip® GX- Qubit quantification | NucleoMag® NGS Clean- up and Size select LabChip® GX- Qubit quantification | NucleoMag® NGS Clean- up and Size select Ion Library Quantitation Kit on LC480 | | |
| Chip | 316 v2 | 316 v2 | 316 v2 | | |
| Total Bases | 480M | 513M | 668M | | 554M |
| |  |  |  | | |
| ISP Loading | 55% | 83% | 93% | 77% | |
| Total Reads | 2 679 776 | 3 079 721 | 4 187 090 | 3315529 | |
| Usable Reads | 78% | 59% | 72% | 70% | |
| % Polyclonal | 15% | 32% | 24% | 24% | |
| MedianLength | 195 bp | 181 bp | 167 bp | 181 bp | |
| AQ17 | 454M | 478M | 616M | 516M | |
| AQ20 | 430M | 445M | 570M | 482M | |
| Perfect | 345M | 355M | 460M | 387M | |
| Torrent Suite | 4.2.1 | 4.2.1 | 4.4.2 | | |

Figure 75. Caractéristiques et performances générales des séries : 1, 2 et 3 (1er design).

1.1.2. ANALYSE DE LA COUVERTURE

Très importante pour la validation du run, une première analyse générale de la couverture est possible avec les critères suivants : profondeur moyenne par base et pourcentage de couverture des bases d'intérêt à une profondeur donnée.

La profondeur moyenne par base est un critère global (figure 76) qui permet d'avoir une notion générale de la profondeur sur l'ensemble des régions ciblées. Elle est souvent notifiée dans les résultats d'analyse, mais ne permet pas de connaître la disparité de la couverture.

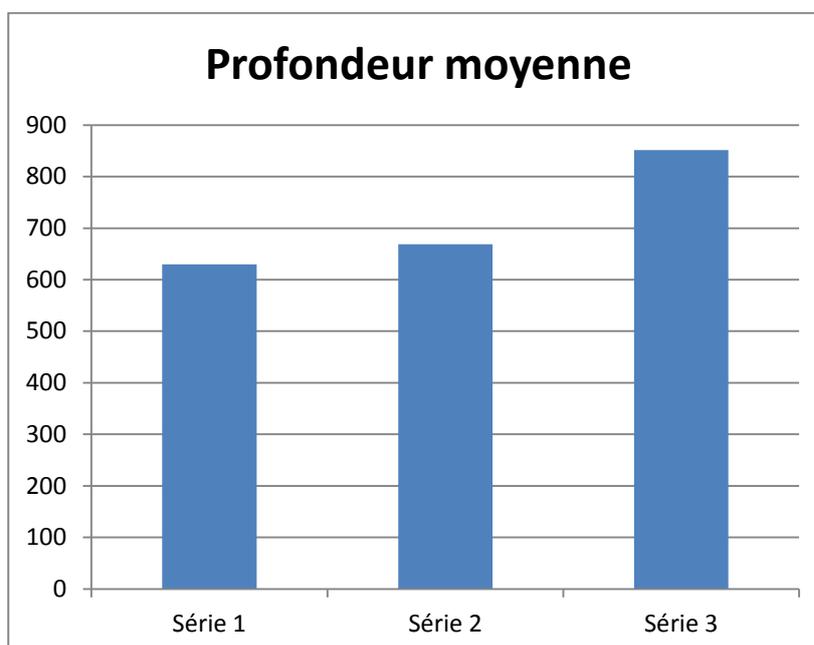


Figure 76. Profondeur moyenne des séries 1, 2 et 3 du 1er design.

La profondeur moyenne s'améliore avec les séries (plus de 800 pour la série 3). Elle est dépendante du nombre de bases totales séquencées et donc du chargement de la puce. Elle peut paraître assez élevée pour la dernière série, et laisse à penser que plus d'échantillons pourraient être analysés sur une même puce.

Un des autres critères analysé : le pourcentage de couverture des bases d'intérêt à une profondeur donnée (notre limite est de 100X), révèle une certaine hétérogénéité de la profondeur sur l'ensemble des bases du design (figure 77).

Si une profondeur de plus de 800X est une moyenne et celle de 100X parait un minimum, toutes les bases n'ont pas une profondeur suffisante, même pour la série 3. En effet, une proportion non négligeable d'environ 10% des bases d'intérêt ont une profondeur inférieure à 100X. Et pour les séries 1 et 3, environ 5% des bases sont en dessous de 20X. Au final, la série 2 est celle avec le moins de base sous la limite des 100X de profondeur.

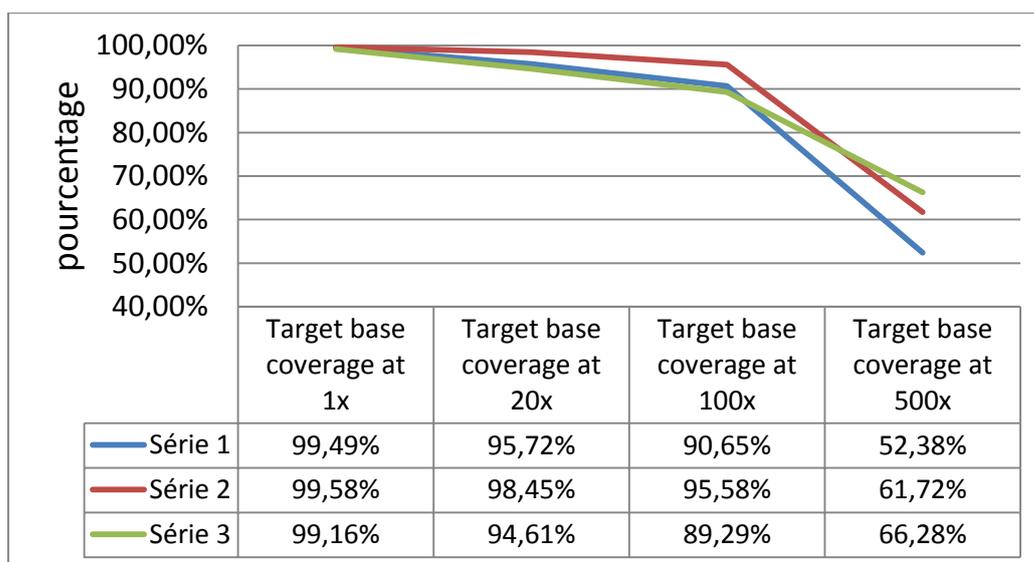


Figure 77. Taux de couverture des bases d'intérêt à une profondeur donnée (Bed CDS +/- 20).

Il apparaissait déjà évident avec cette figure que de nombreuses régions ont une profondeur de couverture insuffisante.

L'alignement est une des raisons possible d'une perte d'information globale (par l'élimination de *reads*). Le plus grand nombre possible de lectures doit s'aligner sur la région d'intérêt afin d'améliorer la couverture. Il était important pour ces premiers passages de vérifier que l'alignement sur les régions pseudogéniques du gène *NF1* soit minime et ne capte pas un nombre conséquent de lectures.

Le pourcentage de lectures hors des régions d'intérêt (ou *off-target*) augmente avec les séries, et donc avec le chargement de la puce et le nombre totale de lectures, ce qui est normal. Il est composé des séquences qui s'alignent sur les régions pseudogéniques et non pseudogéniques. Le pourcentage de lectures sur les régions non pseudogéniques et hors région d'intérêt est obtenu par soustraction (le pourcentage total des *off-target* moins celui des régions pseudogéniques obtenu par le fichier bed « pseudogènes »).

Le pourcentage de lectures qui s'alignent sur les pseudogènes est stable à environ 5% sur l'ensemble des séries. Le pourcentage de lectures alignées hors pseudogènes et région d'intérêt est quant à lui variable. Il augmente avec le nombre de lectures totales produites (plus fort à la série 3). On peut estimer qu'un plus grand nombre de lectures de mauvaises qualités est produit lorsque le chargement est presque parfait (série 3), ce qui augmente l'alignement aspécifique (figure 78).

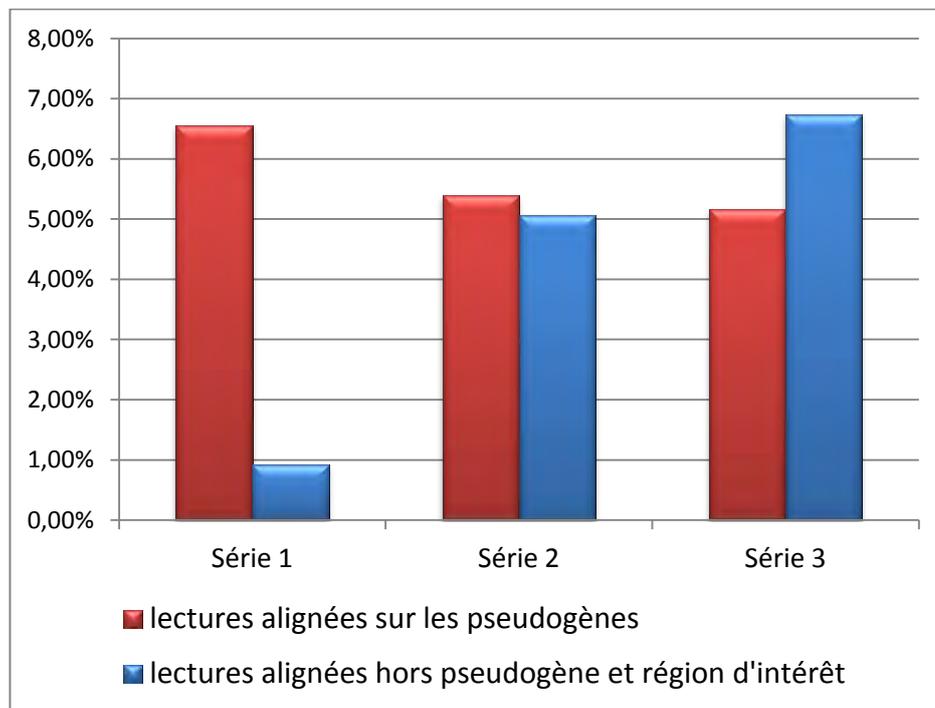


Figure 78. Pourcentage de lectures alignées hors région d'intérêt.

Ces résultats sont cohérents et acceptables pour notre méthode. La part des lectures non informatives car non alignées sur la région d'intérêt reste faible. L'alignement influence donc peu la couverture.

1.1.3. ANALYSE DE LA PROFONDEUR

Une analyse de la profondeur en détail permet de comptabiliser le nombre d'exons ayant au minimum une base couverte avec une profondeur inférieure à une limite fixée à l'avance. L'étude avec pour limite une profondeur de 100X ou 200X a été réalisée grâce au fichier *base à base* sur le gène *NF1*. Elle permet d'estimer le nombre d'exons qui seront *de facto* séquencés par méthode Sanger. L'objectif est que ce nombre soit le plus faible possible.

Pour une limite de 200X, la série 1 possède 290 régions non suffisamment couvertes au total, 182 pour la 2^{ème} et 286 pour la 3^{ème}. Les répartitions par exons sont hétérogènes et dépendantes des séries (figure 79).

Ce nombre élevé d'exons à séquencer par Sanger diminue environ de moitié avec un seuil abaissé à 100X. Il passe ainsi à 152 régions pour la série 1, à 85 pour la série 2, et à 181 pour la série 3.

Une moitié des régions initiales possèdent donc des bases avec une profondeur minimum entre 100 et 200X.

Cela indique que le nombre de patients séquencés simultanément sur une même puce (Ion 316™) ne peut être réellement augmenté, auquel cas, le nombre de régions non suffisamment couvertes deviendrait trop élevé.

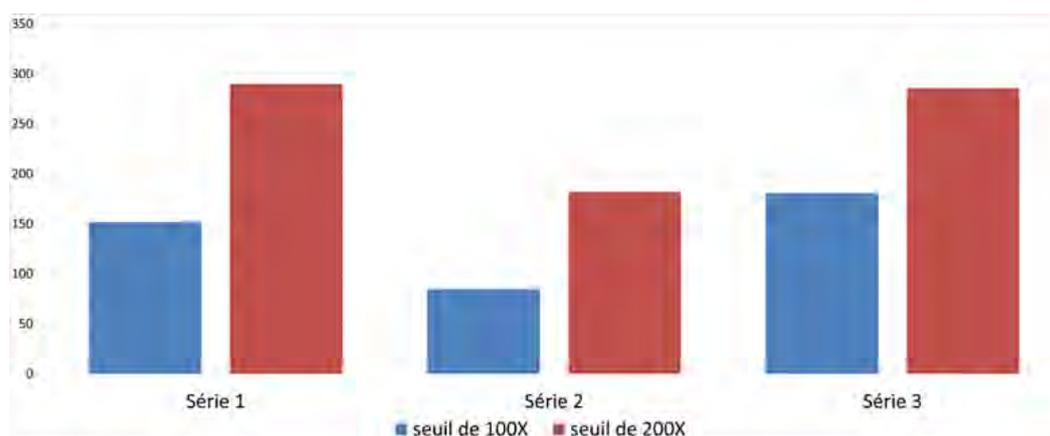


Figure 79. Nombre d'exons par série avec minimum une base à un seuil de profondeur inférieur de 100X ou 200X.

Si le seuil limite de profondeur retenu est de 100X, la moyenne de séquençage par run est de 140 régions, soit 8.75 séquences par patient (252 en moyenne et 15,75 par patient pour un seuil à 200X). En s'en tenant aux régions traduites (hors régions UTR), ce chiffre diminue à 116 amplicons de moyennes, soit 7.25 amplicons par patient. Ce chiffre est trompeur car pour les séries 1 et 3, la banque de deux patients ont été mal fabriquées. Les patients concernés ont ainsi pour presque tous les exons, des bases inférieures au seuil fixé.

Pour le seuil de 100X, en supprimant les mauvais échantillons, on arrive à une moyenne de 73 régions sous le seuil par série, soit une moyenne d'environ **5 régions** à séquencer en Sanger par échantillon.

Ce chiffre reste élevé à nos yeux et doit être pris en compte dans le coût et temps d'analyse supplémentaires.

Le graphique 81 avec un seuil à 100X de profondeur indique que pour les trois séries, 4 régions ressortent constamment et pour presque tous les patients : les 2 régions UTR en 5' et 3', et les exons 4b et 23-1 du gène *NF1*.

Ces régions sont en fait couvertes par des couples d'amorces défailants qui ne s'amplifieront dans aucune série. Ce problème provient à priori du design initial (compétition, aspécificité du couple) ou de la production des pools d'amorces par Life Technologies.

Il faut un nouveau design pour escompter couvrir suffisamment ces régions (entrepris avec le second design).

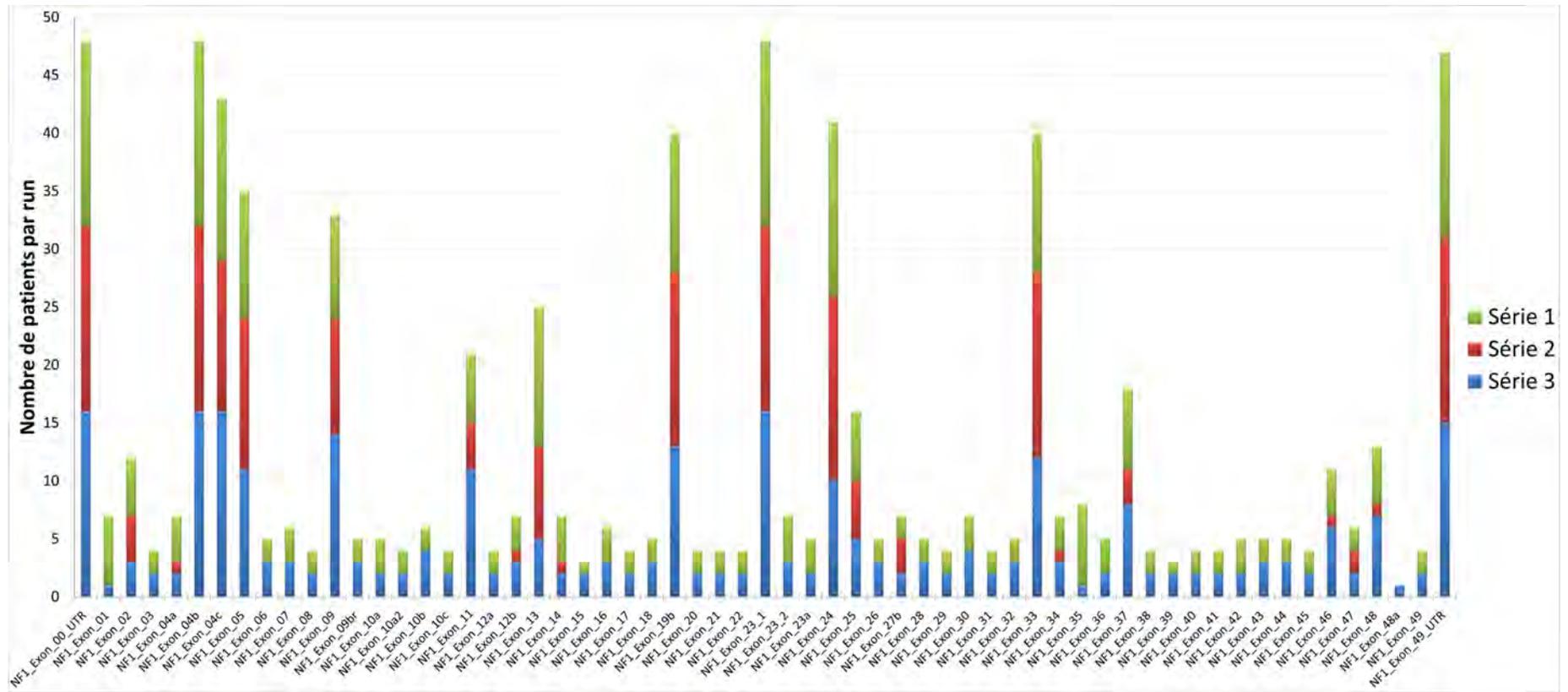


Figure 80. Compilation du nombre de patients ayant au moins 1 base <200 dans une région du gène NF1.

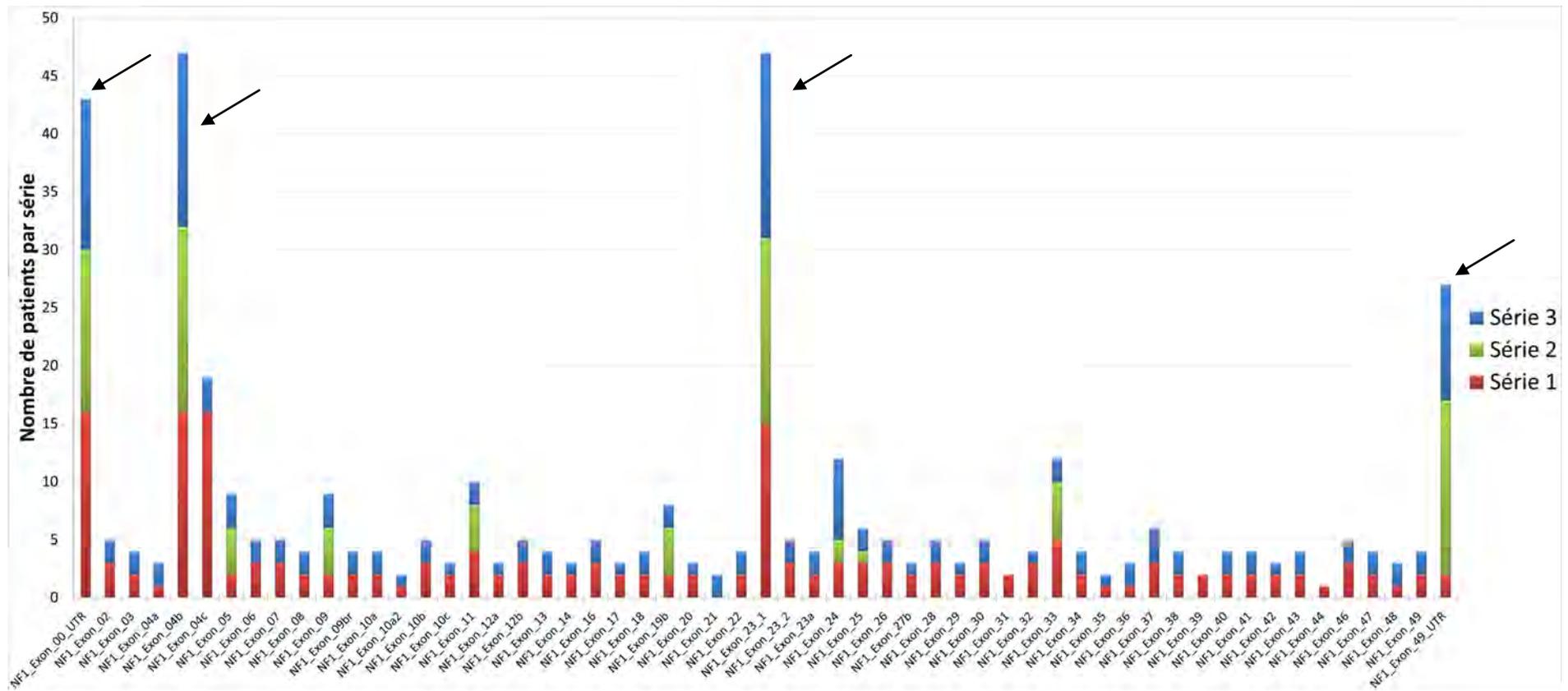


Figure 81. Compilation du nombre de patients ayant au moins 1 base <100 dans une région du gène NF1.

Les flèches indiquent les régions avec un problème de couples d'amorces défailants.

1.1.4. ANALYSE DE LA DETECTION DES VARIANTS

Au total, sur les 3 séries du premier design, 952 variants ont été trouvés par le *Variant Caller*[™]. La répartition est largement favorable aux SNV (758 variants). Les insertions sont majoritaires devant les délétions (131 et 56).

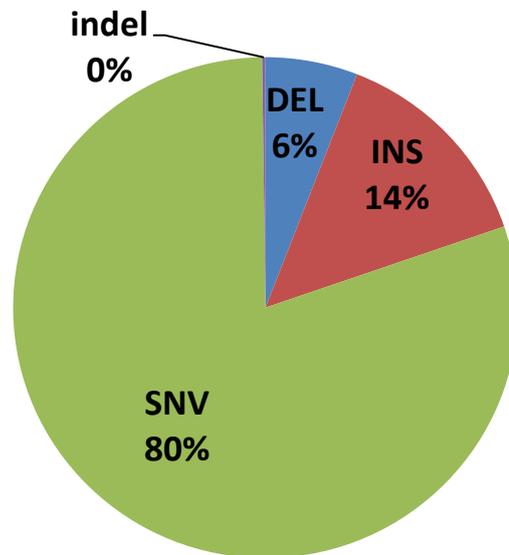


Figure 82. Distribution du type de variants retrouvés par les 3 premières séries (MNP : Multiple nucleotide polymorphism).

De nombreux variants ont été vérifiés par méthode Sanger. La moitié des variants (425) sont cependant non analysables par cette méthode (les amorces de PCR utilisées dans notre laboratoire ne couvrant pas ces régions).

Si le même variant été retrouvé chez plusieurs patients d'une même série, il n'a été vérifié que chez un patient. Le résultat retrouvé à ensuite été étendu à tous les autres patients, ou « extrapolé ». Tous les variants vus une seule fois ont été vérifiés par Sanger. Cela a permis de classer les variants : vrais positifs ou faux positifs.

Un faible nombre de variants connus (Sanger) n'ont pas été détectés par la méthode NGS, ce sont les faux négatifs NGS.

1.1.4.1. Les variants retrouvés en NGS et Sanger (vrais positifs)

Pour une grande majorité, le séquençage Sanger était en accord avec le Variant Caller (variant vrai positif ou VP).

Au total, 745 variants sont classés comme vrais positifs. Environ 40%, soit 297 variants, ont été vérifiés en Sanger. Le reste des résultats a été extrapolé.

Ils correspondent pour la plupart à des polymorphismes connus du gène *NF1*. En effet, 81.48 % des variants VP vérifiés en Sanger représentent les 10 SNP les plus fréquents (avec un rs connu et validé dans la base dbSNP, tableau 26).

Les critères : score qualité, hétérozygotie et profondeur de couverture sont donnés par le Variant Caller™ dans le fichier VCF. D'après le tableau 25, ces critères sont assez hétérogènes pour les VP.

| | Qualité | hétérozygotie | Profondeur de couverture |
|----------------|---------|---------------|--------------------------|
| min | 19,5 | 28,3 | 6 |
| max | 15002,7 | 100 | 3998 |
| moyenne | 2500,9 | 55,2 | 975,8 |

Tableau 25. Différents critères des variants VP.

Le score qualité varie d'un extrême à l'autre de 19,54 à plus de 15 000. Il est fonction de l'hétérozygotie mais surtout de la profondeur de couverture (les première et deuxième lignes représentent chacune un seul variant).

En moyenne, un variant VP est retrouvé avec un score qualité élevé et une profondeur largement suffisante de 975X.

En représentant le score qualité en fonction de la profondeur, on peut deviner 3 types différents de variants (figure 83).

Le score qualité pour les homozygotes mutés est très élevé à profondeur similaire par rapport aux hétérozygotes mutés.

Les hétérozygotes mutés mais se situant dans une région d'homopolymères ou faiblement couverte, ont un score plutôt faible, prenant en compte les bases voisines du variant.

Le modèle de régression calculé suit la distribution proposée par les hétérozygotes mutés qui sont largement majoritaires (en rouge sur la figure 83).

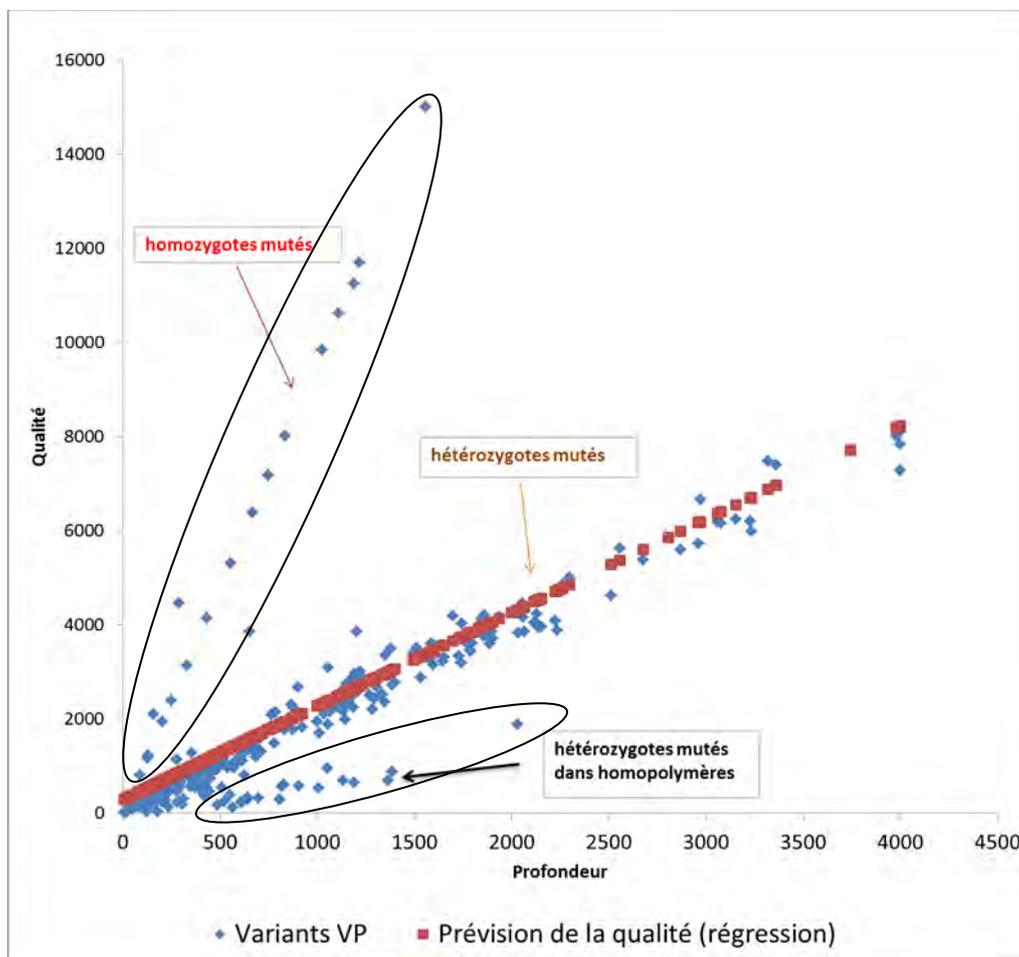


Figure 83. Représentation de la qualité en fonction de la profondeur pour les variants VP. En rouge, la prévision de la qualité, selon l'analyse de la régression

| Mutation | Total | RS connu (dbSNP) | MAF (dbSNP) | Proportion de VP vérifiés |
|----------------|-------|--------------------------|-------------|---------------------------|
| 29422382G>T | 1 | | | 0,34% |
| 29483108C>T | 2 | | | 0,67% |
| 29483145G>A | 1 | | | 0,34% |
| 29486070CA>- | 1 | | | 0,34% |
| 29486152G>A | 25 | rs2952976 | 0.451 | 8,42% |
| 29496957T>A | 1 | | | 0,34% |
| 29508508G>T | 1 | | | 0,34% |
| 29508775G>A | 24 | rs1801052 | 0.450 | 8,08% |
| 29509616T>G | 1 | | | 0,34% |
| 29528178G>T | 2 | | | 0,67% |
| 29541339A>T | 8 | | | 2,69% |
| 29541437T>C | 19 | rs2905876 | 0.478 | 6,40% |
| 29545988->T | 24 | rs67472948 | 0.444 | 8,08% |
| 29546129C>A | 1 | | | 0,34% |
| 29546175T>C | 26 | rs2905880 | 0.478 | 8,75% |
| 29548872T>C | 1 | | | 0,34% |
| 29550494TAAC>- | 2 | | | 0,67% |
| 29553478->C | 1 | | | 0,34% |
| 29553485G>A | 26 | rs2285892 | 0.471 | 8,75% |
| 29554557A>C | 1 | | | 0,34% |
| 29556837T>C | 2 | | | 0,67% |
| 29559932C>A | 29 | rs2066736 | 0.424 | 9,76% |
| 29562641C>T | 1 | | | 0,34% |
| 29562696A>G | 1 | | | 0,34% |
| 29563008C>T | 1 | | | 0,34% |
| 29585383C>T | 1 | | | 0,34% |
| 29585518A>G | 1 | | | 0,34% |
| 29586098A>T | 1 | | | 0,34% |
| 29587341G>C | 3 | | | 1,01% |
| 29587544C>G | 1 | | | 0,34% |
| 29588844->A | 1 | | | 0,34% |
| 29588889T>G | 1 | | | 0,34% |
| 29652919T>A | 1 | | | 0,34% |
| 29654553C>T | 1 | | | 0,34% |
| 29654876T>A | 25 | rs2285894 | 0.443 | 8,42% |
| 29663499C>G | 1 | | | 0,34% |
| 29663935A>T | 1 | | | 0,34% |
| 29664645T>A | 5 | | | 1,68% |
| 29665752ACTT>- | 1 | | | 0,34% |
| 29665757C>A | 1 | | | 0,34% |
| 29667610CTT>- | 1 | | | 0,34% |
| 29670065G>A | 1 | | | 0,34% |
| 29670190C>G | 17 | rs7405740 | 0.123 | 5,72% |
| 29676137G>- | 1 | | | 0,34% |
| 29676285AAAC>- | 1 | | | 0,34% |
| 29679246G>A | 27 | rs964288 | 0.442 | 9,09% |
| 29679401A>G | 1 | | | 0,34% |
| 29685905A>G | 1 | | | 0,34% |
| Total général | 297 | tous les 10 SNP en jaune | 242 | 81,48% |

Tableau 26. Variants vérifiés en Sanger (VP). En jaune les 10 SNPs les plus fréquents.

1.1.4.2. Les variants délétères (VP)

Parmi les variants VP, 28 sont conclus comme délétères et responsables de la pathologie. La plupart sont des substitutions (20/28), le reste des délétions (6/28) ou des insertions (2/28).

Dans les substitutions, 40 % des variants sont des faux-sens, 35 % des non-sens, et les 25 % restants sont des mutations modifiant l'épissage.

| Critère | | profondeur maximum | profondeur minimum | profondeur moyenne | qualité maximum | qualité minimum | moyenne qualité |
|--|----|--------------------|--------------------|--------------------|-----------------|-----------------|-----------------|
| Nombre de variants délétères trouvés: | 28 | 3360 | 17 | 1089 | 7402 | 31,62 | 2422 |
| dont substitution | 20 | 3360 | 17 | 1029,1 | 7402 | 31,62 | 2202,6 |
| dont INS | 2 | 1853 | 110 | 981,5 | 3763 | 110,81 | 1937,2 |
| dont DEL | 6 | 1845 | 456 | 1005,8 | 4125 | 854,83 | 1943 |

Tableau 27. Caractéristiques des variants délétères trouvés dans les 3 séries.

On remarque dans le tableau 27 que la profondeur minimale pour un variant délétère est de 17X, ce qui est faible (score qualité associé de 31.62).

Si notre limite de profondeur de 100X est confortable pour un rendu de résultat avec une plus grande confiance, une profondeur plus faible peut suffire pour la détection de certains variants. Il faut noter qu'une profondeur de 30X est souvent recommandée dans la littérature.

Les variants délétères sont en outre bien répartis sur l'ensemble des exons du gène *NF1*, sans former de cluster de mutation apparent. Cependant, par deux fois, le même variant est trouvé chez deux échantillons (c.1185+1G>T et c.1756_1759delACTA).

Tous les variants sont retrouvés sous forme hétérozygotes. Pas plus d'un variant par patient est estimé comme délétère, ce qui est en accord avec le caractère dominant de la NF1.

On remarque que l'hétérozygotie est très proche de 50% pour tous les variants délétères et n'est jamais inférieur à 40%.

| Exon | Mutation | Qualité | Hétérozygotie (%) | Profondeur |
|-----------|--------------------|---------|-------------------|------------|
| Exon_01 | c.55G>T | 4166,28 | 49,7 | 2058 |
| Exon_02 | c.204+1G>A | 4468,33 | 51 | 286 |
| Exon_03 | c.247_248delCA | 1520,58 | 51,2 | 689 |
| Exon_04b | c.528T>A | 31,62 | 41,2 | 17 |
| Exon_04c | c.654+1G>T | 257,79 | 40,2 | 194 |
| Exon_06 | c.821 T>G | 1214,06 | 49 | 604 |
| Exon_08 | c.1185+1G>T | 978,3 | 48,8 | 490 |
| Exon_08 | c.1185+1 G>T | 3411,6 | 52,1 | 1494 |
| Exon_10c | c.1634C>A | 3476,18 | 48,5 | 1786 |
| Exon_11 | c.1646T>C | 443,88 | 54 | 176 |
| Exon_12a | c.1756_1759delACTA | 1617,32 | 52 | 710 |
| Exon_12a | c.1756_1759delACTA | 4125,02 | 52,1 | 1845 |
| Exon_13 | c.2033dupC | 110,81 | 45,5 | 110 |
| Exon_15 | c.2342A>C | 6195,58 | 49,4 | 3058 |
| Exon_22 | c.3721C>T | 2378,89 | 49,6 | 1159 |
| Exon_22 | c.3776A>G | 3864,31 | 49,5 | 1896 |
| Exon_23-1 | c.3943C>T | 2210,67 | 45,9 | 1284 |
| Exon_24 | c.4132C>T | 7401,53 | 51,3 | 3360 |
| Exon_24 | c.4267A>G | 123,56 | 54,3 | 46 |
| Exon_25 | c.4318 A>T | 1070,67 | 47,5 | 568 |
| Exon_27a | c.4630dupA | 3763,46 | 49,5 | 1853 |
| Exon_28 | c.4854T>A | 821,75 | 51,5 | 367 |
| Exon_29 | c.5242C>T | 1331,69 | 48 | 696 |
| Exon_33 | c.6364+3A>T | 3700,45 | 50,5 | 1742 |
| Exon_37 | c.6789_6792delTTAC | 896,41 | 48,5 | 456 |
| Exon_37 | c.6792C>A | 637,58 | 50,5 | 295 |
| Exon_38 | c.6949_6951delCTT | 1703,97 | 45,4 | 1008 |
| Exon_40 | c.7128delG | 2882,74 | 48,5 | 1533 |

Tableau 28. Variants délétères retrouvés dans les 3 premières séries.

Au final, le taux de mutations est d'environ **58%** (28/48), en accord avec le taux moyen de détection de séries passées par la stratégie actuelle du laboratoire.

1.1.4.3. Les Variants non retrouvés en Sanger (faux positifs).

Un certain nombre de variants ne sont pas confirmés et sont appelés faux positifs (FP). Au total, 166 variants sont étiquetés comme tel, dont 30 après analyse par méthode Sanger.

Ils se retrouvent souvent chez plusieurs patients d'une même série, et correspondent à une erreur itérative au cours du séquençage. Certains sont spécifiques au design et au gène *NF1*, et d'autres semblent aléatoires du séquençage PGM™.

| Critère | | profondeur maximum | profondeur minimum | profondeur moyenne | qualité maximum | qualité minimum | moyenne qualité |
|-------------------------------------|----|--------------------|--------------------|--------------------|-----------------|-----------------|-----------------|
| Nombre de variants délétères | 30 | 957 | 6 | 304,7 | 4068,7 | 10,48 | 1329,3 |
| dont substitution | 10 | 372 | 6 | 86 | 450,42 | 10,48 | 96,27 |
| dont INS | 15 | 957 | 122 | 310,5 | 4068,7 | 690,18 | 1372,4 |
| dont DEL | 5 | 303 | 16 | 105 | 2449,7 | 75,07 | 748,5 |

Tableau 29. Caractéristiques des variants faux positifs trouvés dans les 3 séries et vérifiés en Sanger.

La répartition du type de variant chez les FP est différente de celle des VP. Un plus grand nombre d'insertion est retrouvé (50% des FP), et moins de substitution.

La profondeur minimum est très faible sauf pour les insertions. Pour ces dernières, la profondeur moyenne est la plus élevée, de même que la qualité. La profondeur n'est donc pas, dans ce cas, la cause des erreurs de détection (Tableau 29).

La répartition n'est plus homogène sur l'ensemble des exons du gène *NF1* mais bien plus concentrée dans quelques exons « à problème ». Ce sont en grande partie, des régions spécialement chargées en répétitions (homopolymères)

| Exon | Mutation | Qualité | hétérozygotie | Profondeur |
|----------|-------------|---------|---------------|------------|
| Exon_03 | 29486050->A | 2527 | 100 | 619 |
| Exon_03 | 29486050->A | 3165 | 100 | 677 |
| Exon_03 | 29486050->A | 1033 | 100 | 947 |
| Exon_03 | 29486050->A | 4069 | 100 | 957 |
| Exon_04c | 29508415T>- | 1161 | 100 | 149 |
| Exon_04c | 29508415T>- | 1788 | 100 | 235 |
| Exon_04c | 29508415T>- | 2450 | 100 | 303 |
| Exon_04c | 29508431T>G | 10,48 | 33,3 | 6 |
| Exon_04c | 29508446->T | 932,9 | 100 | 122 |
| Exon_04c | 29508446->T | 690,2 | 83,8 | 148 |
| Exon_04c | 29508446->T | 1348 | 100 | 167 |
| Exon_04c | 29508446->T | 1515 | 100 | 195 |
| Exon_04c | 29508446C>T | 33,08 | 85,7 | 7 |
| Exon_04c | 29508446C>T | 94,61 | 100 | 12 |
| Exon_04c | 29508446C>T | 100,3 | 100 | 12 |
| Exon_04c | 29508448T>- | 140,6 | 18,8 | 16 |
| Exon_04c | 29508448T>A | 140,6 | 81,2 | 16 |
| Exon_06 | 29509480A>- | 75,07 | 31,9 | 94 |
| Exon_06 | 29509480A>G | 10,75 | 17,4 | 23 |
| Exon_10b | 29541451->T | 2542 | 100 | 382 |
| Exon_10b | 29541451->T | 3190 | 100 | 458 |
| Exon_10b | 29541451->T | 2971 | 100 | 461 |
| Exon_10b | 29541451->T | 3097 | 100 | 476 |
| Exon_10b | 29541451->T | 1139 | 100 | 551 |
| Exon_14 | 29554327G>A | 20,96 | 26,9 | 26 |
| Exon_18 | 29557250->T | 1933 | 100 | 482 |
| Exon_18 | 29557250->T | 3151 | 100 | 843 |
| Exon_23a | 29579966A>T | 33,19 | 32,3 | 31 |
| Exon_39 | 29670052T>A | 68,35 | 20 | 355 |
| Exon_45 | 29684221C>T | 450,4 | 39 | 372 |

Tableau 30. Variants faux positifs retrouvés dans les 3 premières séries et vérifiés en Sanger.

L'hétérozygotie est aussi différente de celle des variants VP. Elle est soit faible (20-30%), soit proche des 100% (fausse homozygotie), mais semble éloignée le plus souvent de 50%.

La plateforme Ion Torrent™ est connue pour produire beaucoup de faux positifs dans les régions avec des répétitions, dû à un problème dans le traitement du signal émis lors du séquençage simultané de plusieurs bases répétées.

Les motifs voisins des variants FP sont notés dans le tableau 31 afin de trouver la cause de ces erreurs du *Variant Caller*™. On voit ainsi de nombreuses régions avec des répétitions d'un nucléotide.

Pour d'autres FP, la profondeur semble être la cause. Si elle est trop faible, elle induit en erreur le Variant Caller™ qui n'a plus une confiance suffisante dans son appel de variant (l'hétérozygotie est plus faible pour les faibles profondeurs de couverture).

| Variant (génomique) | Type | motif voisin du variant: | nombre de patients | qualité moyenne | moyenne profondeur | moyenne hétérozygotie |
|---------------------|------|--------------------------|--------------------|-----------------|--------------------|-----------------------|
| 29486050->A | INS | AAAAAAA | 4 | 2698 | 800 | 100 |
| 29508415T>- | DEL | TTTTT | 3 | 1799 | 229 | 100 |
| 29508431T>G | SNP | GTTTTT | 1 | 10,48 | 6 | 33,3 |
| 29508446->T | INS | CATTAAATTT | 4 | 1121 | 158 | 96 |
| 29508446C>T | SNP | CATTAAATTT | 3 | 76 | 10 | 95,2 |
| 29508448T>- | DEL | CATTAAATTT | 1 | 140,57 | 16 | 18,8 |
| 29508448T>A | SNP | CATTAAATTT | 1 | 140,57 | 16 | 81,2 |
| 29509480A>- | DEL | GGGATTTTGTT | 1 | 75,07 | 94 | 31,9 |
| 29509480A>G | SNP | GGGATTTTGTT | 1 | 10,75 | 23 | 17,4 |
| 29541451->T | INS | TGTTTTGTTTTG | 5 | 2587 | 465,6 | 100 |
| 29554327G>A | SNP | GCAACAGAAACA | 1 | 20,96 | 26 | 26,9 |
| 29557250->T | INS | TTTTTTT | 2 | 2542 | 662,5 | 100 |
| 29579966A>T | SNP | CCACTCCCT | 1 | 33,19 | 31 | 32,3 |
| 29670052T>A | SNP | GGCAATCCGGAA | 1 | 68,35 | 355 | 20 |
| 29684221C>T | SNP | CTTCTCACCAAA | 1 | 450,42 | 372 | 39 |

Tableau 31. Motifs proches des variants FP.

La base en rouge est celle qui porte le variant. Les variants colorés sont situés dans l'exon 4c.

L'exon 4c du gène *NF1* produit beaucoup de FP et il résume à lui seul les problèmes de détection des FP du Variant Caller™.

Il est moyennement ou mal couvert et les profondeurs sont largement insuffisantes pour beaucoup d'échantillons. Le couple d'amorces qui produit l'amplicon qui le couvre dans sa partie 5' a été difficile à mettre au point lors du design et se montre peu performant.

En effet, l'exon 4c présente beaucoup de répétitions (une seule base, en tandem, voire en triplet) qui ont posé problème pour le design des amorces. Le séquençage est ainsi perturbé et entraîne la création de nombreux faux positifs.

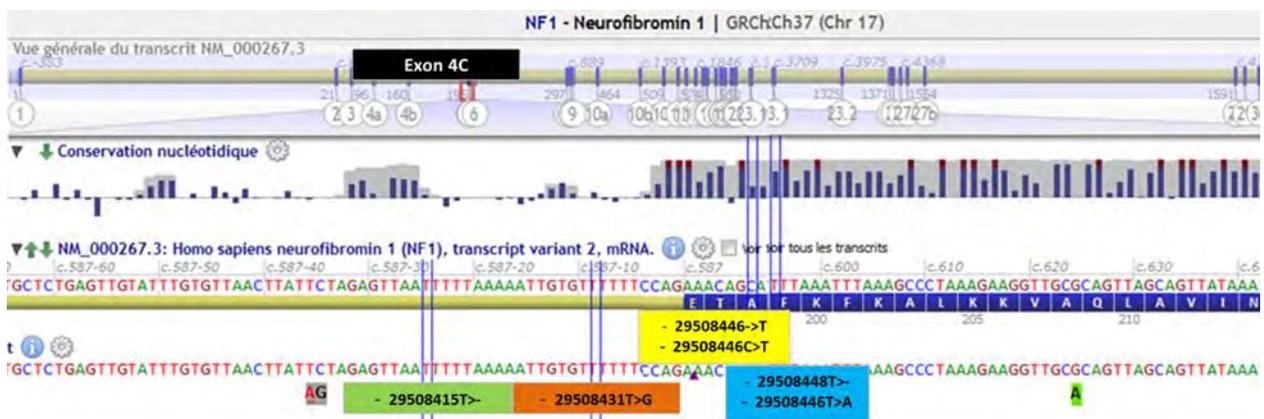


Figure 84. Les variants FP de l'exon 4c du gène *NF1*.

Pas moins de 6 variants FP différents sont situés sur cet exon. Pour la même position génomique, 2 variants différents, FP tous les deux, ont été retrouvés chez deux patients (figure 84).

La fréquence, élevée dans une même série, permet de mettre en évidence assez facilement tous ces FP. Un simple filtre sur l'occurrence permet de faire un tri efficace afin de ne garder que les variants trouvés une ou deux fois dans une série.

1.1.4.4. Les variants non retrouvés en NGS (faux négatifs)

Certains variants, non retrouvés par la méthode NGS, sont visibles par la méthode Sanger. Ils sont au nombre de 8, dont la moitié correspond à des polymorphismes.

L'autre moitié inclue des variants délétères, dont deux ont déjà été trouvés chez d'autres patients. On peut ainsi exclure un problème d'analyse intrinsèque les concernant.

La raison principale de non détection des variants chez 6 des 8 faux négatifs (FN) est un manque de profondeur (figure 85).

| Exon | classé | variant | profondeur |
|----------|------------------|---------------|------------|
| Exon_04b | Variant délétère | c.528 T>A | 0 |
| Exon_05 | rs1801052 | c.702G>A | 10 |
| Exon_05 | rs1801052 | c.702G>A | 5 |
| Exon_10b | rs2905876s | c.1393-32T>C | 8 |
| Exon_10c | SNP (pas de rs) | c.1528-29dupT | 47 |
| Exon_37 | Variant délétère | c.6792G>A | 4 |

Figure 85. Variants faux négatifs des 3 séries avec profondeurs faibles.

La couverture est un point qui peut être contrôlé par l'analyse des profondeurs et le séquençage automatique par méthode Sanger. Cependant, deux autres FN échappent à cette logique et n'auraient pas été séquencés autrement que par NGS.

Ces deux faux négatifs, de séries différentes, sont identiques et correspondent à une duplication d'une cytosine dans une répétition de 7C (c.2033dupC, haplotypes : 7C/8C chez les hétérozygotes).

Ils se situent dans l'exon 13 et sont apparemment bien couvertes (>200X sur la base portant le variant pour les deux patients).

Ce stretch de cytosines est suivi directement en 3' par un SNP fréquent (rs2285892 : c.2034G>A, MAF de 0.5). Nos deux patients sont porteurs du SNP à l'état hétérozygote.

La répétition d'un nombre important de cytosines semble être la cause de ce FN. Cependant, la mutation délétère a bien été appelée chez un troisième patient porteur du même variant.

Il faut préciser que le fait d'avoir 3 variants identiques dans 3 séries est non volontaire.

Chez ce dernier patient, le variant VP était couvert à 110X, soit deux fois moins que pour les deux FN. La différence principale réside dans le fait que ce patient bien séquencé n'est pas porteur du SNP consécutif au stretch.

Cette erreur correspond plus vraisemblablement à une erreur indépendante de la profondeur de couverture.

Elle pourrait être due à un défaut de design du panel. En aval un chevauchement entre deux amplicons se fait à proximité du variant. Ce chevauchement est assez court (17 bases), et le stretch de cytosine est ainsi situé sous l'amorce 5' de l'amplicon 2.

Une deuxième raison est le problème d'analyse du au SNP adjacent à la duplication. En effet, il est plus compliqué pour le Variant Caller™ d'appeler deux bases mutées qui sont adjacentes.



Figure 86. Représentation de l'exon 13, avec la répétition de C, le SNP adjacent et les amplicons le couvrant.

Visualisation à partir d'Alamut™ avec un des deux patients porteurs du FN.

Pour ces FN, aucune solution n'a été trouvée, ni par notre équipe, ni par les bio-informaticiens de Life Technologies. Parmi leurs propositions, une était d'activer le paramètre *allow_complex* de l'algorithme Freebayes du Variant Caller™.

Ce paramètre permet de détecter plusieurs variants dans un voisinage proche mais dans un « *block* » de taille différente.

Ce paramétrage fonctionna pour un autre panel du laboratoire, pour lequel une délétion était proche d'un SNP dans le gène *APC*.

Cette modification ne changea pas notre détection des deux FN. Et même si le nombre de faux positifs augmente drastiquement avec ce paramètre, il fut gardé pour les essais suivants.

Ces deux échecs de détection nous ont amenés à nous poser des questions sur notre design, spécialement le chevauchement entre amplicons, qui parait trop faible avec ce premier essai.

1.1.4.5. Calcul de la sensibilité et valeur prédictive positif de la méthode

Grâce à ces trois premières séries, un calcul de sensibilité est possible.

On la définit comme la proportion de mise en évidence de la mutation délétère responsable de la pathologie.

Si on considère les 2 variants faiblement couverts et les 2 duplications d'une cytosine comme délétères, alors la sensibilité est d'environ **85%** (24/28).

Si on considère que les 2 variants faiblement couverts seront séquencés en Sanger, la sensibilité est d'environ **93%** (26/28).

En prenant compte des faux positifs totaux (166/952 variants), on peut calculer la valeur prédictive positive.

Elle se définit comme la probabilité que le variant est vrai (vérifiable en Sanger) lorsqu'il est appelé par le Variant Caller. Elle est alors estimée par ces trois séries à environ **86%**.

1.2. COMPARAISON SERIE 3 ET SERIE 4

L'intérêt de la série 4 était double : continuer l'essai sur le protocole modifié pour la moitié des échantillons, et de réduire les régions insuffisamment couvertes par une puce de plus grande capacité.

La puce Ion 318™ a une capacité double par rapport à une 316™. Pour un même nombre d'échantillons analysés, la profondeur et la couverture doivent s'améliorer. Cependant, pour les couples d'amorces inefficients, le rattrapage n'est pas possible, le problème se situant en amont du séquençage (design, fabrication des amorces).

Malheureusement, l'essai du protocole modifié fut un échec avec la série 4, tout du moins pour une des trois PCR d'amplification.

Chez tous les échantillons utilisant le protocole modifié, la PCR multiplexe réalisée avec le pool d'amorces B ne fonctionna pas.

Sur la figure 87, ces échantillons sont représentés en orange. On voit clairement qu'une majorité des amplicons se sont mal amplifiés. L'homogénéité et les profondeurs sont bien supérieures pour les échantillons avec le protocole normal (en bleu).

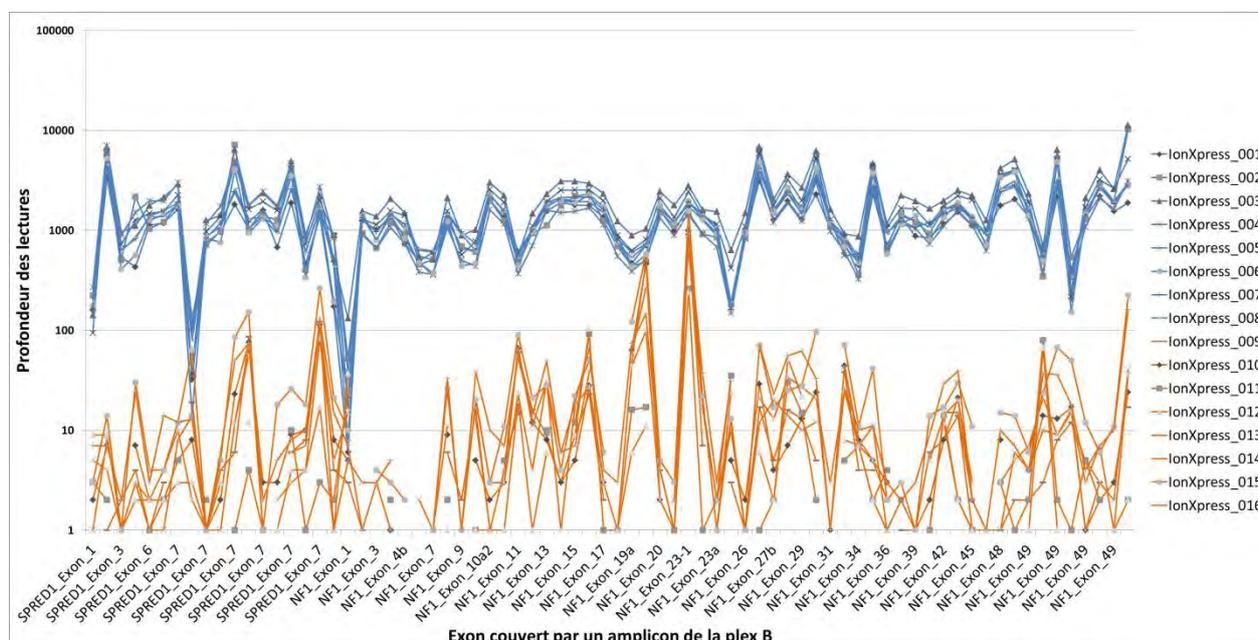


Figure 87. Profondeur des lectures de différents amplicons de la plex B. En bleu, les échantillons avec protocole normal, en orange le protocole modifié

Une étude de profondeur avec un seuil fixé à 100X sur les 8 premiers échantillons (protocole normal) a été réalisée. La moyenne par patient d'exons insuffisamment couverts est de **4,25**. Comme pour les premiers séquençages, l'exon 4b et 4c sont les plus touchés pour les régions codantes. De même, les régions UTR en 5' et 3' souffrent aussi d'un manque de couverture pour la majorité des échantillons. Pour ces échantillons, l'analyse de la profondeur est donc comparable dans l'emploi des puces 316™ et 318™.

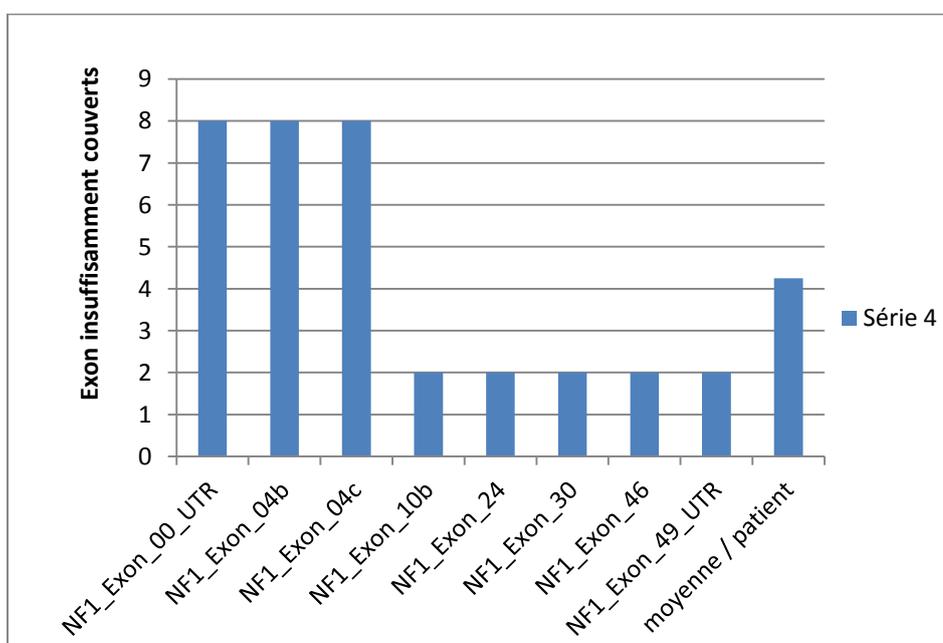


Figure 88. Résultats de l'analyse de profondeur pour les 8 premiers échantillons de la série 4.

1.3. RESULTAT DU PROTOCOLE MODIFIE

Le protocole modifié utilise moitié moins du réactif limitant *Ion AmpliSeq HiFi Master Mix*. Il permet de baisser fortement le coût de fabrication des banques, très dépendante de la fabrication de la librairie.

Nous l'avons testé dans la série 3 sur la deuxième moitié des échantillons. Les concentrations des librairies finales ont été dosées pour la première fois lors de la série 3 par Lightcycler®. Il a aussi servi pour la série 4, qui est la même série que la série 3 mais séquençée grâce à une puce 318™.

Chez 3 des 8 patients de la série 3 avec un protocole modifié, la concentration finale était faible et bien en deçà de la moyenne de la série (proche des 100 vs 1149 pM en moyenne).

De plus, lors de l'analyse des profondeurs, les résultats des échantillons 11 et 12 étaient passables et le manque de couverture flagrant. Presque toutes les régions devaient être séquencées par Sanger (55 et 49 régions).

Dans la série 4, les bibliothèques finales des échantillons 11 et 12 étaient aussi très faibles (13.6 et 42 pM).

Après ces mauvais résultats, le protocole modifié n'a plus été utilisé.

Cependant, il pourrait être testé à nouveau avec les nouveaux réactifs Hi-Q et/ou dans un protocole différent cette fois. Il consisterait à mettre la moitié de tous les réactifs de la première étape, et non juste de *Ion AmpliSeq HiFi Master Mix*. Ce protocole est validé par le laboratoire d'Anatomo-pathologie de l'hôpital Edouard Herriot pour un petit panel de mutations « *hotspots* ».

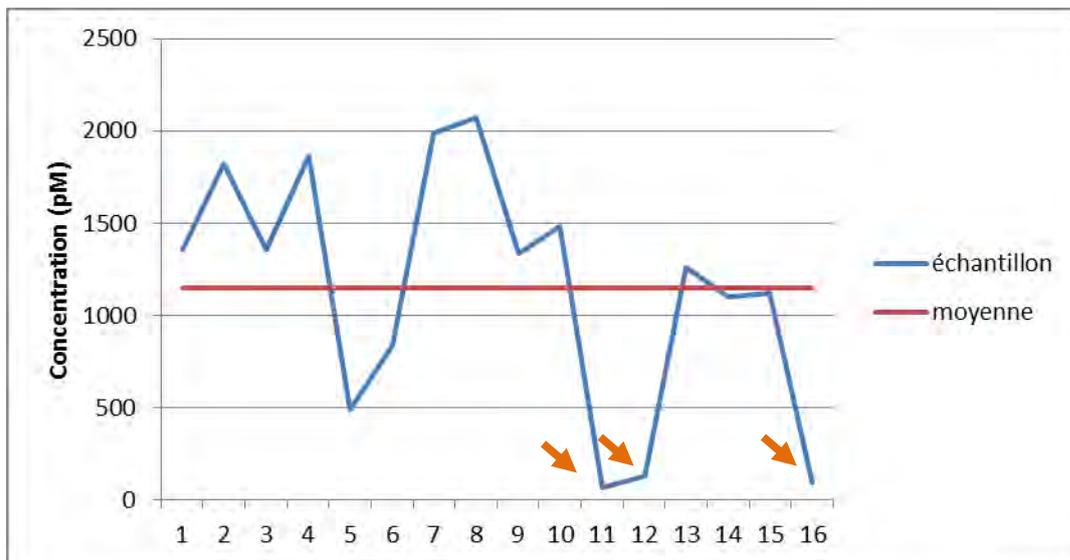


Figure 89. Concentration finale des banques (pM) de la série 3. Les flèches orange indiquent les patients avec une bibliothèque de faible concentration.

2. RESULTATS DU SECOND DESIGN AMPLISEQ™

Pour cette seconde série d'essais, l'objectif est d'analyser la couverture et la détection des variants afin de quantifier les performances du second Design sur les gènes *NF1* et *SPRED1*.

Notre attention s'est portée sur l'analyse de la couverture et de la profondeur ainsi que des variants. Pour les faux-positifs, des projections de certains critères (fréquence allélique, score qualité, profondeur) nous permettent de distinguer avec une bonne précision la nature du variant.

2.1. COMPARAISON DES SERIES

2.1.1. PERFORMANCES GENERALES DU SEQUENÇAGE

Par rapport aux séries de 16 patients sur une puce 316™, la capacité de bases séquencées par patient est double dans les séries de 24 patients sur une puce 318™.

Le protocole issu du premier design a été assimilé rapidement en pratique grâce à l'expérience acquise des premières séries. Le tableau suivant résume les bons résultats retrouvés et inscrits dans le rapport de run :

| | Série 1 | Série 2 | Série 3 | Série 4 | Moyenne |
|--------------------------------|-----------|-----------|-----------|-----------|------------------|
| Total Bases (G) | 1,4 | 1,6 | 1,4 | 1,6 | 1,5 |
| isp loading (%) | 92 | 91 | 95 | 94 | 93 |
| total reads | 7 314 189 | 7 776 040 | 6 984 915 | 8 163 621 | 7 559 691 |
| usable reads (%) | 71 | 76 | 66 | 78 | 73 |
| polyclonal (%) | 24 | 20 | 30 | 18 | 23 |
| median read length (b) | 205 | 213 | 208 | 208 | 208,5 |
| AQ17 (G) | 1,4 | 1,4 | 1,3 | 1,5 | 1,4 |
| AQ20 (G) | 1,2 | 1,3 | 1,2 | 1,4 | 1,3 |
| perfect (G) | 1,0 | 1,0 | 1,0 | 1,1 | 1,0 |
| Torrent suite (version) | 4.4.3 | 4.4.3 | 4.4.3 | 4.4.3 | 4.4.3 |

Tableau 32. Performances générales des séries : 1, 2,3 et 4 (2nd design).

La moyenne de bases séquencées sur puce 318™ (1.5 Giga) est deux fois plus grande que pour une puce 316™, en accord avec les capacités de la puce. Le chargement est très bon et le taux de remplissage de la puce est stable à plus de 90%.

Le taux de lectures utilisables est similaire aux passages du premier design (73 vs 70 % de moyenne). Le taux d'ISPs polyclonales reste lui aussi raisonnable (avec un écart important possible inter-runs). Enfin, la longueur moyenne des lectures est plus grande de 30 bases en faveur du second design, ce qui correspond parfaitement à l'écart moyen des inserts entre les deux bibliothèques.

Ces séries, de bonne qualité, ont toutes été validées techniquement.

2.1.2. ANALYSE DE LA COUVERTURE ET DES PROFONDEURS

Nous attendions des meilleurs résultats de l'analyse des profondeurs du second design par rapport au premier. Le critère important reste le nombre moyen de séquences contrôles réalisées par série, avec toujours une limite de profondeur fixée de 100X. L'analyse a été réalisée cette fois sur les deux gènes *NF1* et *SPRED1*.

Les résultats des 4 séries sont inscrits dans le tableau suivant :

| Série | Nombre d'exons (hors UTR) avec 1 base <50X | Nombre d'exons (avec UTR) avec 1 base <50X | Nombre d'exons (hors UTR) avec 1 base <100X | Nombre d'exons (avec UTR) avec 1 base <100X |
|----------------|--|--|---|---|
| série 1 | 64 | 101 | 104 | 151 |
| série 2 | 43 | 75 | 60 | 102 |
| série 3 | 46 | 82 | 78 | 117 |
| série 4 | 179 | 232 | 250 | 308 |
| Moyenne | 83 | 122,5 | 123 | 169,5 |

Tableau 33. Analyse des profondeurs par série.

La série 4, contrairement au nombre de bases séquencées, est la série avec le nombre d'exons à contrôler en Sanger le plus grand. Ce nombre est plus du triple de celui de la série 2. Il démontre un certain problème sur cette série.

En prenant en compte les régions UTR 5' et 3', la moyenne d'exons à séquencer en Sanger est d'environ 170 (limite 100X). Hors UTRs, ce qui correspond aux exons réellement séquençable par notre laboratoire, ce chiffre descend à 123. En ne prenant pas en compte la dernière série, il tombe à 80 exons à contrôler.

L'étude a aussi été réalisée avec comme limite une profondeur de 50X. Environ un tiers d'exons en moins nécessitent un séquençage Sanger avec cette limite. Les deux tiers des exons qui ont une base avec une profondeur inférieure à 100X ont en réalité une base avec une profondeur inférieure à 50X. Cette proportion élevée reflète la mauvaise efficacité de certains amplicons. Ils sont quasiment absents du séquençage de manière récurrente, et questionnent sur l'optimisation du design.

En prenant en compte les régions d'intérêts de manière individuelle, on s'aperçoit que le défaut de profondeur est, comme pour le premier design, constant pour certaines régions et donc certains amplicons.

La figure 90 démontre deux choses :

- Que les régions 5'UTR de NF1 ainsi que **l'exon 6** et la région 3'UTR de *SPRED1* sont déficients pour toutes les séries.
- Que la série 4 est bien anormalement représentée, et ses défauts de couverture sont répartis sur l'ensemble des exons des deux gènes. Un défaut dans la fabrication de la librairie est suspecté pour un ou plusieurs échantillons.

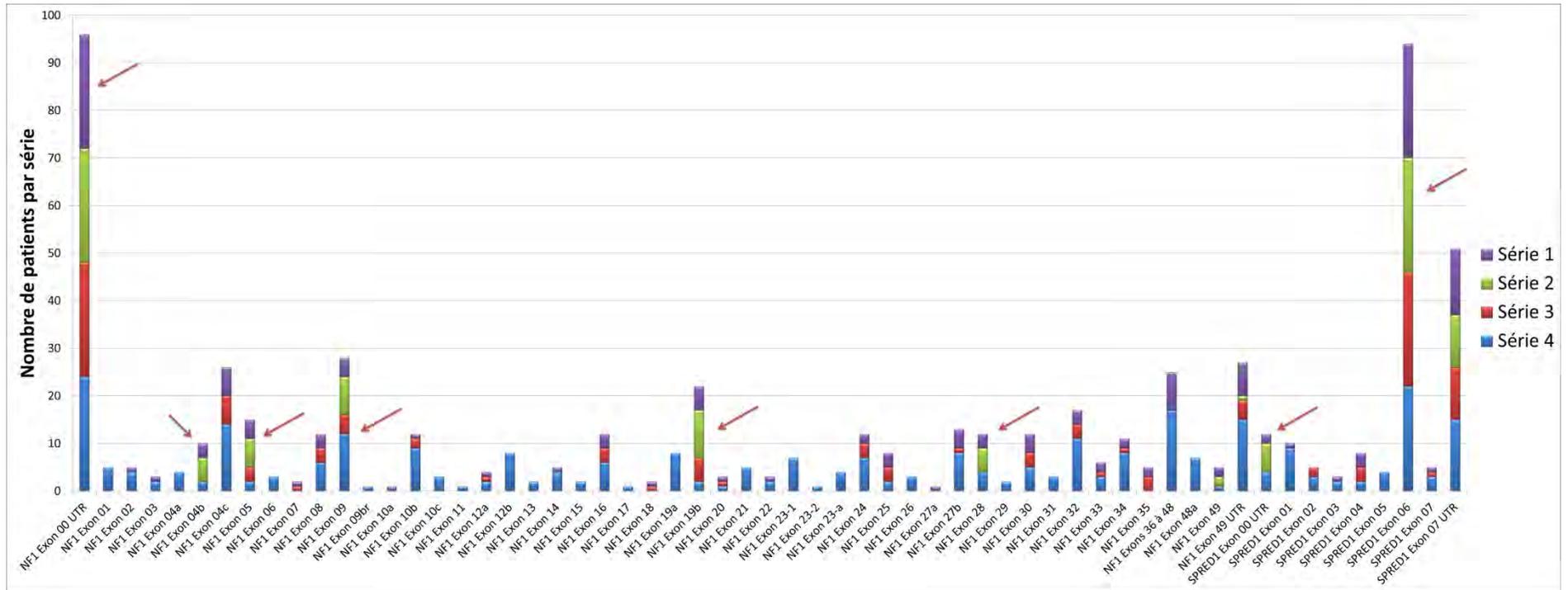


Figure 90. Nombre de patients ayant au moins 1 base <100 dans une région du gène NF1 ou SPRED1 (+/- 20 pb).

Les flèches rouges représentent les régions pour lesquelles le rajout d'un couple d'amorce du design 1 a été testé sur 2 échantillons du run 4.

Le nombre de régions touchées par un défaut de couverture et par échantillon varie de 4 pour la série 2 à presque 13 pour les échantillons de la série 4 (avec UTR, et limite de 100X).

La moyenne pour l'ensemble des séries est de 7 régions à séquencer par méthode Sanger par patient. En ne prenant pas en compte les régions UTR, cette moyenne descend à **5 exons** par patient.

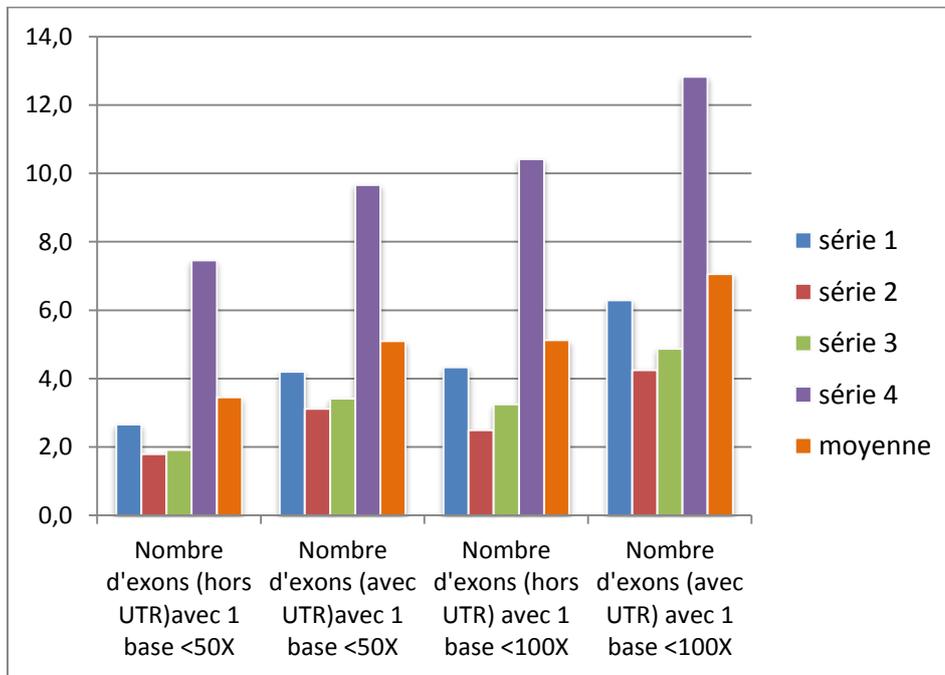


Figure 91. Nombre de régions moyenne avec une base <100X de profondeur par patient.

Le nombre élevé de défauts de couverture de la série 4 est en grande partie expliqué par un échec dans la fabrication de librairie de 2 échantillons. Ils ont chacun 45 et 49 régions d'intérêt à contrôler. Sans les prendre en compte, la moyenne de 5 exons par patient à vérifier en Sanger baisse à 4,3 exons.

2.1.3. ANALYSE DE LA DETECTION DES VARIANTS

Au total sur les 4 séries du second design, **4055** variants ont été appelés par le *Variant Caller*[™] (42,2 variants par patient en moyenne).

La répartition est largement favorable aux SNVs (3177 variants). Les insertions sont majoritaires devant les délétions (477 et 328). Plusieurs variants de type insertion-délétion (37) sont aussi retrouvés.

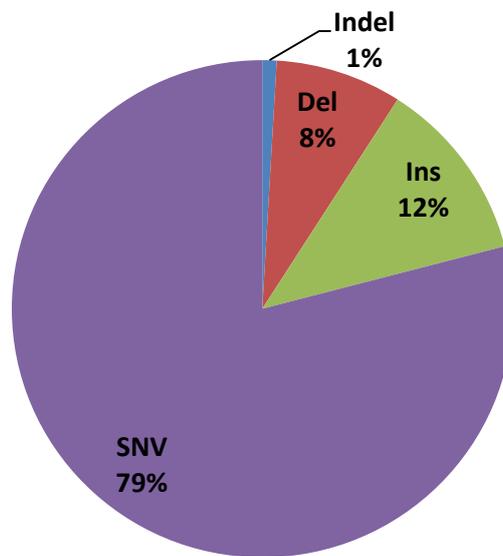


Figure 92. Distribution du type de variants retrouvés par les 4 séries du design 2.

Cette distribution est très proche de celle retrouvée pour le design 1.

Environ 52 % des variants appelés sont non couverts par nos amorces PCR et n'ont donc pas été séquencés par méthode Sanger.

Pour les deux gènes, 212 variants au total (de 146 variants différents) ont été vérifiés par séquençage Sanger sur 1895 variants atteignables par nos amorces (11% du total atteignable).

Comme pour le premier design, la conclusion pour le reste des variants a été extrapolée, et tous les types de variant ont été séquencés au moins une fois.

2.1.3.1. Les variants retrouvés en NGS et Sanger (vrais positifs)

Parmi les 1895 variants couverts par nos amorces PCR, environ 80% (1504) sont considérés comme des VP.

Le nombre de VP séquencés est de 164 variants répartis en 110 variants uniques. Le reste des VP est extrapolé à partir du séquençage Sanger d'un des représentants du variant.

| | Hétérozygotie (%) | qualité | profondeur | biais de brin |
|----------------|-------------------|---------|------------|---------------|
| minimum | 21,9 | 49,7 | 8,0 | 0,500 |
| maximum | 100,0 | 38419,8 | 4000,0 | 0,998 |
| moyenne | 78,3 | 8853,7 | 1470,7 | 0,510 |
| médiane | 100,0 | 6075,0 | 1224,0 | 0,500 |

Tableau 34. Différents critères des variants VP du design 2.

Le tableau ci-dessus reprend certains critères donnés par le *Variant Caller*[™] lors de l'appel de variants.

On peut y lire que la plus faible hétérozygotie d'un VP est de 21.9 %. La profondeur minimum de détection pour un VP est de 8X. Pour ces 2 variants détectés avec des critères plutôt défavorables, la vérification par Sanger a été effectuée.

Pour la qualité la plus faible, le variant inscrit dans le tableau 34 se situe dans un homopolymère et a une profondeur faible (VP extrapolé).

Il est à noter que la profondeur maximale est similaire à celle du design 1, avec un variant séquencé à 4000X.

2.1.3.2. Les variants délétères (VP)

Parmi les variants VP, 70 sont conclus comme délétères et responsables de la pathologie (70/73). La plupart d'entre eux sont des substitutions (41/70), le reste des délétions (18/70), des insertions (8/70) ou des insertions-délétions (3/70).

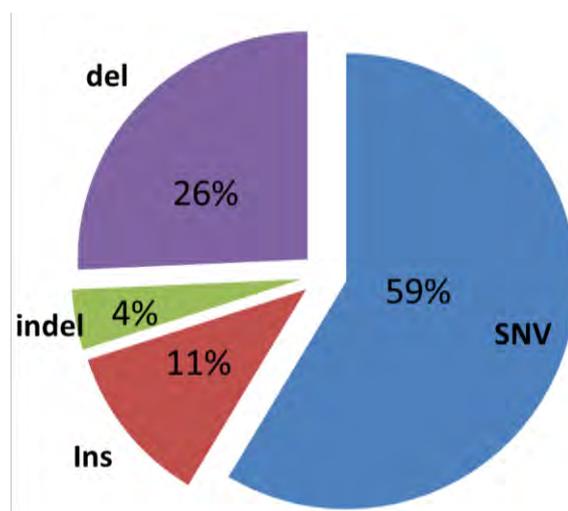


Figure 93. Distribution du type de variants délétères retrouvés par les 3 premières séries.

Dans les substitutions, 40 % des variants sont des faux-sens, 36 % des non-sens, et les 24 % restants sont des mutations entraînant une modification de l'épissage. Des taux quasi-similaires ont été trouvés pour le premier Design.

Les caractéristiques des critères qualité pour les VP sont les suivants :

| Critère | | profondeur maximum | profondeur minimum | profondeur moyenne | qualité maximum | qualité minimum | moyenne qualité |
|-----------------------------|----|--------------------|--------------------|--------------------|-----------------|-----------------|-----------------|
| Nombre de variants trouvés: | 70 | 3995 | 22 | 1739,2 | 8166,8 | 83,7 | 3493,6 |
| dont substitution | 40 | 3995 | 25 | 1644,9 | 7649,1 | 83,7 | 3149,7 |
| dont INS | 8 | 3479 | 354 | 1708,1 | 7431,4 | 338,8 | 3565,2 |
| dont DEL | 18 | 3987 | 779 | 2126,3 | 8166,8 | 1504,6 | 4550,7 |
| dont INDEL | 3 | 1581 | 22 | 541,7 | 3307,0 | 147,2 | 1200,5 |

Tableau 35. Caractéristiques des variants délétères trouvés dans les 4 séries.

Une nouvelle fois, on remarque que la profondeur minimum pour la détection des VP délétères est faible. Le biais de brin minimum est de 0.5 et maximum de 0.65. Les VP ne semblent pas avoir un biais de brin important.

Le tableau 35 compile toutes les variants délétères des 4 séries. En prenant en compte les séries 1 et 3 qui sont prospectives, ces mutations semblent bien réparties sur l'ensemble des exons du gène *NF1* et *SPRED1*.

Il faut noter que la série 3 a permis de détecter 2 variants de *SPRED1*. Le taux de mutation, pour les deux séries prospectives, est donc de 4% pour le gène *SPRED1*. Pour le gène *NF1*, le nombre de variants sur ces 2 séries est de 23 (12 + 11), soit un taux de mutation de **48%** pour les séries prospectives, en accord avec les précédentes séries de patient (58% pour le design 1). Aucun variant n'a été retrouvé à plusieurs reprises dans ces deux séries.

2.1.3.3. Les variants non retrouvés en Sanger (faux positifs)

Des 1895 variants couverts par nos amorces PCR, environ 20% (391) sont des FP ou étiquetés comme tel et répartis en 45 variants uniques.

Le nombre de FP séquencés est de 64 variants (dont 8 chez *SPRED1*). Le reste des VP est extrapolé à partir du séquençage Sanger d'un des représentants du variant.

Les caractéristiques des FP ayant été vérifiés en Sanger sont situées ci-dessous :

| Critère | | profondeur maximum | profondeur minimum | profondeur moyenne | qualité maximum | qualité minimum | moyenne qualité |
|------------------------------|----|--------------------|--------------------|--------------------|-----------------|-----------------|-----------------|
| Nombre de FP trouvés: | 64 | 2874 | 6 | 646,3 | 5409,3 | 6,0 | 2874,0 |
| dont substitution | 15 | 1205 | 10 | 273,2 | 5409,3 | 10,2 | 5409,3 |
| dont INS | 22 | 2869 | 6 | 764,9 | 4769,0 | 10,2 | 1194,0 |
| dont DEL | 17 | 2642 | 27 | 924,4 | 5166,9 | 16,7 | 624,7 |
| dont INDEL | 10 | 2874 | 9 | 472,5 | 2843,4 | 10,2 | 707,2 |

Tableau 36. Caractéristiques des variants faux positifs vérifiés en Sanger dans les 4 séries du design 2.

La répartition des types de mutation est différente de celle des VP. Le taux d'insertion, de délétion et d'insertion-délétion est supérieur aux taux rencontrés avec les VP. On peut noter que 19 FP sont notés comme homozygotes par le *Variant Caller*TM.

Comme pour le design 1, la répartition des FP n'est plus homogène sur l'ensemble des exons du gène NF1. Les exons « à problème » sont surtout les exons 4c et 34 de *NF1* ainsi que l'exon 1 de *SPRED1*.

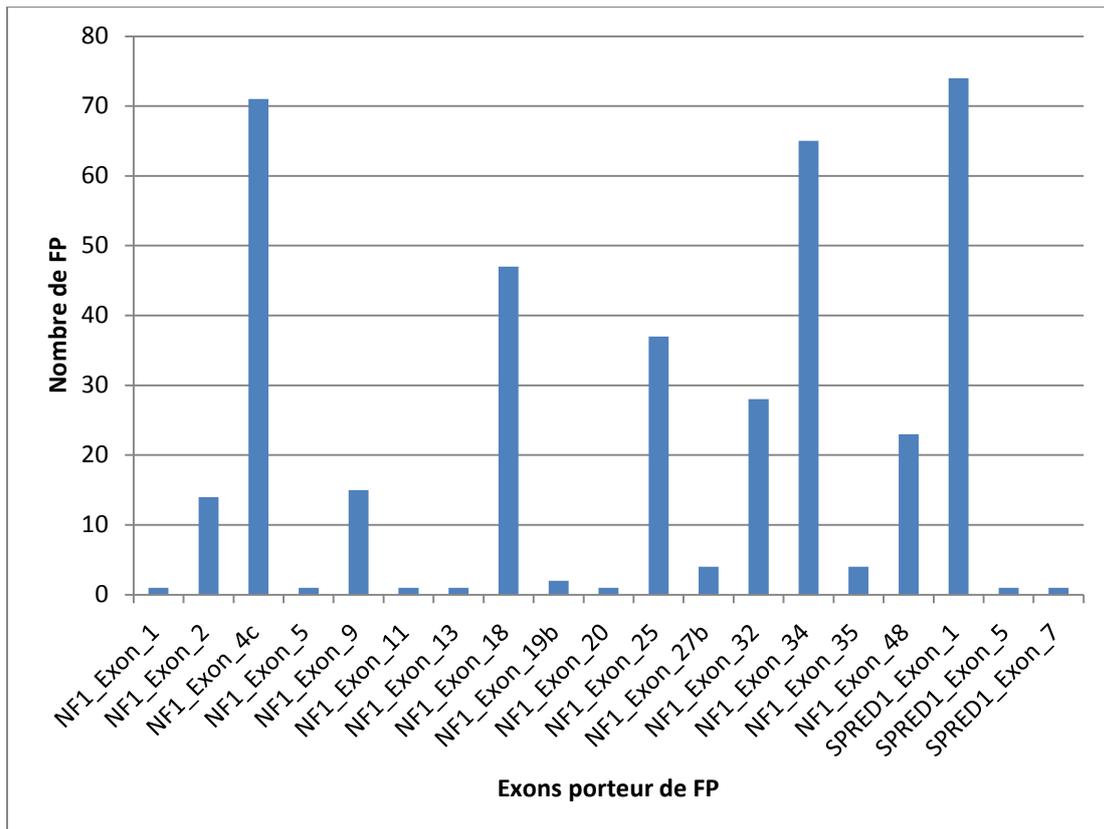


Figure 94. Nombre de FP par exons (vérifiés et extrapolés) sur l'ensemble des 4 séries.

| Série | Gène | Exon | Type | Variant | protéique | hétérozygotie | Qualité | Pronfondeur | strand bias |
|-------|------|----------|-------|----------------------|--------------------|---------------|---------|-------------|-------------|
| 2 | NF1 | Exon_1 | SNP | c.55G>T | p.Glu19* | 52.1 | 3424.49 | 1504 | 0.5185 |
| 3 | NF1 | Exon_1 | SNP | c.58C>T | p.Gln20* | 51.6 | 3697.71 | 1658 | 0.5348 |
| 1 | NF1 | Exon_2 | DEL | c.162delT | p.Ile55* | 50.3 | 2110.91 | 1000 | 0.5062 |
| 2 | NF1 | Exon_2 | SNP | c.204+1G>A | p.? | 49.1 | 1837.06 | 914 | 0.5249 |
| 4 | NF1 | Exon_3 | SNP | c.239A>G | p.Tyr80Cys | 49.9 | 5004.48 | 2413 | 0.5002 |
| 2 | NF1 | Exon_3 | DEL | c.247_248delCA | p.Gln83Valfs*23 | 49.5 | 5111.56 | 2515 | 0.5101 |
| 1 | NF1 | Exon_4b | SNP | c.574 C>T | p.Arg192* | 48.2 | 1850.24 | 958 | 0.5123 |
| 2 | NF1 | Exon_4b | SNP | c.528T>A | p.Asp176Glu | 52.7 | 1461.8 | 620 | 0.5000 |
| 3 | NF1 | Exon_4b | SNP | c.586+5G>C | p.? | 42.6 | 2210.58 | 1517 | 0.5020 |
| 4 | NF1 | Exon_4b | SNP | c.480-1G>A | p.? | 49.5 | 1740.21 | 853 | 0.5076 |
| 2 | NF1 | Exon_4c | SNP | c.654+1G>T | p.? | 35.2 | 1500.62 | 1634 | 0.5026 |
| 4 | NF1 | Exon_4c | Indel | c.587-5_587-1delinsA | p.? | 40.9 | 147.23 | 22 | 0.5881 |
| 4 | NF1 | Exon_4c | indel | c.587-5_587-1delinsA | p.? | 45.5 | 147.23 | 22 | 0.5000 |
| 1 | NF1 | Exon_5 | SNP | c.730+1G>A | p.? | 50.5 | 2622.3 | 1232 | 0.5169 |
| 3 | NF1 | Exon_6 | SNP | c.731-2A>G | p.? | 47.7 | 5287.79 | 2820 | 0.5015 |
| 4 | NF1 | Exon_6 | DEL | c.733_741del | p.Cys245_Glu247del | 50.1 | 5926.39 | 2841 | 0.5086 |
| 1 | NF1 | Exon_7 | DEL | c.983_984delGT | p.Cys328* | 51.5 | 4144.76 | 1868 | 0.5066 |
| 1 | NF1 | Exon_7 | DEL | c.1019_1020 delCT | p.Ser340Cysfs*12 | 42.8 | 2231.56 | 1514 | 0.5236 |
| 4 | NF1 | Exon_7 | Ins | c.998dup | p.Tyr333* | 50.7 | 7431.41 | 3479 | 0.5011 |
| 2 | NF1 | Exon_8 | SNP | c.1185+1G>T | p.? | 47.3 | 1844.03 | 998 | 0.5057 |
| 4 | NF1 | Exon_9 | SNP | c.1198C>T | p.Gln400* | 48.9 | 1538.35 | 771 | 0.5067 |
| 4 | NF1 | Exon_9 | DEL | c.1202del | p.Asn401Ilefs*11 | 51.4 | 2969.32 | 1363 | 0.5180 |
| 4 | NF1 | Exon_9 | DEL | c.1202del | p.Asn401Ilefs*11 | 51.6 | 4903.13 | 2218 | 0.5214 |
| 3 | NF1 | Exon_10a | SNP | c.1381C>T | p.Arg461* | 53.0 | 5484.97 | 2321 | 0.5052 |
| 4 | NF1 | Exon_10a | SNP | c.1318C>T | p.Arg440* | 50.0 | 6209.58 | 3002 | 0.5063 |
| 4 | NF1 | Exon_10b | SNP | c.1446A>G | p.Tyr489Cys | 54.9 | 1018.73 | 395 | 0.5133 |
| 1 | NF1 | Exon_10c | DEL | c.1541_1542delAG | p.Gln514Argfs*43 | 53.4 | 4664.36 | 1945 | 0.5147 |
| 2 | NF1 | Exon_11 | SNP | c.1646 T>C | p.Leu549Pro | 48.5 | 2206.36 | 1133 | 0.5129 |
| 2 | NF1 | Exon_12a | DEL | c.1756_1759del | p.Thr586Valfs*18 | 49.8 | 8128.4 | 3952 | 0.5158 |

| | | | | | | | | | |
|---|-----|-----------|-------|--------------------|----------------------|-------|---------|------|--------|
| 2 | NF1 | Exon_13 | INS | c.2033dup | p.Ile679Aspfs*21 | 100.0 | 7202.89 | 1258 | 0.5000 |
| 2 | NF1 | Exon_13 | INS | c.2033dup | p.Ile679Aspfs*21 | 43.2 | 1114.1 | 2428 | 0.5163 |
| 2 | NF1 | Exon_15 | SNP | c.2342A>C | p.His781Pro | 47.5 | 6562.7 | 3534 | 0.5003 |
| 4 | NF1 | Exon_16 | DEL | c.2674del | p.Ser892Alafs*10 | 49.7 | 8166.81 | 3987 | 0.5196 |
| 4 | NF1 | Exon_18 | INS | c.3088_3089dup | p.Phe1031Hisfs*6 | 44.5 | 1631 | 775 | 0.5062 |
| 4 | NF1 | Exon_19a | INS | c.3197+9dup | p.? | 54.5 | 338.752 | 354 | 0.5353 |
| 3 | NF1 | Exon_19b | SNP | c.3216C>G | p.Ser1072Arg | 60.0 | 83.7023 | 25 | 0.6469 |
| 4 | NF1 | Exon_19b | SNP | c.3277G>A | p.Val1093Met | 49.3 | 1043.94 | 511 | 0.5121 |
| 1 | NF1 | Exon_21 | SNP | c.3578T>G | p.Phe1193Cys | 51.0 | 2207.04 | 1009 | 0.5062 |
| 1 | NF1 | Exon_21 | INS | c.3606_3607dupTG | p.Asp1203Valfs*13 | 54.0 | 2878.54 | 1167 | 0.5070 |
| 4 | NF1 | Exon_21 | SNP | c.3703C>T | p.Gln1235* | 48.2 | 7649.06 | 3995 | 0.5060 |
| 2 | NF1 | Exon_22 | SNP | c.3721C>T | p.Arg1241* | 49.0 | 3436.04 | 1721 | 0.5067 |
| 2 | NF1 | Exon_22 | SNP | c.3776A>G | p.Asn1259Ser | 33.5 | 2102.24 | 2602 | 0.5125 |
| 2 | NF1 | Exon_23.1 | SNP | c.3943C>T | p.Gln1315* | 47.3 | 4541.09 | 2466 | 0.5317 |
| 4 | NF1 | Exon_23.2 | indel | c.4017_4023delinsT | p.Leu1340_Gln1341del | 50.1 | 3306.99 | 1581 | 0.5161 |
| 2 | NF1 | Exon_24 | SNP | c.4267A>G | p.Lys1423Glu | 48.6 | 1508.71 | 768 | 0.5264 |
| 1 | NF1 | Exon_25 | SNP | c.4277 A>G | p.Gln1426Arg | 47.1 | 1282.96 | 699 | 0.5281 |
| 2 | NF1 | Exon_25 | SNP | c.4318 A>T | p.Met1440Leu | 51.9 | 1942.31 | 861 | 0.5057 |
| 2 | NF1 | Exon_27a | INS | c.4630dupA | p.Thr1544As | 49.8 | 6290.52 | 3065 | 0.5004 |
| 3 | NF1 | Exon_27a | SNP | c.4537C>T | p.Arg1513* | 51.5 | 4424.88 | 1989 | 0.5033 |
| 4 | NF1 | Exon_27b | SNP | c.4772+1G>A | p.? | 50.0 | 3233.14 | 1552 | 0.5073 |
| 3 | NF1 | Exon_28 | SNP | c.4865T>A | p.Val1622Glu | 49.7 | 4215.82 | 2052 | 0.5078 |
| 4 | NF1 | Exon_28 | DEL | c.4914_4917del | p.Lys1640Glyfs*36 | 49.4 | 3086.61 | 1528 | 0.5154 |
| 1 | NF1 | Exon_29 | SNP | c.5483A>T | p.Asp1828Val | 47.4 | 4481.64 | 2422 | 0.5263 |
| 4 | NF1 | Exon_30 | SNP | c.5749+2T>G | p.? | 47.5 | 6174.56 | 3323 | 0.5053 |
| 4 | NF1 | Exon_31 | SNP | c.5839C>T | p.Arg1947* | 49.2 | 3624.22 | 1799 | 0.5501 |
| 2 | NF1 | Exon_33 | SNP | c.6364+3A>T | p.? | 52.8 | 1865.88 | 790 | 0.5235 |
| 4 | NF1 | Exon_34 | SNP | c.6407C>G | p.Ser2136* | 47.5 | 3624.17 | 1949 | 0.5230 |
| 2 | NF1 | Exon_37 | DEL | c.6789_6792delTTAC | p.Tyr2264Thrfs*5 | 49.7 | 6116.45 | 2983 | 0.5214 |
| 2 | NF1 | Exon_38 | DEL | c.6949_6951delCTT | p.Leu2317del | 43.2 | 4504.97 | 3013 | 0.5632 |

| | | | | | | | | | |
|---|--------|---------|----------|---------------------|-------------------|------|---------|------|--------|
| 1 | NF1 | Exon_40 | DEL | c.7182_7183delAC | p.Leu2395Thrfs*5 | 48.1 | 4327.4 | 2262 | 0.5013 |
| 2 | NF1 | Exon_40 | DEL | c.7128delG | p.Tyr2377Thrfs*20 | 49.3 | 2046.67 | 1147 | 0.5111 |
| 3 | NF1 | Exon_40 | INS | c.7127-11_7127-8dup | p.? | 42.8 | 1634.52 | 1139 | 0.5339 |
| 3 | NF1 | Exon_41 | SNP | c.7285C>T | p.Arg2429* | 40.1 | 1639.69 | 1296 | 0.5113 |
| 1 | NF1 | Exon_42 | SNP | c.7486 C>T | p.Arg2496* | 48.7 | 5259.15 | 2679 | 0.5028 |
| 3 | NF1 | Exon_43 | SNP | c.7583C>G | p.Ser2528* | 48.9 | 3906.62 | 1971 | 0.5186 |
| 3 | NF1 | Exon_46 | DEL | c.7917del | p.Leu2639Phefs*19 | 47.4 | 1504.61 | 818 | 0.5372 |
| 4 | NF1 | Exon_5 | Faux Neg | c.670delG | p.Val224* | | | | |
| 4 | NF1 | Exon_5 | Faux Neg | c.670delG | p.Val224* | | | | |
| 2 | NF1 | Exon_13 | Faux Neg | c.2033dup | p.Ile679Aspfs*21 | | | | |
| 2 | SPRED1 | Exon_2 | SNP | 131T>A | p.Val44Asp | 50.4 | 4093.39 | 1936 | 0.5138 |
| 3 | SPRED1 | Exon_2 | SNP | c.88G>A | p.Gly30Arg | 45.9 | 1293.64 | 748 | 0.5140 |
| 3 | SPRED1 | Exon_3 | DEL | c.305del | p.Thr102Serfs*19 | 47.7 | 4742.64 | 2540 | 0.5021 |
| 2 | SPRED1 | Exon_7 | DEL | 1048_1060del13 | p.Gly350Metfs*52 | 37.9 | 7226.66 | 779 | 0.5000 |

Tableau 37. Variants délétères du design 2.

En rouge les patients FN NGS, en fond bleu les mutations retrouvées dans *SPRED1*. En orange les duplicats de 3 patients de la série 4 (avec test d'amorces).

2.1.3.4. Les variants non retrouvés en NGS (faux négatifs)

Au total, 3 variants délétères (2 distincts) ne sont pas retrouvés par le séquençage NGS. Ils s'expliquent par deux causes différentes :

- **Défaut de couverture :**

La délétion c.670delG, située dans l'exon 5 du gène *NF1* est portée par deux échantillons en duplicat dans la série 4.

L'exon 5 fait partie des exons les plus souvent insuffisamment couverts (figure 90). Pour nos échantillons, une couverture trop faible est à l'origine du défaut dans l'appel de variant.

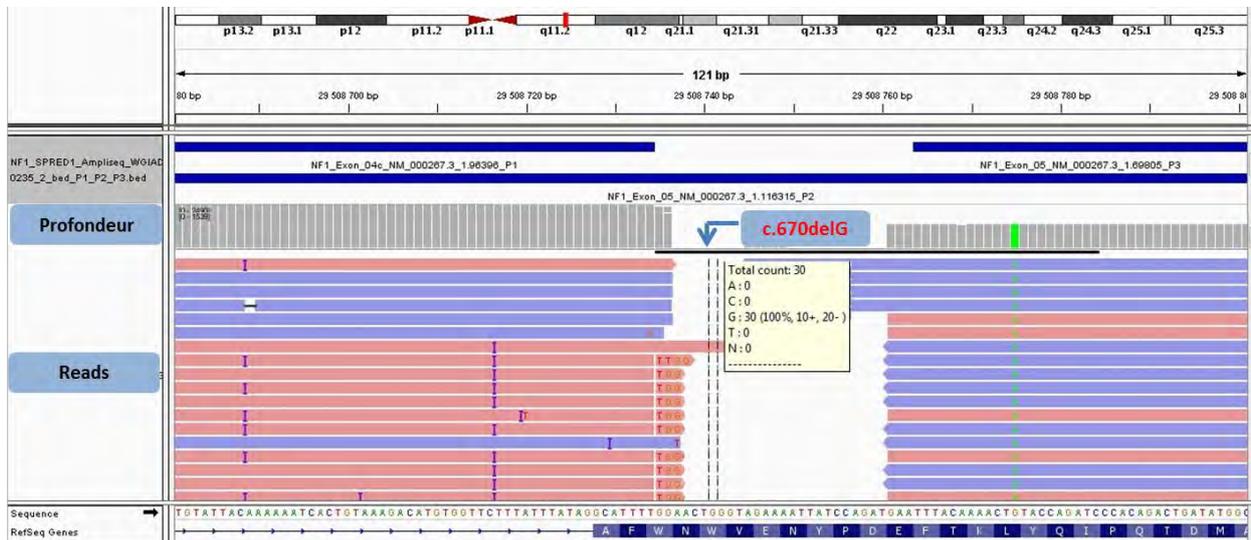


Figure 95. Faux positif c.670delG du gène *NF1* dû à un défaut de couverture (IGV).

Une profondeur de 30X est insuffisante dans ce cas. Par ailleurs on remarque qu'aucune des 30 lectures n'est porteuse de la délétion d'après la lecture sur Alamut™.

Pour les deux échantillons, le défaut de couverture entraîne systématiquement la vérification par le séquençage Sanger, le variant sera donc détecté.

- **Défaut dans l'appel de variant du *Variant Caller*[™] :**

Les trois variants identiques c.2033dupC de l'exon 13 du gène *NF1*, qui posaient déjà problème avec le premier design, ont été séquencés à nouveau avec le second design. Les 3 échantillons font partis de la série 2. Ils portent le code-barre IonXpress 3, 4 et 5.

Pour rappel, la duplication d'une cytosine se fait à la suite d'un stretch de 7 cytosines. Un SNP adjacent, situé en 3' du stretch, semble provoquer un défaut d'appel de variant par le plug in *Variant Caller*[™].

Les échantillons 3 et 4 sont hétérozygotes pour le SNP (c.2034G>A) tandis que l'échantillon 5 est lui homozygote sauvage à cette position.

Pour le premier design, les échantillons 3 et 4 étaient retrouvés FN, le 5^{ème} étant appelé correctement.

Pour le design 2, seuls les échantillons 4 et 5 ont leurs duplications renseignées dans le fichier résultat VCF de variants :

- Comme pour le design 1, l'échantillon 5 non porteur du SNP est appelé correctement, avec une bonne fréquence allélique du variant.
- Pour l'échantillon 4, le SNP est appelé de manière conventionnelle (fréquence allélique de 47%). Par contre, pour la duplication, l'échantillon est évalué comme porteur du variant de manière homozygote (fréquence allélique de 100%). La duplication passe donc les filtres qualité du *Variant Caller*[™], mais sa fréquence allélique est sur-estimée.

L'appel de variant pour cet échantillon s'est fait sur une profondeur de 1258X d'après le fichier vcf produit par le *Variant Caller*[™]. Or, d'après le fichier *base à base*, la profondeur à cette position était le double. Il semble que les lectures non porteuses de la duplication aient été écartées lors de l'appel de variant, expliquant ainsi la fausse homozygotie de l'échantillon pour ce variant.

Pour l'échantillon 3 (similaire à l'échantillon 4), le SNP est aussi appelé (fréquence allélique de 38%) mais pas la duplication. Cependant, aucune distinction visible directement permet de comprendre la différence de traitement dans l'appel du variant entre les échantillons 3 et 4.

A contrario du design 1, la duplication c.2033dupC est située à cheval entre deux amplicons. Le design ou les nouveaux réactifs de séquençage ont peut-être permis d'augmenter la sensibilité de détection (2/3 patients, même si un variant a été retrouvé faussement homozygote, la correction sera faite lors de la vérification par séquençage Sanger).

Le stretch de 7C semble tout de même poser un problème d'analyse au Variant Caller™. A nouveau, la question de l'implication du SNP se pose, puisque pour le patient 5, aucun soucis d'appel de variant n'est à déplorer sur deux séries et deux designs différents.

Nous avons demandé aux équipes de Life Technologies de réaliser des essais pour affiner les paramètres qui permettront d'atteindre une sensibilité améliorée. Un changement de paramétrage du Variant Caller™ ou une mise à jour améliorant ses performances pourront permettre d'améliorer ces résultats. Le passage à la version 5.0.2 du Variant Caller™ permet de retrouver la duplication hétérozygote pour l'échantillon 4.

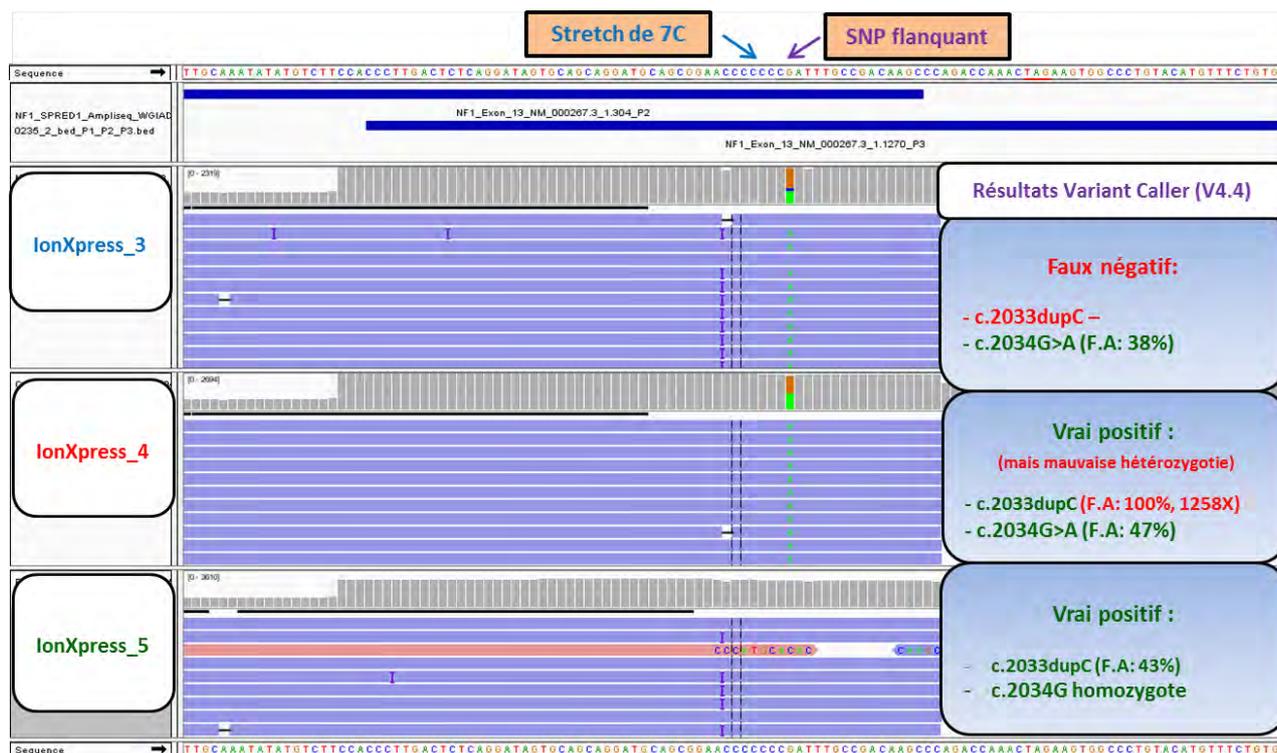


Figure 96. Visualisation de 3 échantillons porteurs d'une duplication dans l'exon 13 du gène NF1.

Une analyse toute particulière devra donc être apportée à l'exon 13 du gène *NF1* afin d'éviter les FN. Une lecture des variants rejetés par le Variant Caller ou une lecture visuelle sur Alamut™ peut permettre de limiter ces échec et évitent de séquencer automatiquement l'exon 13 de chaque échantillon.

2.1.3.5. Calcul de la sensibilité et valeur prédictive positif de la méthode.

Le calcul de la sensibilité et de la VPP est réalisé de la même manière que pour le design 1 (à partir des variants délétères).

Seules 3 variants, tous dans le gène *NF1*, ne sont pas retrouvés sur un total de 73.

La VPP se calcule en prenant en compte tous les VP et FP (vérifiés et extrapolés).

Les résultats sont inscrits dans le tableau ci-dessous :

| Mutations délétères | Théoriques (n) | Mutations retrouvées | Sensibilité de détection (%) | VPP (%) |
|---------------------|----------------|----------------------|------------------------------|-------------|
| Total | 73 | 70 | 95,9 | 82,2 |
| gène <i>NF1</i> | 69 | 66 | 95,7 | 77,8 |
| gène <i>SPRED1</i> | 4 | 4 | 100,0 | 90,5 |

Tableau 38. Calcul de sensibilité et VPP pour le design 2.

La sensibilité globale aux deux gènes *NF1* et *SPRED1* est proche des **96%**, et donc comparable à la sensibilité des méthodes de référence. Aucun variant non détecté

La VPP est plus faible pour le gène *NF1* que pour *SPRED1* à cause d'un nombre conséquent de faux positifs dus aux **répétitions** contenues dans le gène.

2.1.3.6. Arguments pour l'analyse biologique des variants

La VPP pour le gène *NF1* semble moyenne, et 1 variant sur 4 serait en réalité un FP. Afin de discriminer au mieux les FP des VP, nous avons déterminé quelques caractéristiques propres aux FP, qui sont à prendre en compte de la part du biologiste.

Avec l'aide de notre pipeline d'analyse bio-informatique (Annotations complètes d'Alamut™, base dynamique des FP avec FileMaker™ etc.) et la prise en compte de certains critères, le

biologiste peut considérer facilement et rapidement certains variants comme de probables FP. La vérification, pour certains d'entre eux, serait alors superflu et le nombre de séquençage Sanger de contrôle serait ainsi réduit.

Hétérozygotie et profondeur :

La figure ci-dessous révèle que l'hétérozygotie est un caractère discriminant entre les VP et les FP. On peut voir que très peu de FP ont une hétérozygotie supérieure à 40%.

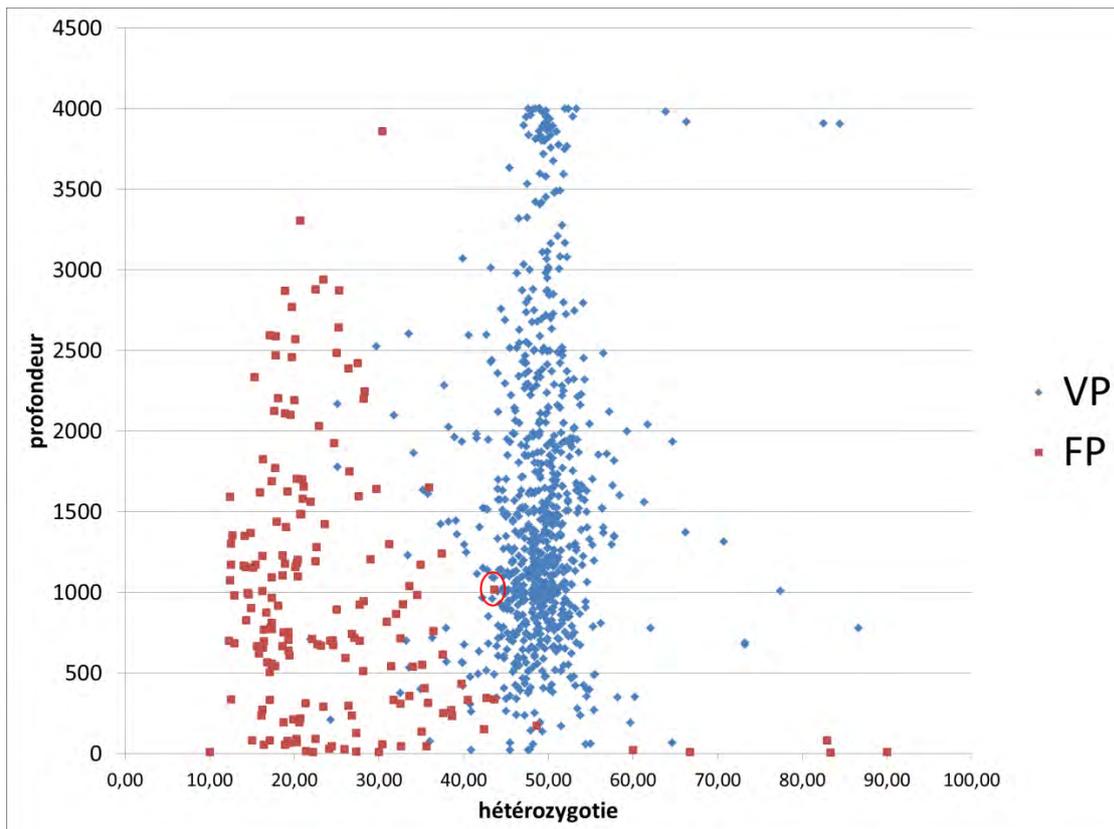


Figure 97. Profondeur des variants en fonction de l'hétérozygotie (VP et FP appelés comme hétérozygotes).

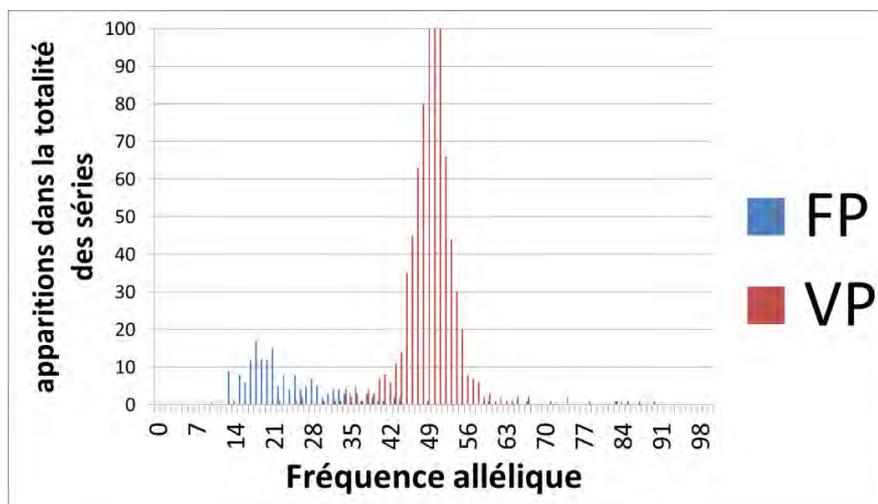


Figure 98. Distribution de la fréquence allélique des variants VP et FP dans les 4 séries.

Cependant, ici ne sont pris en compte que les variants appelés par le Variant Caller™ comme hétérozygotes. Un large nombre de FP possède une fréquence allélique proche des 100% (souvent avec biais de brin).

Dans le cadran nord-est délimité par les flèches vertes horizontales (profondeur à 1000X) et verticales (hétérozygotie à 40%), aucun FP n'est représenté.

Le seul variant FP situé à la limite de cette zone, avec une hétérozygotie à 43% et une profondeur à 1018X, a été vérifié en Sanger et confirmé comme FP.

A la vue de cette figure, le biologiste peut avoir plus confiance dans les variants appelés avec une forte hétérozygotie (entre 40 et 60%) et une profondeur suffisante de 1000X.

Biais de brin :

Le biais de brin est un des paramètres incontournables des filtres de l'étape d'appel de variant. Un variant retrouvé toujours sur le même brin indique fortement un FP et une erreur intrinsèque à la fabrication des librairies ou du séquençage.

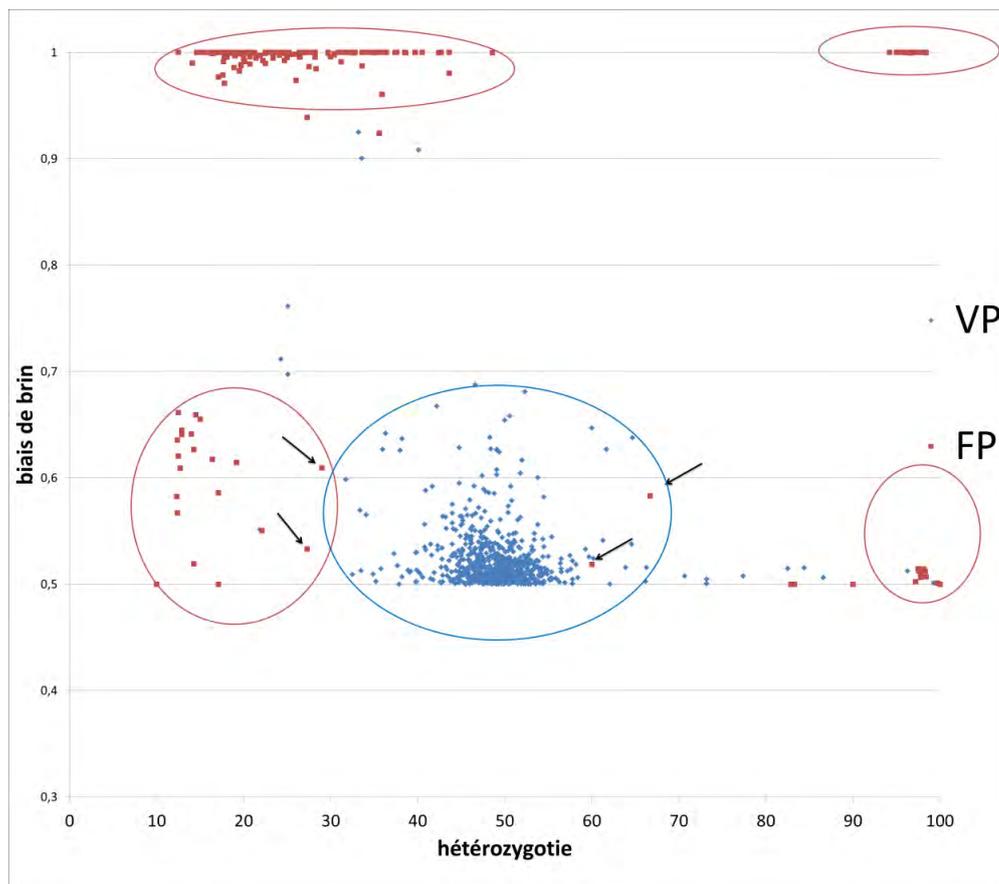


Figure 99. Biais de brin en fonction de l'hétérozygotie chez tous les variants VP et FP (flèches noirs : les FP vérifiés en Sanger).

Comme on peut le voir sur la figure 99, 4 « zones » de FP peuvent être distinguées (cercles rouges) selon les critères:

- Fort biais de brin, faible hétérozygotie (en haut à gauche de la figure).
- Fort biais de brin, grande hétérozygotie (100 % aux deux, en haut à droite).
- Biais de brin normal, faible hétérozygotie (comme vu plus haut, au milieu à gauche).
- Biais de brin normal, grande hétérozygotie (100%, au milieu à droite).

Au final peu de variants VP ont un biais de brin important. S'il peut exister, il n'est presque jamais de 100%.

Cette notion est à prendre en compte par le biologiste lors de l'analyse de variants, d'autant plus que pour notre méthode, aucun filtre ne joue sur le biais de brin (fixé à 1).

2.2. RESULTATS DES DUPLICATS DE LA SERIE 4.

Lors de la série 4, un mélange d'amorces provenant de la série 1 a été rajouté aux échantillons IonXpress_008 et 013. Les amplicons produits par ces couples d'amorces devaient éviter les défauts de couverture des régions fréquemment insuffisamment couvertes.

Sur la figure 100 sont représentés les nombres de lectures correspondantes aux couples d'amorces rajoutés pour les 2 échantillons tests et la moyenne des 22 autres échantillons (vert).

Pour le design 2, les amplicons les moins efficaces étaient ceux de l'exon 5' UTR de *NF1* et l'exon 06 de *SPRED1*. On peut voir (flèches oranges) que la moyenne pour ces deux régions des échantillons normaux est presque nulle.

Notre test obtient de très bons résultats pour ces deux régions à problème, avec une couverture jusqu'à 3500 lectures pour la région UTR de *NF1*.

Pour les amplicons avec une efficacité moindre mais pas nulle, la hausse est plus mesurée (exceptée pour *NF1* exon 09 et 49). De plus, l'ajout d'amorces supplémentaires pour certains n'entraîne aucun changement et le nombre de lectures est même similaire (*NF1* exon 04b et 05).

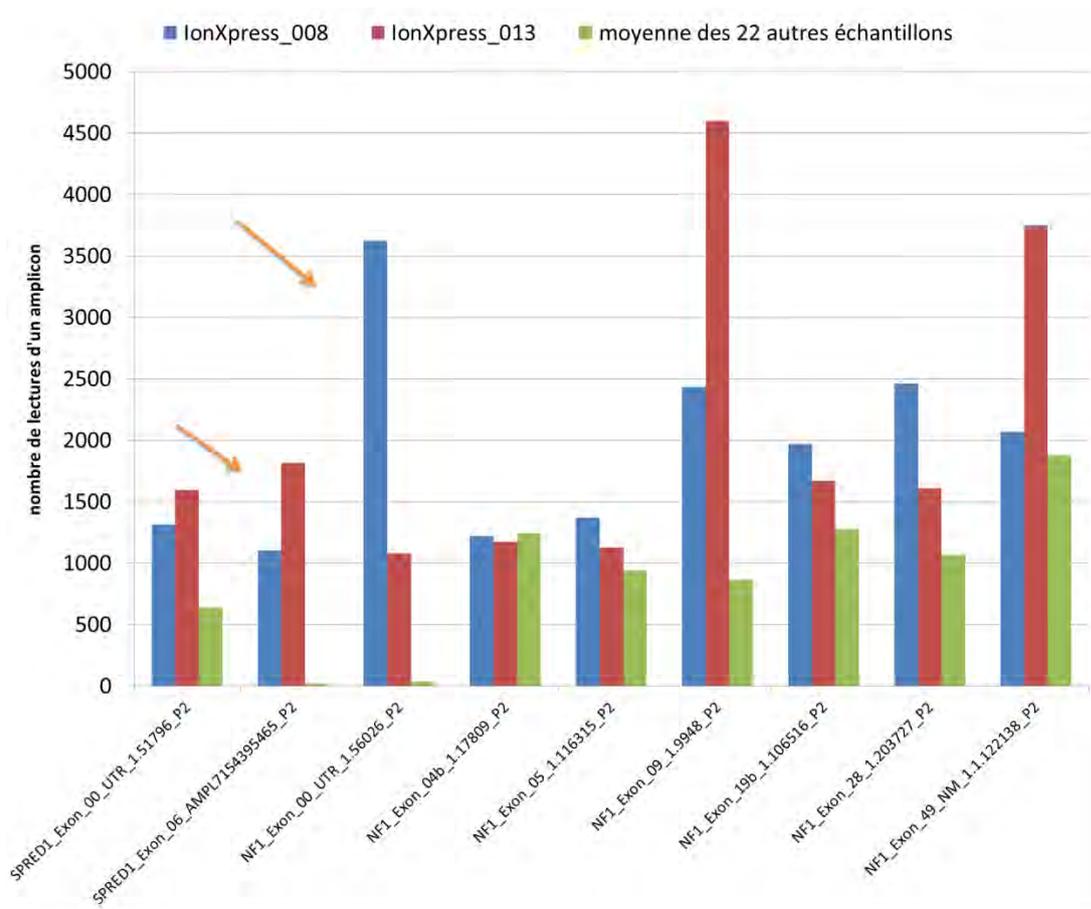


Figure 100. Comparaison du nombre de reads selon l'ajout ou non d'amorces du design 1.

Pour rappel, le couple d'amorce de l'exon 06 de SPRED1 est le même dans le design 1 et 2. Cependant, l'ajout d'amorces dans un autre pool du design 2 change radicalement le nombre de lectures produites. On peut donc en conclure en un défaut de production ou de design des amorces.

Cette solution n'est pas exempte de défauts (étape manuelle supplémentaire avec risque d'erreur, augmentation du volume du pool et dilution des autres couples), mais semble efficace. Elle peut être poursuivie, uniquement pour les exons qui demandent un séquençage systématique Sanger (spécialement l'exon 6 de SPRED1), ou pour tous les autres amplicons peu efficaces.

Un retour vers Life Technologies et la commande d'un nouveau design avec un changement dans l'attribution des pools d'amorces est cependant conseillé.

III. DISCUSSION

1.1.1. LA TECHNOLOGIE AMPLISEQ™.

1.1.1.1. Raison du choix

Le choix de la librairie s'est porté sur la méthode Ampliseq™ (Life Technologies) pour plusieurs raisons.

La technologie Ion Torrent™ fut la première plateforme de séquençage haut-débit mise en place au CHU de Lyon. Le PGM™ était un automate de paillese prometteur, permettant de s'initier rapidement aux techniques de NGS. Plusieurs équipes étaient à l'initiative de ce projet, ce qui permettait un échange continu afin de résoudre les premiers problèmes rencontrés et optimiser nos capacités d'analyse. La méthode Ampliseq™ est une méthode liée au PGM™ et correspond à la création de librairie par PCR multiplex.

Les méthodes par capture ne nous semblaient pas le choix le plus judicieux, en particulier à cause de la présence de nombreuses régions pseudogéniques de *NF1* dans le génome humain. Les profondeurs de couverture semblaient insuffisantes avec les premiers essais d'autres laboratoires. Par ailleurs, une équipe utilisant la technologie SureSelect™ (Agilent Technologies) publia l'analyse haut-débit d'un panel de 109 gènes dont *NF1*¹⁸⁵. La couverture était hétérogène (un exon 1 jamais amplifié) et la profondeur moyenne assez faible à 192X. La méthode par capture était à l'époque plus chère et fastidieuse à mettre en place. Elle montrait aussi des limitations dans la couverture et une sensibilité inférieure à notre méthode d'analyse par méthodes dHPLC-Sanger (environ 92 %).

Life Technologies fournissait des solutions adaptées à toutes les étapes, du design jusqu'au séquençage. Cela a permis un échange facile et productif avec leurs équipes techniques devant chaque problème rencontré.

La plateforme Ampliseq™ Designer, utilisée pour les deux designs, est un outil intuitif et automatisé. Il permet la création rapide d'un design de librairies en quelques clics seulement. L'algorithme employé est efficace pour la plupart des séquences ciblées. Le design peut être délivré rapidement et le prix d'une librairie est compétitif (3.000 réactions dans une commande).

1.1.1.2. Limitations de la stratégie Ampliseq®

Cependant, l'Ampliseq™ Designer montre ses limites pour des régions à faible spécificité de séquence (motifs répétés). L'Ampliseq™ Designer propose un paramétrage limité de la librairie et ne permet pas d'optimisation fine. Après l'échec d'un design ou si besoin d'un paramétrage manuel et spécifique, Life Technologies propose l'aide de ses techniciens à travers un service appelé « White Glove ». Après une évaluation de notre premier design, des améliorations semblaient nécessaires et nous avons fait appel à ce service.

D'une part, les chevauchements entre deux amplicons étaient beaucoup trop courts pour permettre un appel des variants optimal. Durant les premiers essais, des variants de type insertion ou délétion et se trouvant à l'intersection de deux amplicons n'étaient pas détectés par le Variant Caller™. Un plus grand chevauchement été nécessaire pour augmenter la sensibilité de la méthode. Ce travail a pu être réalisé avec le deuxième design qui présente un chevauchement moyen de 50 bases (cahier des charges).

D'autre part, l'efficacité d'amplification des cibles était très hétérogène. Certaines réactions PCR ne marchaient dans aucune série. Pour d'autres, la PCR semble peu robuste et entraîne des défauts de couverture récurrents sans qu'une raison particulière ne se démarque réellement.

Récemment pour l'analyse *BRCA* faite dans notre laboratoire, nous avons rencontré une baisse drastique de la qualité de librairies réalisées par capture à partir d'un ADN ayant un rapport de DO 260/230 trop faible. Même si la technologie Ion Torrent™ semble moins dépendante de la qualité de l'ADN séquencé qu'une méthode de capture, la pureté initiale de l'ADN joue probablement un rôle. D'autres critères, comme la performance d'un thermocycleur, la consommation des réactifs et la qualité de la purification (automate ou manuelle) sont à prendre en considération devant l'échec de fabrication d'une plex, ou d'un échantillon complet.

L'algorithme de l'Ampliseq™ Designer semble donc perfectible. Sa capacité à former les pools d'amorces n'est pas toujours optimale comme le montre notre essai d'ajout d'amorces du design 1 dans les pools du design 2. L'équipe d'Anatomo-pathologie de l'hôpital Edouard Herriot utilise également un design Ampliseq™ personnalisé avec des échecs de couples d'amorces similaires aux nôtres. L'analyse *in silico* ne prédisait aucune amplification pour le couple d'amorce déficient mais proposé par Life Technologies. Un essai de réaction PCR monoplex avec le couple d'amorce en question a montré que l'amplification ne se faisait dans aucune condition PCR.

L'essai de plusieurs designs pour un même panel de gènes permet ainsi d'optimiser la couverture et évite le séquençage Sanger des régions mal couvertes. Le prix de ces designs est

raisonnable. C'est une part faible dans le coût unitaire d'un séquençage, mais ces mises au point sont longues et fastidieuses.

1.1.2. LA FABRICATION DE LIBRAIRIE.

Plusieurs essais ont été nécessaires afin d'optimiser la fabrication de librairie avec notre design comprenant 3 pools d'amorces, et donc 3 réactions d'amplification de cibles par patient.

Notre choix de mettre en commun ces 3 réactions juste après la première étape semble judicieux. Il nous permet de ne traiter qu'un seul tube au lieu de 3 pour toutes les étapes suivantes. Les défauts de couverture ne semblent pas venir de cette modification de protocole. Pour le design 2, les amplicons les moins efficaces proviennent tous de la plex B. Cette plex B (comme la A) a une concentration plus grande que celle de la plex C à la sortie de l'étape d'amplification des cibles. Cependant, le volume de prise d'essai est le même pour les 3 plexs, et la plex C ne semble pas souffrir de défauts récurrents d'amplification. Le protocole mise en place est donc valide. Il ne diminue pourtant pas les coûts de fabrication, qui restent dépendant du réactif limitant de l'étape d'amplification des cibles.

Les essais de protocole modifié avec moitié moins du réactif limitant « 5x Ion AmpliseqHiFi Master Mix » ont été des échecs sur les quelques échantillons testés. La librairie finale était faible et des défauts de couverture plus nombreux étaient présents. De nouveaux tests pourraient être effectués, avec ce même protocole ou un autre légèrement différent. En effet, l'équipe d'Anatomo-pathologie utilise un protocole avec des volumes divisés par deux pour tous les réactifs, ne modifiant ainsi aucunement la concentration propre à l'AmpliseqHiFi Master Mix.

Cette économie de réactif est vivement conseillée, puisque le prix de la fabrication de librairie compte fortement dans le coût total du séquençage d'un échantillon. Pour nos designs à 3 pools d'amorces, une série à 24 échantillons consomme environ les 2/3 de l'AmpliseqHiFi Master Mix disponible dans un kit.

Si les premiers essais ont permis de mettre l'accent sur l'importance de certains réactifs dans la réussite de la technique, comme l'alcool absolu utilisé, d'autres tests doivent être effectués afin de valider l'utilisation des thermocycleurs du service de génétique moléculaire. Actuellement, nous utilisons le thermocycleur Veriti localisé dans le laboratoire d'hémostase. Nous avons validé dans notre service le contrôle de nos thermocycleurs avec des températures inhabituelles pour une réaction PCR classique, mais proches de celles utilisées par exemple lors de l'étape de ligation. Le prochain test devra se faire sur une machine qui possède

la capacité de moduler la température du couvercle afin d'obtenir une homogénéisation de la température, essentielle pour l'étape de ligation des codes-barres. L'utilisation des thermocycleurs de notre service rendra la tâche technique plus facile pour le technicien.

La fabrication de librairie nécessite plusieurs étapes de manipulation classique pour un technicien averti. Elle reste manuelle pour beaucoup d'étapes et demande tout de même une certaine exigence. D'une part, le nombre de patients séquencés dans une même série ne peut pas être trop grand, au risque de devoir manipuler un grand nombre de tubes et de provoquer des erreurs techniques. D'autre part, l'étape de ligation des barres-codes est cruciale et demande la plus grande vigilance au technicien. Plusieurs inversions de barre-code sont à déplorer, pour plusieurs techniciens de notre laboratoire. Facilement discernable lors de séries rétrospectives, une mauvaise distribution de ces courtes séquences est difficilement analysable après séquençage. La nécessité d'une vérification par Sanger des variants délétères permet cependant de repérer ces erreurs techniques et de procéder à un nouveau séquençage pour les échantillons concernés, ou un éventuel traitement informatique des fichiers d'analyse.

Toutefois, l'erreur demeure invisible dans le cas d'un défaut de couverture d'un échantillon mal barre-codé et pourrait provoquer la non détection de variants délétères (contrôle Sanger réalisé sur le mauvais échantillon). Cette étape doit donc être réalisée avec rigueur. Des contrôles doivent être mis en place afin d'éviter les erreurs. Une solution pourrait être de demander à un second technicien de valider la distribution des codes-barres.

Une automatisation des étapes de fabrication de la librairie est aussi bienvenue. Mise au point sur un automate Hamilton, les étapes de purification permettent un gain en temps technique et une homogénéité des performances. La dernière version utilisée (V3) emploie un « shaker », et semble proposer de bonnes performances. L'Ion Chef™ System permettrait d'automatiser toutes les étapes de fabrication de librairie. Son emploi pourrait être utile mais deux conditions limitent son utilisation. D'une part il est situé sur le site du Centre de Biologie et Pathologie Est (C.B.P.E), d'autre part, il n'est pas adapté pour la création d'une librairie à 3 pools.

Les étapes d'amplification clonale et de séquençage se font sur le site du C.B.P.E. Une technicienne dévouée à plusieurs équipes prenait jusqu'à présent en charge ces étapes, qui nécessitent une bonne connaissance pratique et une solide expérience (chargement de la puce par exemple). Le déploiement en routine de notre panel et d'autres méthodes nécessite la formation de techniciens qualifiés, et une organisation bien définie du matériel et réactifs entre les différentes équipes qui utilisent la plateforme.

1.1.3. L'ANALYSE DES DONNEES

Le volume de données généré par les techniques de séquençage haut-débit a rendu nécessaire pour le biologiste une connaissance fine des étapes de bio-informatique.

Ce nouveau besoin rend obligatoire l'utilisation de matériels et logiciels performants et sensibles. Le traitement des données doit se faire avec des ordinateurs dédiés aux performances élevées. Les logiciels utilisés doivent être adaptés aux besoins analytiques mais aussi aux capacités matérielles et intellectuelles.

La plateforme Ion Browser™ est efficace. Elle possède un visuel agréable et permet le lancement de plusieurs scripts de façon automatisée. Cette unique plateforme permet de gérer les listes de patients, de lancer des analyses et de récupérer les résultats. Enfin, elle permet aussi la gestion de l'export et du stockage des données produites.

Elle n'est cependant pas exempte de défauts. De nombreuses mises à jour, nécessaires mais fréquentes, sont déployées par Life Technologies™ pour l'Ion Browser. Sans chamboulements particuliers, ces mises à jour posent tout de même un problème dans la gestion que l'on a des données. Un changement de version de l'Ion Browser peut ainsi s'accompagner d'un changement de format des listes de patients (changement du nombre de colonnes par exemple), d'un changement de versions des scripts (qui demanderait une nouvelle validation à chaque changement), ou encore d'une remise à zéro du temps imparti maximum à une analyse Variant Caller™ (24h par défaut, que nous modifions à 72h pour le fichier VCF *base à base*). Ces changements de version, outre certains nouveaux bugs minimes apportés, demande une vigilance particulière aux personnes travaillant avec l'Ion Browser™. Par les changements apportés, les changements et corrections de nos procédures sont chronophages. Pour ces raisons, nous n'avons pas utilisé la solution de Life Technologies pour la lecture et le rendu des variants (phase d'annotation et de validation des variants) : l'Ion Reporter™.

A la place, une solution interne a été pensée et optimisée par nos soins. Même si elle n'est pas parfaite, elle permet une gestion efficace, rapide, et peu coûteuse avec des logiciels déjà présents au laboratoire (Excel®, Alamut™, FileMaker™). Sa mise au point demanda une grande charge de travail, mais elle permet d'avoir une solution complète dans le traitement et la validation des variants et la recherche de CNV.

La recherche de grands réarrangements peut se faire par méthode « *reads count* » grâce à une macro Excel. Proche du principe de la MLPA, elle met en rapport la quantité de lectures produites et normalisée entre un échantillon cible et d'autres échantillons contrôles. Pour

l'instant, les essais réalisés indiquent une bonne sensibilité de la méthode pour les délétions totales et partielles de plusieurs exons. Elle semble encore limitée pour la détection de délétions d'une courte région (un exon seulement par exemple). La faible robustesse de certaines réactions PCR initiales peut expliquer la difficulté à normaliser les données d'un seul exon. De futurs essais permettront d'améliorer la sensibilité de cette méthode. La MLPA est encore réalisée pour les échantillons séquencés sans variant délétère mis en cause.

Notre solution interne demande quelques étapes manuelles de copier/coller et d'import/export de données. Ces étapes peuvent faire l'objet d'erreurs de manipulation et doivent donc être réalisées par une personne correctement formée au traitement de données. De même, la base de données FileMaker, qui est géré par le biologiste, utilise des scripts et des paramétrages qu'il faut maîtriser et pour lesquels il faut s'assurer de leur bon fonctionnement.

Du traitement bio-informatique dépend les performances d'analyse. Nous souhaitons avoir une solution la plus sensible possible et adaptée au cadre d'un rendu diagnostique. Les résultats du Variant Caller™ sont de bonne facture, et la sensibilité évaluée de notre deuxième design est équivalente aux méthodes actuelles de séquençage (95%). Le nombre limité d'échantillons testés nécessite une attention encore soutenue même après validation de la méthode. Les quelques échecs retenus ont deux origines différentes :

- Le défaut de couverture est une cause facilement identifiable et qui est résolue par notre analyse minutieuse des profondeurs. Cette analyse permet d'avoir sur les profondeurs, une information liée à l'étape d'appel de variant. En effet, il nous a semblé judicieux de prendre en compte la valeur de profondeur qui se rapporte aux lectures « utiles » au Variant Caller™, c'est-à-dire qui passent les filtres qualités de ce dernier. Par l'obtention du fichier *base à base*, l'information sur les profondeurs « utiles » est récupérée, tout comme les raisons de « no-call » de certaines bases.
- L'autre raison d'un échec d'appel de variant est interne et dépendante du *Variant Caller™*. La solution peut demander une optimisation des paramètres d'analyse. Dans ce cas, l'échange avec les techniciens de Life Technologies est important. Leur retour permet de mieux comprendre à quelle étape l'appel de variant a échoué.

Aucun algorithme d'appel de variant n'est parfait et l'utilisation de plusieurs logiciels en parallèle est souvent la meilleure façon d'augmenter la sensibilité de détection de variants.

L'utilisation de logiciel commerciaux est une solution. Le logiciel NextGENe® (Softgenetics®) permet de réaliser l'alignement des séquences, l'appel de variant, l'analyse des profondeurs et la recherche de grands réarrangements (aussi par « reads count »). Ce logiciel clé-en-main souffre d'une utilisation et d'une prise en mains parfois difficile, spécialement dans la lecture des résultats. Son paramétrage demande une bonne connaissance du logiciel, et les filtres à employer lors de l'appel de variants restent compliqués à optimiser.

Néanmoins, une seconde analyse en parallèle de notre pipeline « maison » avec ce logiciel permettrait peut-être d'obtenir une meilleure sensibilité. Il serait judicieux d'utiliser une seconde méthode d'analyse, au minimum pour les échantillons négatifs du *Variant Caller*TM. Le temps devant être investi à la mise au point du logiciel puis à la lecture des résultats est tout de même à prendre en considération.

D'autres outils dits « académiques » sont largement utilisés par la communauté scientifique. Souvent employé sous Linux, ce sont pour la plupart des scripts (Python, R, Perl, Java) qui sont actionnables par ligne de commandes. Ils demandent une connaissance approfondie en bio-informatique. Ces outils se démocratisent et de nombreuses solutions permettent d'apprendre les bases et leur utilisation pratique (D.U, MOOC, ou blogs internet). Très évolutifs, ils demandent tout de même une certaine maîtrise comme pour le paramétrage qui est fonction de la méthode de séquençage utilisée, ou des versions à mettre à jours.

L'enchaînement de scripts peut se faire de façon automatisée (Galaxy par exemple). Aux H.C.L, des solutions d'analyse intégrant un « *pipeline* » académique se mettent en place grâce au service bio-informatique. Un certain manque de recommandation ou de procédure d'analyse homogène au niveau national et international se fait sentir. Ce « péchés de jeunesse » du séquençage haut-débit limite une comparaison précise des performances inter-laboratoires et une optimisation globale des résultats.

Une dernière solution est d'externaliser l'analyse bio-informatique, le biologiste n'ayant plus que la validation des variants et l'interprétation biologique à effectuer. Des entreprises privées, comme Sophia Genetics (Suisse), propose une prise en charge complète, du traitement des données FASTQ à la validation de l'appel de variant, à travers une interface graphique agréable. Après plusieurs essais de validation selon un protocole fixe, le « *pipeline* » d'analyse spécifique de notre demande est créé. L'analyse se fait en quelques heures et les résultats sont visibles sur n'importe quel poste informatique.

Cette solution, évoluant rapidement avec le retour de leurs clients, permet de gagner du temps lors d'une production de séquençage soutenue. Même si le stockage des données est

intégré à l'offre, le prix est dépendant du nombre d'analyse et rapidement devenir vite conséquent. Point non négligeable dans une future phase active d'accréditation du laboratoire, leur solution d'analyse a obtenu la certification de marquage CE-IVD. Leurs infrastructures informatiques (serveur) sont également localisées en France, en accord avec la gestion des données de santé définie par le Code de la santé publique. Ces solutions externes ne doivent tout de même pas empêcher le biologiste de réaliser des essais internes pour son propre compte. L'objectif est de mieux connaître la technique utilisée et ses erreurs inhérentes, mais aussi de comprendre l'analyse réalisée par ces entreprises commerciales, qui le plus souvent ne dévoilent pas totalement leur méthode.

1.2. MISE EN PRODUCTION DE LA METHODE

Les résultats des essais de validation ont confirmé une sensibilité équivalente voire supérieure aux techniques actuelles de séquençage (exceptée pour la recherche de grands réarrangements qui reste en développement).

La validation technique et biologique, à travers des critères de qualité et notre « *pipeline* » d'analyse interne, est en place et performante. Une mise en production de la méthode est donc réalisable.

La méthode permet de séquencer 24 patients en une semaine environ : deux jours de fabrication de bibliothèques, un jour d'amplification clonale, un jour de séquençage, et un jour d'analyse des données.

La validation nécessite en moyenne un contrôle par méthode Sanger de 5 exons par patient. En comparaison, l'analyse isolée du gène *NF1* peut prendre de 4 à 6 semaines avec les techniques classiques pour une série de 15 patients.

La mise en production, outre une baisse du coût patient (200€ environ + 30€ de contrôle Sanger, voir annexe 2), permet donc une hausse spectaculaire du débit. Cette méthode apporte aussi, par l'analyse conjointe des gènes *NF1* et *SPRED1*, jusqu'ici analysés en séquentielle, un gain diagnostique important. Dans des pathologies pour lesquelles le diagnostic est principalement clinique, mais avec des symptômes parfois confondants, le séquençage des deux gènes limitera l'errance diagnostique. La confrontation de la clinique aux résultats moléculaires reste primordiale et constitue le principal apport du biologiste généticien.

Des perspectives d'amélioration demandent cependant un suivi avant une production en routine, et portent sur :

- Le design de la librairie, avec soit un nouveau design qui diminuerait les défauts de couverture, ou une optimisation de l'ajout d'amorces dans des pools prédéfinis.
- L'optimisation de certaines étapes de fabrication de banques.
- Finalisation et optimisation de l'analyse des grands réarrangements.
- Apporter des solutions aux défauts de détection du Variant Caller™.
- La création d'une série d'échantillons et d'un ensemble de données (*data set*) « contrôles », qui permettront de valider les changements : de réactifs et de protocole, ou de logiciel d'analyse.

Les perspectives d'avenir des technologies de séquençage haut-débit de 2^{ème} ou de 3^{ème} génération sont énormes, pour un coût toujours plus bas. Si les analyses d'exomes ou de génomes entiers posent encore des difficultés multiples aux généticiens (d'ordre techniques mais surtout d'interprétation), nul doute que l'évolution prochaine des technologies mènera à une analyse génétique toujours plus globalisée et multiple. L'avènement des « Omics » permettra une connaissance poussée des divers mécanismes cellulaires. Le biologiste généticien aura un rôle central, il permet de faire le lien entre les données techniques, technologiques, et cliniques. Sa plus-value viendra de l'interprétation d'un flux toujours plus important et complexe de données récoltées.

La mise en production de notre nouvelle méthode de séquençage n'est qu'un début dans la révolution génétique qui arrive. Pour les Rasopathies, dont la NF1 et le syndrome de Légit font partie, l'analyse complète de tous les effecteurs de la voie Ras-MAPK pourrait apporter, dans un futur proche, une nouvelle vision de ces pathologies et redéfinir leur champ d'action.

IV. CONCLUSIONS

MEMOIRE SOUTENU PAR M. BUISSON Adrien

La neurofibromatose de type 1 (NF1), ou maladie de Von Recklinghausen est l'une des maladies génétiques les plus fréquentes. Elle présente des manifestations cliniques variables associant des atteintes cutanées, neurologiques et certains cancers. Le diagnostic repose sur des critères cliniques établis et connus. Le déficit en neurofibromine en est la conséquence et elle est due à une atteinte d'un gène suppresseur de tumeur : *NF1 (17q11.2)*. Le syndrome de Legius, découvert en 2007, se caractérise principalement par la présence de signes cutanés similaires à ceux de la NF1. Le gène mis en cause est le gène *SPRED1*, qui doit depuis être analysé pour éliminer ce diagnostic différentiel.

La confirmation par l'analyse moléculaire est parfois utile dans la prise en charge des patients. Elle est proposée principalement en cas de manifestations incomplètes (patient jeune), pour les formes atypiques, pour les formes en mosaïque, et dans le cadre des demandes de diagnostic prénatal ou préimplantatoire.

Actuellement, la stratégie du laboratoire repose sur une analyse séquentielle de ces 2 gènes. L'analyse du gène NF1 est composée d'une recherche de grands réarrangements par technique M.L.P.A et d'un criblage par dHPLC des mutations, confirmées par séquençage Sanger. Pour le gène SPRED1, un séquençage systématique des exons par méthode Sanger puis une recherche par M.L.P.A sont réalisés. Cette méthode reste longue et coûteuse, mais elle permet d'atteindre une sensibilité proche des 95%.

Depuis quelques années, des séquenceurs de 2^{ème} génération permettent un séquençage haut-débit, avec un gain de temps et un « coût patient » abaissé. Notre laboratoire a mis au point deux designs de librairie NGS grâce à l'outil Ion Ampliseq™ Designer. Le séquençage

a été effectué par un automate Ion PGM™, qui utilise la technologie par semi-conducteur Ion Torrent™ (Life Technologies®). Cette méthode nous permet de réaliser l'analyse simultanée de tous les exons des gènes NF1 et SPRED1 ainsi que leurs régions 5' et 3' non traduites.

Afin de valider notre étude, les variants de patients précédemment génotypés ont été comparés aux données récoltées par NGS. Les résultats démontrent une bonne sensibilité de la méthode, avec la mise en évidence de la majorité des variants polymorphiques ou causals. L'analyse de séries de nouveaux patients ont été réalisées par séquençage haut débit et en parallèle par les techniques standards. La concordance des résultats permet la validation de la méthode avant mise en production. Le séquençage de référence Sanger reste à ce jour nécessaire pour la confirmation des variants et pour l'analyse des exons mal couverts en NGS. La recherche de grands réarrangements est encore à l'étude et reste à améliorer.

Cette nouvelle approche permet d'améliorer le temps de rendu du résultat et le coût de l'analyse reste modéré. L'objectif de notre travail est de remplacer à terme la méthode de criblage actuelle par la méthode NGS.

Le président du Jury, Pr. VINCIGUERRA Christine

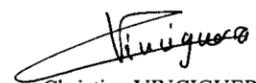


VU ET PERMIS D'IMPRIMER

Lyon, le 22 octobre 2015

Vu, la Directrice de l'ISPB – Faculté de Pharmacie de Lyon

Pour le Président de l'Université Claude Bernard



Professeur Christine VINCIGUERRA

ANNEXES

Annexe 1: Coordonnées génomiques des régions pseudogéniques du fichier Bed « pseudogènes » (Hg.19)

| Chromosome | Début (génomique) | Fin (génomique) | Nom donné |
|------------|----------------------|--------------------|-----------|
| chr15 | 21122012 | 21141048 | NF1P1 |
| chr15 | 22133181 | 22152165 | NF1P2 |
| chr21 | 15373246 | 15377600 | NF1P3 |
| chr14 | 19484443 | 19495427 | NF1P4 |
| chr18 | 14152996 | 14157343 | NF1P5 |
| chr22 | 16345912 | 16355362 | NF1P6 |
| chr14 | 20078032 | 20089290 | NF1P7 |
| chr2 | 131946021 | 131955578 | NF1P8 |

Annexe 2 : Coût global pour le séquençage d'un échantillon d'une série de 24 échantillons sur une puce Ion 318™. Le coût ne comprend pas les réactifs usuels aux manipulations techniques (cônes, tubes, utilisation des automates etc.).

| Etape du séquençage | Réactifs | Coût par patient (€) | Coût par patient protocole modifié (€) |
|---|--|----------------------|--|
| Préparation des échantillons | Dosage ADN par QuBit® (kit High Sensitivity) | | 0,30 |
| Fabrication de la librairie | Ion™ Ampliseq Librairie 2.0 | 123,36 | 61,68 |
| | Panel custom White Glove | | 2,52 |
| | Code-barres | | 2,47 |
| | Dosage post-amplification par QuBit® | | 0,90 |
| | Billes Macherey Nagel | | 0,83 |
| | Dosage qPCR | | 2,90 |
| Amplification clonale (OT2 + ES) | Ion PGM™ Hi-Q™ OT2 Kit | | 4,62 |
| | Dynabeads® MyOne™ Streptavidin C1 | | 1,60 |
| Séquençage | Ion PGM™ Hi-Q Seq Kit | | 9,05 |
| | Ion 318™ Chip Kit v2 | | 15,22 |
| | Ion Sphere™ Quality Control Kit | | 0,67 |
| Prix global | Total HT par échantillon (€) | 164,44 | 102,76 |
| | Total TTC par échantillon (€) | 197,33 | 123,31 |

Le coût par échantillon est d'environ :

- **197€** TTC pour le protocole normal.
- Il descend à **123€** TTC pour le protocole modifié avec moitié moins de réactif limitant (non utilisé).

Le référentiel des actes innovants hors nomenclature de biologie et d'anatomocytopathologie (RIHN) valable jusqu'au 03-2016 cote les nouveaux actes de séquençage NGS.

L'acte N351 : Forfait séquençage haut débit (NGS) > 20 kb et < 100 kb (cas index) correspondant à une cotation de : **BHN 5570** soit **1503,9€**.

Il comprend le forfait "accueil cas index" (BHN 370), le forfait analytique > 20 kb et < 100 kb (BHN 4000) et le forfait "interprétation" (BHN 1200).

Annexe 3: Dosage par qPCR sur LC480 des banques.

Kit Ion library taqMan Quantification Kit réf : 4468802

Etapes:

- Dilution des banques au 1/100 (2µl d'échantillon + 198µl d'eau)
- standard E. Coli DH10B (68pM) dilué au 1/10 (6.8pM), 1/100 (0.68pM) et 1/1000 (0.068pM) dans l'eau (Prévoir dilution : 2µL de prise d'essai et 18µL d'eau)
- blanc avec du TE
- tout en duplicate ou pas de duplicates (très peu de changement)
- possibilité de reprendre des points de gamme d'anciens Run à partir du logiciel d'analyse.

Mix:

| | |
|--------------------------|-------|
| 2X TaqMan mastermix | 10 µl |
| 20X Ion TaqMan assay | 1 µl |
| Banque ou standard ou TE | 9 µl |

- Ne pas oublier de sceller la plaque avec le film plastique (insérer avec dans le LC)

Programme: (Cf: Procédure Quantification Banque NGS sur LightCycler 480 (V2))

50°C - 2mn
95°C - 20s
95°C - 1s } **40X**
60°C - 20s }
40°C - 10s

- Noter sur la boîte combien de réactions ont été utilisées
- Après la manip, jeter la plaque, éteindre l'ordinateur, remplir le cahier.

Annexe 4: Procédure Quantification Banque NGS (Ion library TaqMan) sur LightCycler 480.

Etape 1 :

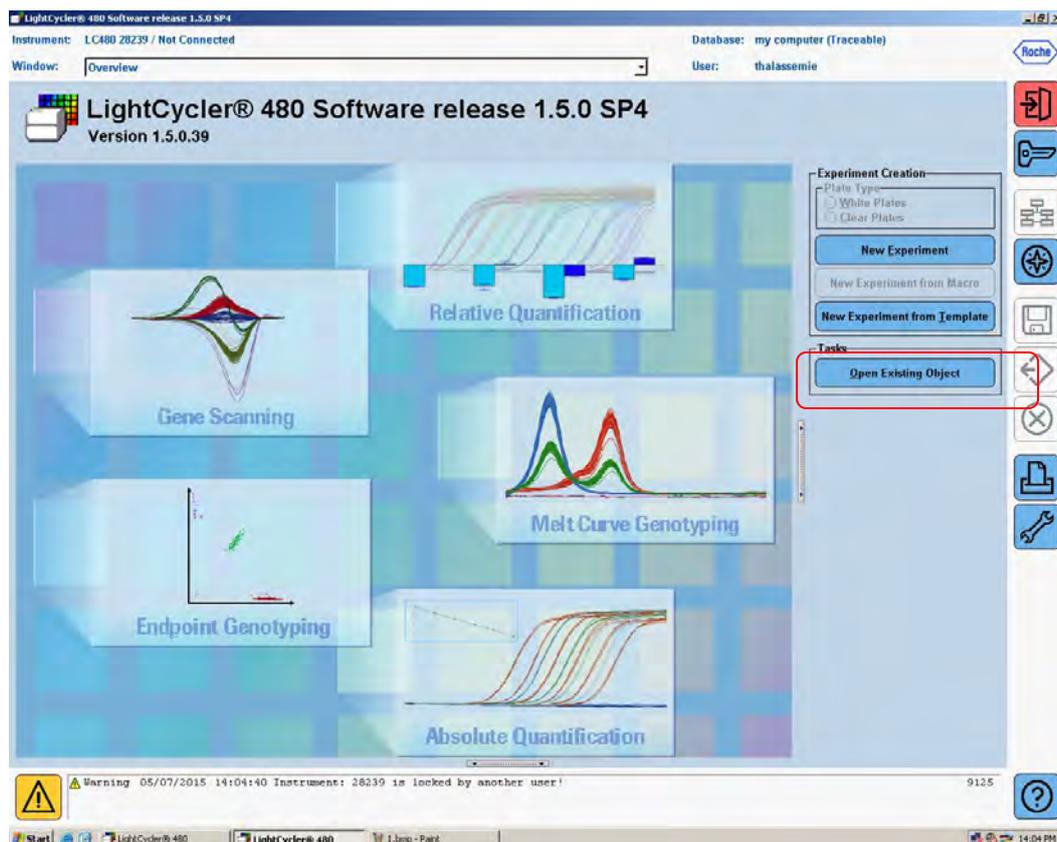
- Allumer l'ordinateur, l'écran, le lightCycler

Etape 2 :

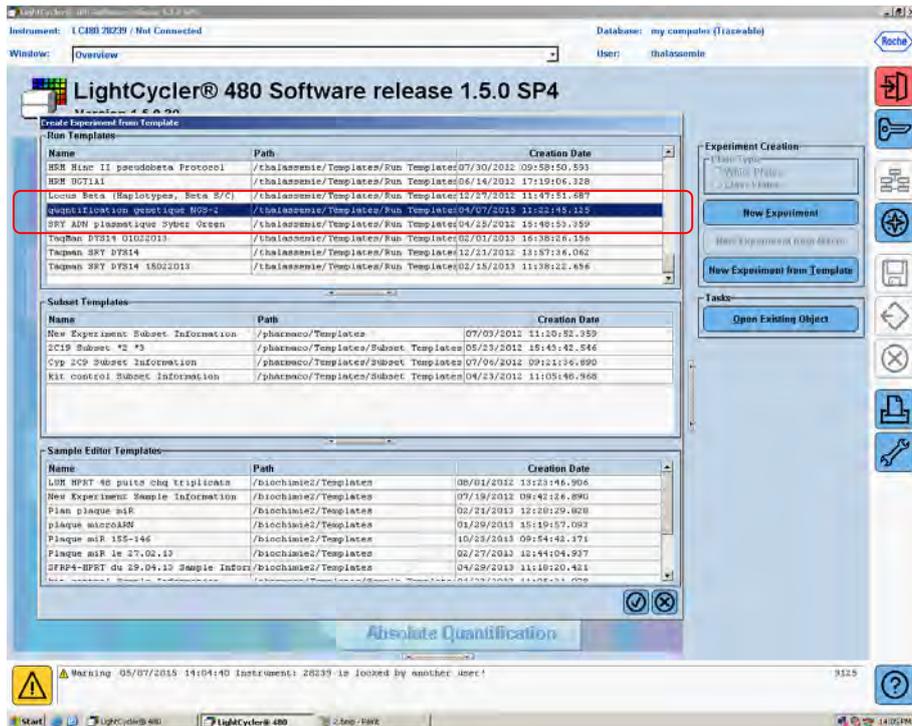
- Ouvrir le logiciel : **LC 480 Software**

Etape 3 :

- Le logiciel s'ouvre, cliquer sur « New experiment from template » (laisser « Plate type » : en « White plates »)



Chercher dans le menu déroulant : quantification génétique NGS-2

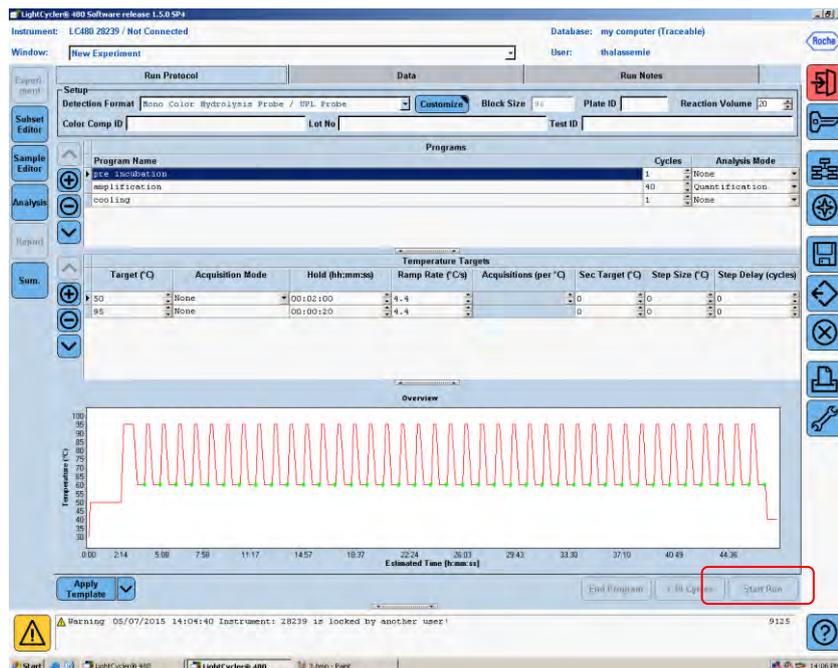


Puis valider

Etape 4 :

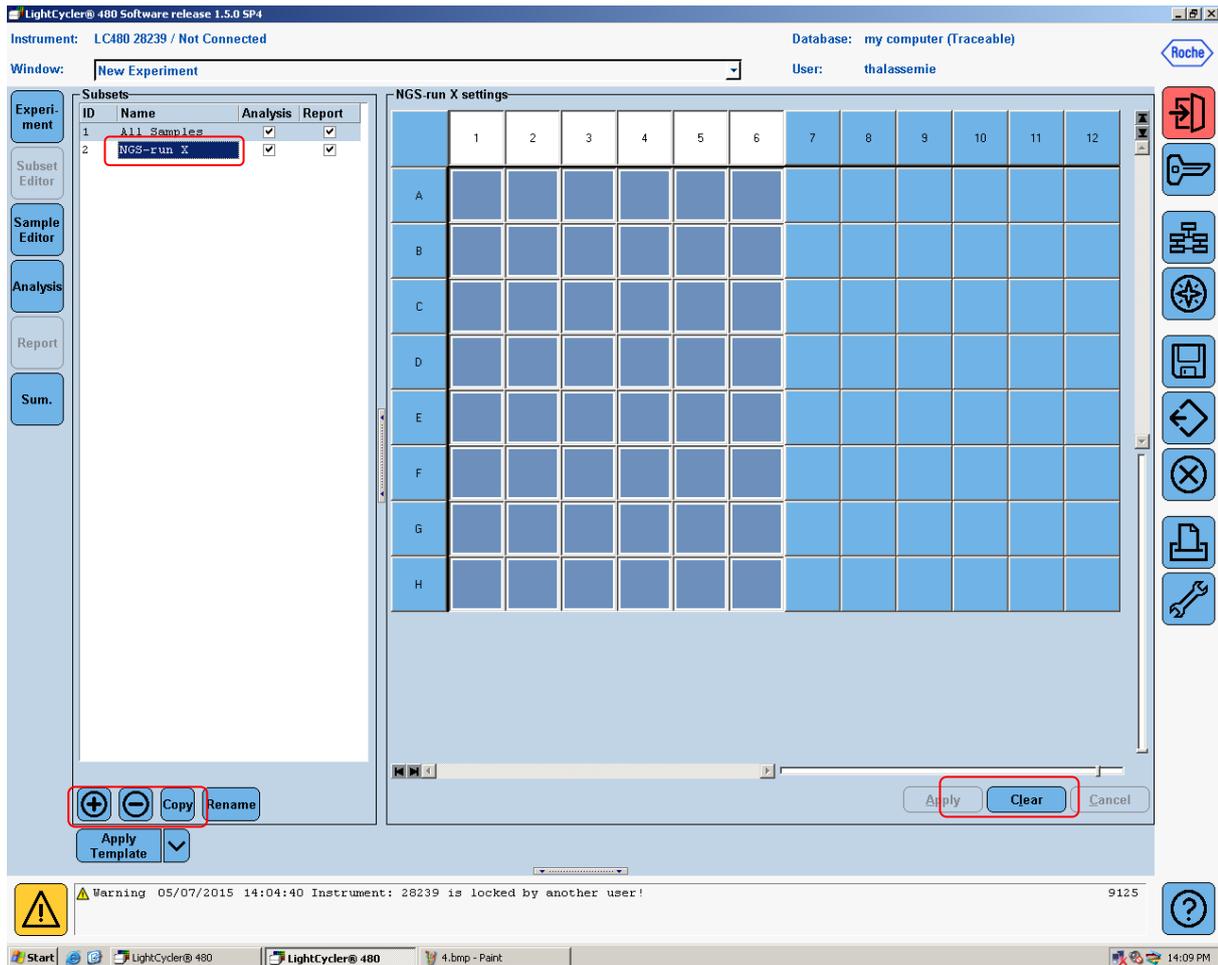
- Mettre la plaque dans le lightCycler, puis cliquer sur « Start Run »
- Attention, la plaque doit être scellée avec un film platique et insérée avec dans le LC
- Enregistrer le run dans « Thalassemie /experiment »

(Nb : le run démarre avant que le plan run ne soit fait)



Etape 5 :

- Cliquer sur « subset Editor »
- Cliquer sur le « + » pour créer un nouveau subset, sélectionner les cases utilisées, puis renommer le « new subset » avec le nom que vous voulez.
- Finir par « apply »



Etape 6 :

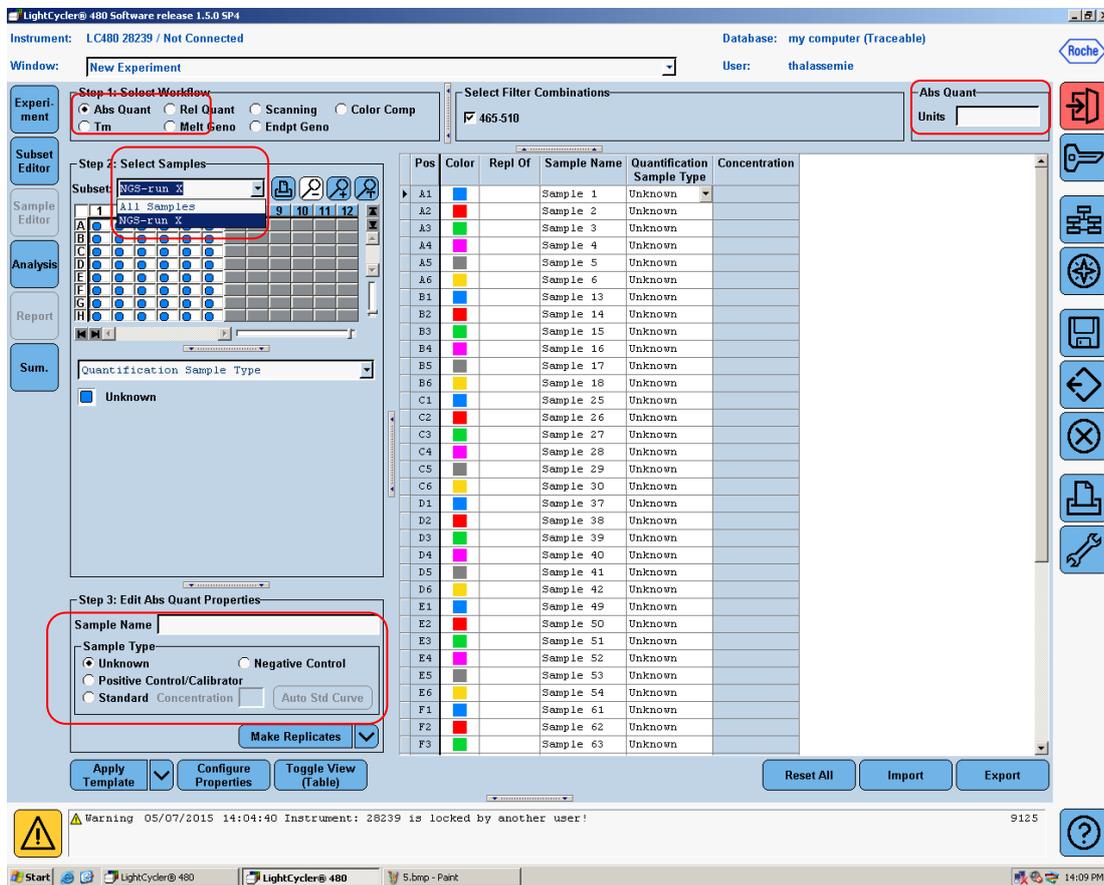
- Cliquer sur « Sample Editor »
- Sélectionner dans «step 1 » : « Abs Quant »
- Puis sélectionner son subset dans « Select Samples »
- Et déterminer en sélectionnant un par un les samples en définissant leur statut

Ex : Standard : renseigner la concentration en mettant les unités (pM) en haut à droite dans la case « Units »

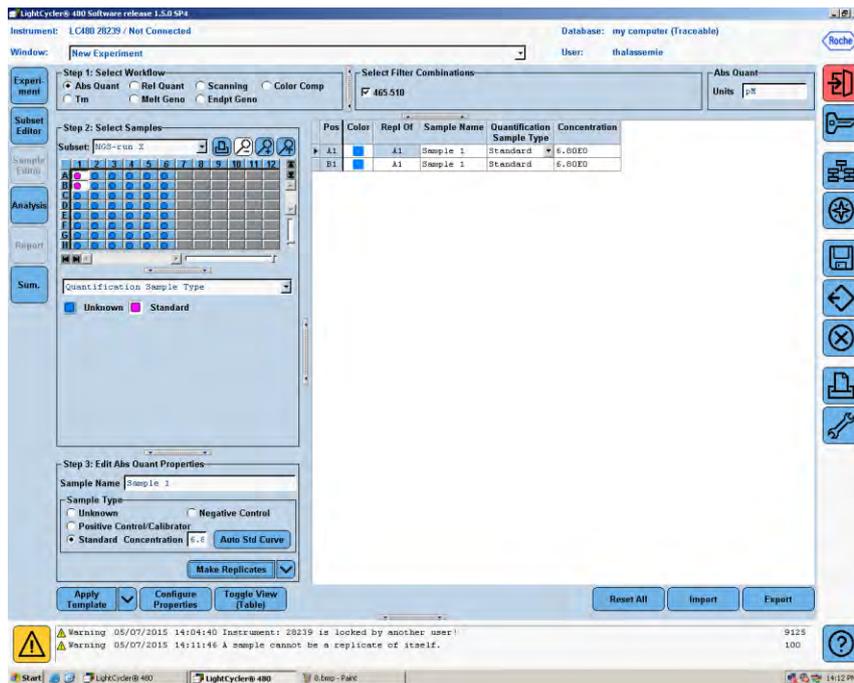
Patient : laisser en « unknown »

Blanc : « negativ control »

(Nb : pour faire des répliquas : sélectionner les 2 puits, renseigner le nom dans la case « sample name », puis cliquer sur « make replicates », les 2 lignes doivent avoir le même nom)

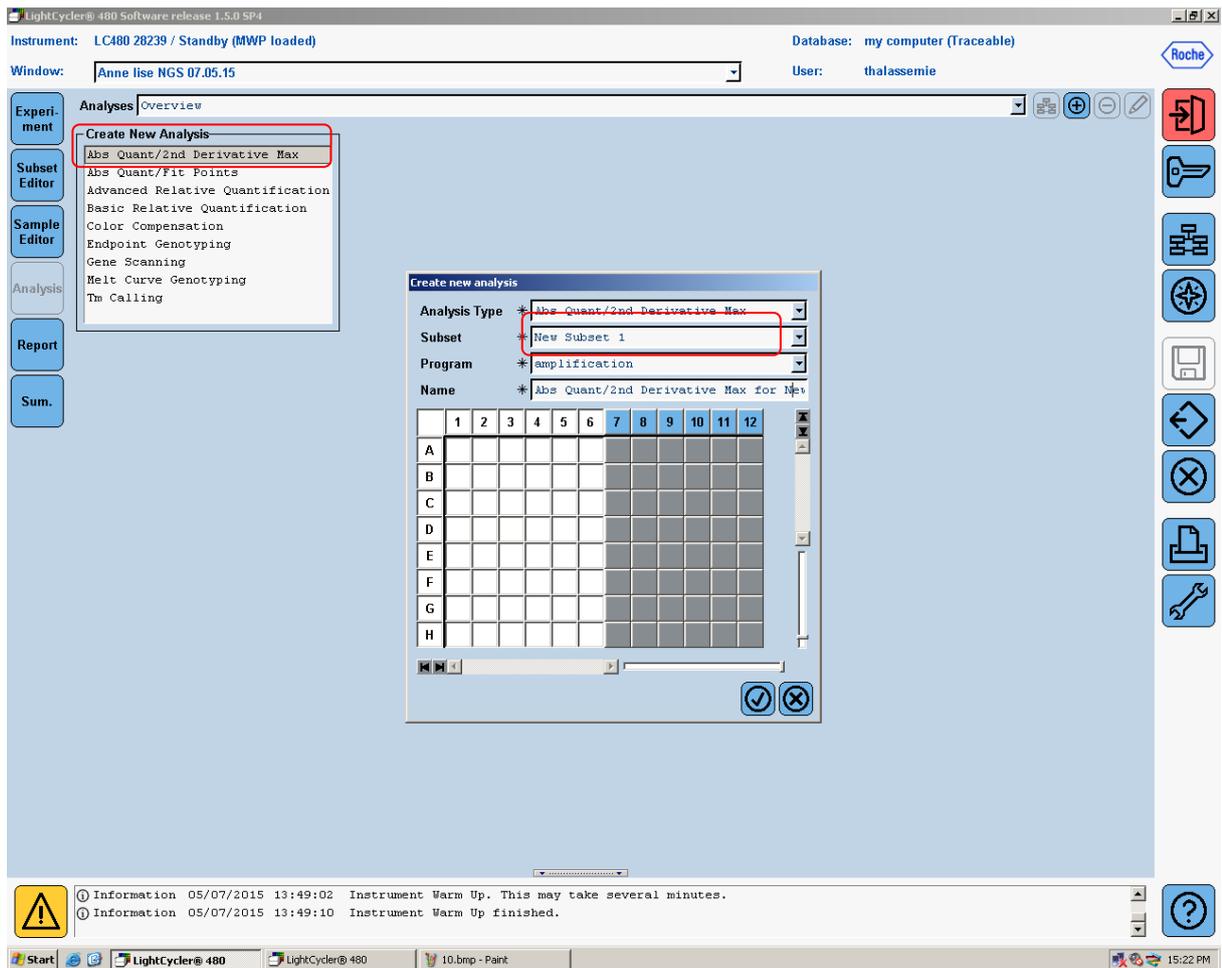


Ex de duplicates :



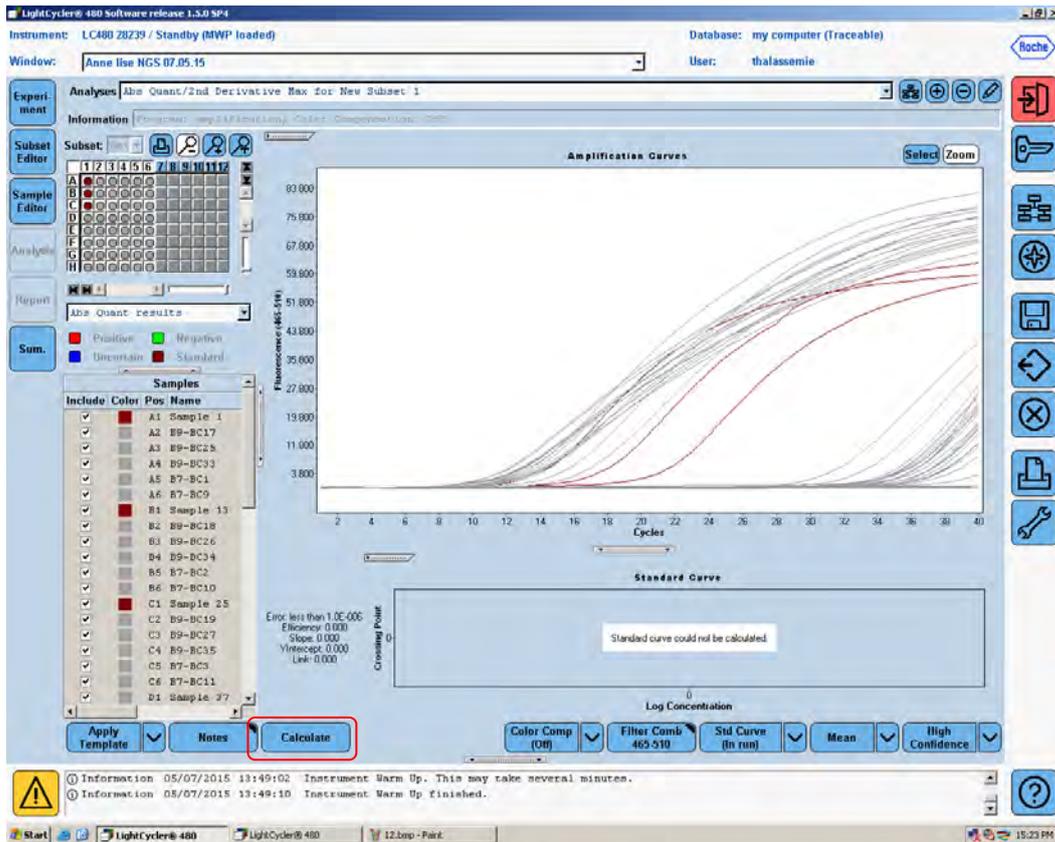
Etape 7 : Une fois l'analyse finie :

- Cliquer sur « Analysis », une fenêtre s'ouvre, renseigner n'importe quel caractère, puis valider.
- Cocher « abs quant / 2nd derivative Max »
- Une fenêtre s'ouvre, sélectionner le subset d'intérêt, puis valider

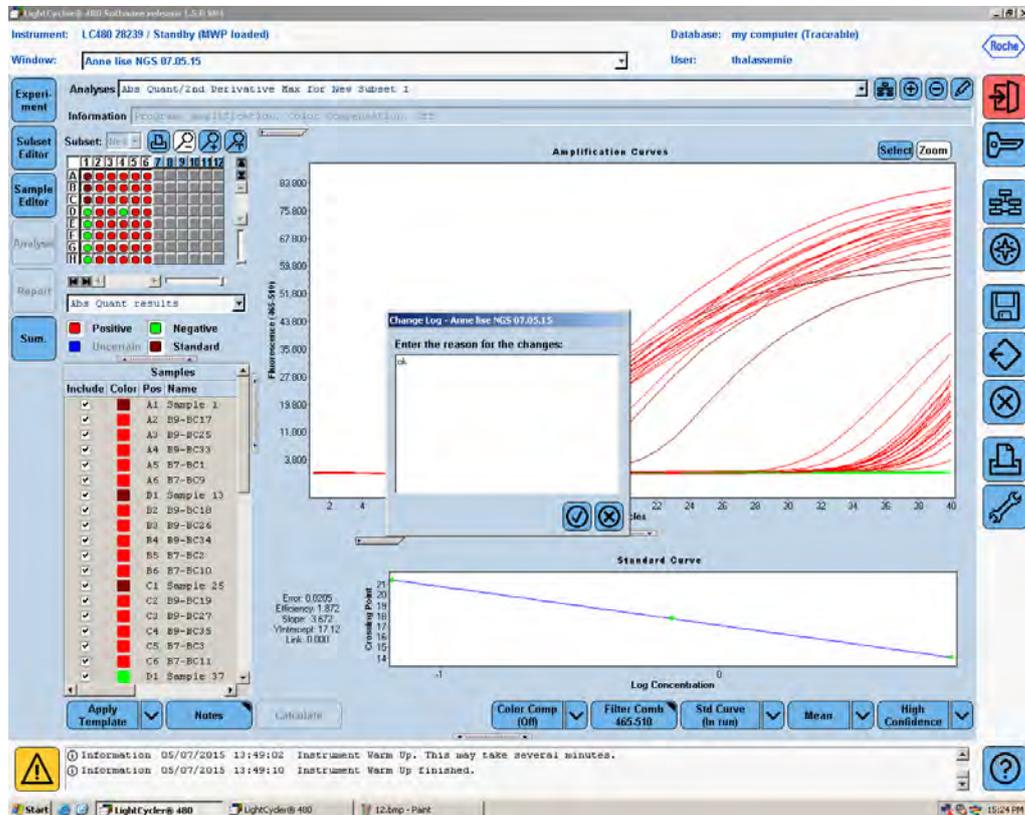


Etape 8 :

- L'analyse se présente à vous, cliquer sur « Calculate » pour générer le rapport d'analyse.



- Sauver en cliquant sur la disquette, ne fenêtre s'ouvre, renseigner n'importe quel caractère, puis valider.



Etape 9 :

- Dans l'onglet « report », sélectionner son subset d'intérêt
- Sélectionner les fichiers d'intérêt en cochant dans la colonne de gauche
- Cliquer sur « Generate », puis imprimer ou cliquer sur PDF pour l'enregistrer.

(Nb : attention l'ordinateur n'est pas sur le réseau, prendre une clé USB)

The screenshot shows the LightCycler 480 Software interface. The main window is titled 'View Report' and displays a report for 'Anne lise NGS 07.05.15'. The report includes the following sections:

- Experiment**

| | | | |
|---------------|--------------------------------|--------------------|---------------------|
| Creation Date | 05/07/2015 13:44:41 | Last Modified Date | 05/07/2015 15:25:09 |
| Operator | thalassemie | Owner | thalassemie |
| Start Time | 05/07/2015 13:49:01 | End Time | 05/07/2015 14:29:39 |
| Run State | Completed | Software Version | LCS480 1.5.0.39 |
| Macro | | Macro Owner | |
| Macro Status | | | |
| Templates | quantification genetique NGS-2 | Plate ID | 05391630 |
| Test ID | A-L NGS 6 mai 15 | Lot ID | |
| Color Comp ID | | | |
| Run Notes | | | |
- Programs**

| Program Name | Cycles | Analysis Mode | None | | | | |
|----------------|------------------|-----------------|------------------|-----------------------|-----------------|----------------|---------------------|
| pre incubation | 1 | Analysis Mode | None | | | | |
| Target (°C) | Acquisition Mode | Hold (hh:mm:ss) | Ramp Rate (°C/s) | Acquisitions (per °C) | Sec Target (°C) | Step size (°C) | Step Delay (cycles) |
| 50 | None | 00:02:00 | 4.40 | | 0 | 0 | 0 |
| 95 | None | 00:00:20 | 4.40 | | 0 | 0 | 0 |

| Program Name | Cycles | Analysis Mode | Quantification | | | | |
|---------------|------------------|-----------------|------------------|-----------------------|-----------------|----------------|---------------------|
| amplification | 40 | Analysis Mode | Quantification | | | | |
| Target (°C) | Acquisition Mode | Hold (hh:mm:ss) | Ramp Rate (°C/s) | Acquisitions (per °C) | Sec Target (°C) | Step size (°C) | Step Delay (cycles) |
| 95 | None | 00:00:01 | 4.40 | | 0 | 0 | 0 |
| 60 | Single | 00:00:20 | 2.20 | | 0 | 0 | 0 |

The interface also shows a 'Report Settings' panel on the left with a grid for selecting subsets (A-H, 1-12) and a 'Generate' button at the bottom. The status bar at the bottom indicates 'Instrument Warm Up' messages.

- Penser à retirer la plaque, tout éteindre et se noter dans le cahier.

Procédure V2 , juin 2015, Auteurs : Maltavernes Anne-lise, Buisson Adrien.

Annexe 5: Procédure QuBit®

Contrôle des concentrations sur Qubit 2.0

1. Préparation des tubes 0.5mL : pour les 2 standards de calibration et 4 échantillons
2. Préparation de la « Qubit working solution » : dilution au 1/200ème :

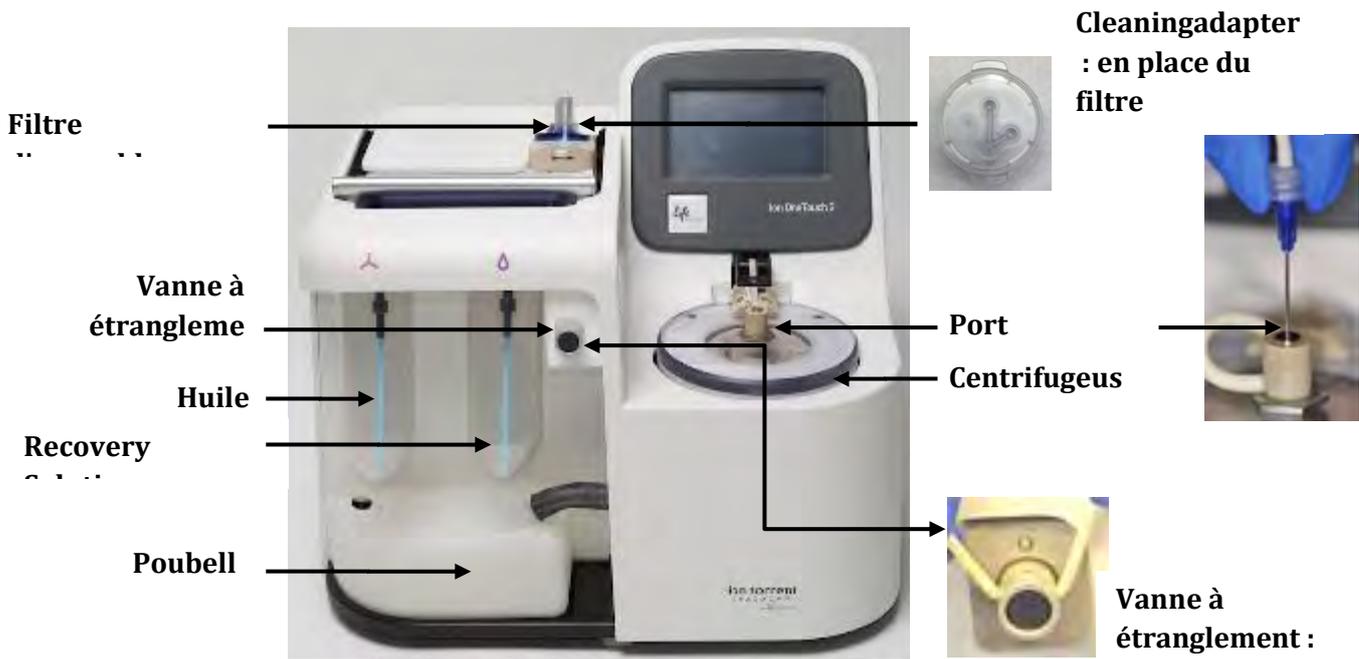
| | | |
|---------------|-----|------|
| Nb éch | 1 | 7 |
| Qubit Reagent | 1 | 7 |
| Qubit Buffer | 199 | 1393 |

3. Charger 190µL de Qubit working solution dans les 2 tubes standards et 199 à 196µL dans les tubes échantillons
4. Ajouter 10µL de Qubit standard dans le tube standard puis mélanger en vortexant qq secondes. Attention aux bulles !
5. Ajouter 1 µL d'ADN dilué à ≈25ng/µL dans les tubes échantillons puis mélanger en vortexant qq secondes. Attention aux bulles !
6. Laisser incuber 2 min à T° ambiante
7. Sur le Qubit : appuyer sur « **DNA** » puis sélectionner « **dsDNA High Sensitivity** ». L'écran de calibration apparaît automatiquement, appuyer sur « **Yes** » pour faire une nouvelle calibration ou « **No** » pour utiliser une ancienne calibration.
8. Nouvelle calibration : insérer le tube contenant le Standard 1 dans le Qubit, fermer le couvercle et appuyer sur « **Read** ». La lecture dure environ 3s puis enlever le standard 1. Puis insérer le tube contenant le Standard 2 dans le Qubit, fermer le couvercle et appuyer sur « **Read** ».
9. Lecture des échantillons : insérer un tube contenant un échantillon puis appuyer sur « **Read** ». Pour l'échantillon suivant appuyer sur « **Read Next Sample** »
10. Pour enregistrer la valeur donnée par le Qubit et faire le calcul de concentration de l'échantillon non dilué : appuyer sur « **Calculate Stock Conc** », sélectionner le volume d'échantillon utilisé pour la dilution, la calcul de concentration se fait.
Pour changer d'unités, appuyer sur « **ng/ml** » et choisir l'unité souhaitée : **ng/µL** en appuyant dessus, appuyer ailleurs sur l'écran pour ressortir.

Annexe 6: : protocole OT Hi-Q et PGM Hi-Q

Amplification clonale de la banque (OneTouch2 Hi-Q)

L'amplification clonale des banques se fait par PCR en émulsion, sur l'appareil OneTouch2 (OT2). Les amplifications clonales se font dans des émulsions contenant un fragment unique d'ADN avec adaptateur + une Ion SphereParticle (ISP) recouverte d'amorces, support des amplicons. Ce sont ces ISP recouvertes d'amplicons qui prendront place dans chaque puits de la puce de séquençage.



- Cleaning

Le « cleaning » ou lavage de l'appareil se fait normalement à la fin de chaque run. Si les runs sont éloignés, il est conseillé d'en faire un avant aussi. Pour la procédure de lavage il faut :

- Que le niveau d'huile soit supérieur à 20ml, si ça n'est pas le cas, remplir au moins à moitié avec de l'ion One TouchOil
- Laisser l'OT2 Amplification plate dans sa loge (si besoin, on peut mettre celle du run à venir)
- Placer la protection au-dessus de l'amplification plate
- Laisser le cleaning adapter, le cas échéant en placer un à la place du filtre d'assemblage
- Placer un tube de 50ml à droite de l'instrument (=poubelle) et y placer la seringue

Sur l'écran OT2, sélectionner « clean » et suivre les instructions sur l'écran

- Préparation du OneTouch2 pour le run

- Sur l'instrument, sélectionner « open lid » puis soulever le couvercle de la centrifugeuse
- Placer **150µl** de Ion One Touch Breaking Solution dans 2 recovery tubes (attention solution visqueuse, de plus si cette solution présente un précipité blanc, la chauffer à 30°C)
- Placer les deux recovery tubes dans la centrifugeuse du OT ainsi que le recovery router en le pinçant pour qu'il atteigne sa loge puis fermer le couvercle de la centrifugeuse
- Enlever le couvercle en plastique au-dessus du bloc chauffant puis pousser la poignée vers l'arrière pour ouvrir le bloc chauffant

- Remplacer l'OT2 amplification plate par une neuve :

- Ecriture « ion » vers le haut et vers le manipulateur
- Le port d'entrée à droite de la plate, doit s'aligner avec le trou le plus à gauche de l'OT2
- Faire passer la canule avec la seringue vers l'avant de l'OT2 et loger la canule sous le clips
- Faire passer la canule dans la vanne à étranglement, il faut l'enclencher en tirant vers le haut
- Placer la seringue dans le port d'injection de la centrifugeuse de l'OT2. En étant à la verticale, la pousser jusqu'au bout
- Rabattre la poignée vers l'avant

- Mise à niveau de l'huile (tube de gauche) :

- Si nouveau kit, remplacer le tube et le sipper par ceux du kit. Sinon laisser en place
- Mélanger le flacon d'huile « Ion OneTouchOil » (450ml) 3 fois par inversion
- Remplir le tube à moitié en évitant de faire des bulles
- Replacer le tube en le vissant fermement

- Mise à niveau de la solution recovery (tube de droite)

- S'assurer que la recovery solution est claire, si ça n'est pas le cas, chauffer à 30°C jusqu'à éclaircissement
- Si nouveau kit, remplacer le tube et le sipper par ceux du kit. Sinon laisser en place
- Mélanger le flacon « Ion PI OT2 Recovery Solution » (450ml) 3 fois par inversion
- Remplir le tube au quart en évitant de faire des bulles
- Replacer le tube en le vissant fermement
- Vider la poubelle (sous les tubes) et la remettre en place
- Vérifier qu'il n'y a pas trop d'huile dans le petit plateau tiroir sous la poubelle

- Préparation et installation de la solution d'amplification

- Préparer la solution d'amplification (zone middlepost-PCR)

- Sortir 1 tube d'Ion PGM Hi-Q Reagent Mix (bouchon violet, -20°). Vortexer 30s, centrifuger, équilibrer à température ambiante.
- Sortir l'Ion PGM Hi-Q Enzyme mix (bouchon marron; -20), centrifuger, laisser sur la glace
- Sortir l'Ion PGM Hi-Q ISPs (bouchon noir ; -20°) et équilibrer à température ambiante, vortexer 1min, centrifuger 2sec puis mélanger par aspiration refoulement
- Vortexer 5s la banque à amplifier, préalablement diluée dans 25µl, centrifuger et laisser sur la glace
- Préparer un filtre d'assemblage et un petit flacon d'Ion OT2 reaction oil pour la suite

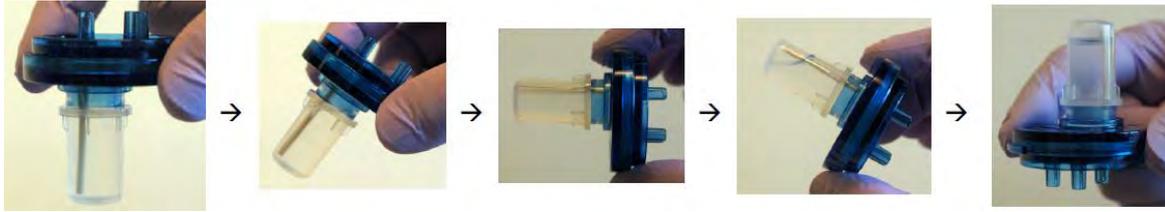
Dans un tube violet contenant 800µl de Ion PGM Hi-Q Reagent Mix ajouter :

- 25µl H₂O
- 50µl d'Ion PGM Hi-Q Enzyme Mix (bouchon marron)
- 25µl de la banque diluée à la bonne concentration
- 100µl Ion PGM Hi-Q ISPs (bouchon noir)

(Remarque : en cas de séquençage de *novo* remplacer 10µl d'eau par 10µl de Ion Calibration)

Charger L'Ion Plus Filter Assembly (filtre d'assemblage) dans les 15min suivant la préparation

- Placer le filtre d'assemblage sur un portoir 15ml (ou un tube 15ml) et repérer le sample port (dans le prolongement de la canule)
- Vortexer 5s l'échantillon préparée, centrifuger puis transférer la solution dans l'assemblage à l'aide d'une P1000 (1000µl) par le sample port (cône bien perpendiculaire dans le sample port)
- Nettoyer si nécessaire avec une serviette en papier
- Transférer lentement 850µl + 850µl d'Ion OT2 Reaction Oil dans l'assemblage, nettoyer si nécessaire
- Retourner le filtre d'assemblage dans le sens des aiguilles d'une montre lorsque la canule inférieure est positionnée à gauche jusqu'à ce que la canule pointe vers le haut



- Lancement du run sur l'OT2

- Insérer l'assemblage en haut de l'appareil OT2 (écriture « ion » dirigée vers le technicien)
- Lancer le run en sélectionnant « RUN » sur l'écran
- Sélectionner le kit OT2 utilisé dans le menu déroulant : Ion PGM Hi-Q OT2 kit -200 ou -400 selon la taille de notre librairie
- Choisir « mode assisté » pour avoir un récapitulatif des étapes de préparation de l'OT2 ou « Expert » pour lancer sans vérifier les étapes (penser à vider la poubelle)

- Le run dure plusieurs heures, il faut récupérer les échantillons au bout de 16h maximum-

- Récolte des ISP

- Suivre les consignes sur l'écran de l'OT2 pour centrifuger les échantillons
- Tout de suite après, sélectionner « open lid » enlever le routeur, transférer délicatement les 2 tubes sur un portoir, ne pas congeler les ISP, procéder immédiatement à la suite
- Pour chacun des tubes, enlever la totalité du surnageant **sauf 100µl**, remettre en suspension les culots à la pipette, ajouter **500µl** d'Ion Wash Solution dans chacun des tubes et mélanger à la pipette.
- Transférer les **2 X 600µl** dans un tube LoBind 1.5mL.

- Stoppingpoint, 2-3jrs +4°C

- Si un précipité blanc apparaît, incubé à 50°C pendant 2 min
- centrifuger 15500g 2min30 et retirer tout le surnageant sauf **100µl**
- Transférer **2µl** dans un tube 0.2 correctement annoté pour qualification et quantification ultérieure des ISP sur le Qubit

-Cleaningdu OneTouch2

- Vérifier le niveau d'huile soit supérieur à 20ml, si ça n'est pas le cas, remplir au moins à moitié avec de l'ion OneTouchOil
- Laisser l'OT2 Amplification plate dans sa loge ainsi que le couvercle

- Remplacer le filtre d'assemblage par un cleaning adapter neuf
- Placer un tube de 50ml à droite de l'instrument qui servira de poubelle
- Placer la seringue du port d'injection dans le tube 50ml

Sur l'écran OT2, sélectionner « clean » et suivre les instructions sur l'écran

Enrichissement des ISP positives sur OneTouch ES



- Préparation de la solution Melt off

Dans un tube 1.5ml annoté « melt off » ajouter :

- **280µl** de solution Tween (Flacon fournie dans le kit-Température ambiante)
- **40µl** de NaOH 1M

320µl de solution Melt off à utiliser dans les 24h

- Lavage des DynabeadsMyOneStreptavidin C1

- Sortir le flacon de DynabeadsMyOneStreptavidin C1 (+4°C)
- Vortexer le flacon de Dynabeads 30s
- Transférer **13µl** de dynabeads dans un tube 1.5ml LoBind
- Placer le tube sur l'aimant 2mn puis jeter le surnageant
- Jeter le surnageant, ajouter **130µl** de Ion PI MyOneBeads Wash Solution (bouchon vert, température ambiante)
- Vortexer 30s, centrifuger, mettre de côté pour la suite

- Remplissage de la barrette de 8 puits pour le OneTouch ES

- Récupérer une barrette, la loger dans la plaque blanche à récupérer sur le OneTouch ES en disposant l'extrémité carrée à gauche, l'extrémité ronde à droite

- **Puit 1** : s'assurer que les **ISP** sont dans **100µl**, mélanger 10X par va-et-vient puis transférer dans P1
- **Puits 2** : transférer **130µl** des **dynabeads** préalablement lavées
- **Puits 3-4-5** : transférer **300µl** d'**Ion OneTouch Wash Solution** (flacon, température ambiante)
- **Puits 6** : reste vide
- **Puit 7** : **300µl** de solution **Melt off** fraîchement préparée
- **Puit 8** : reste vide



Volumes : 100µl 130µl 300µl 300µl 300µl 300µl 300µl

- Préparation et lancement du OneTouch ES

- Placer la barrette (avec la plaque blanche) sur l'appareil
- Mettre un cône dans le Tip loader (cylindre métallique à gauche), détacher le bras mécanique et l'enfoncer dans le Tip loader pour charger le cône puis remettre le bras en place
- Placer un tube 0.2, ouvert, dans la petite loge situé derrière le Tip loader
- Mélanger par va-et-vient les Dynabeads dans le puit 2
- Lancer le run en appuyant sur « start », durée environ 35mn
- A la fin du run, l'instrument affiche « End », presser start/stop pour arrêter l'alarme
- Fermer immédiatement le tube 0.2 d'ISP enrichi, s'assurer que le volume est supérieur à **200µl**, si le volume est très inférieur à 200µl, contacter Life Technologies
- Centrifuger immédiatement le tube **3 minutes à 15550g**, éliminer tout le surnageant **sauf 10µl** (comparer avec un tube témoin)
- Ajouter **200µl** de Ion OT2 Wash solution, mélanger par aspiration refoulement
- Centrifuger de nouveau **3 minutes à 15500g**, vérifier l'absence de billes au fond du tube, puis éliminer tout le surnageant **sauf 10µl**.
- Ajouter de l'Ion OT2 Wash solution pour un volume final de **100µl**, mélanger par aspiration refoulement.

-Stopping point, conserver les ISP jusqu'à 7 jours à 4°C-

Contrôle qualité des ISP positives

Contrôle sur les **2µl** d'ISP mis de côté avant l'enrichissement sur le OneTouch ES

- Hybridation des sondes de marquages

- Décongeler les Ion Probes (-20°C, bouchon bleu), l'annealing buffer (-20°C, bouchon blanc) les standards AF488 et AF647 et le Quality control Wash buffer (flacon à aliquoter pour décongélations futures plus rapides).

- Ajouter **19µl** d'Annealing buffer et **1µl** d'Ion Probes directement dans le tube 0.2 contenant les ISP

- Placer le tube dans un thermocycleur pour hybrider les Ion Probes : **95°-2mn, 37°-2mn**

- Lavage des ISP

- 3 {
- Ajouter au tube d'ISP **200µl** Wash buffer
 - Vortexer, centrifuger 1mn30, 15500g
 - Enlever tout le surnageant **sauf 10µl**
- Après le dernier lavage, ajouter **190µl** de Wash buffer, mélanger par va et vient et transférer dans un tube Qubit

- Dosage des ISP au Qubit

- Sur le Qubit, sélectionner « Ion » puis « AF488 »

- Insérer l'échantillon, fermer le couvercle, sélectionner « Read » puis noter la valeur mesurée

- Faire de même avec le standard AF488 puis le blanc (200µl wash buffer)

- Sélectionner « Home » pour revenir à l'écran d'accueil puis « Ion » et enfin « AF647 »

- Insérer l'échantillon, fermer le couvercle, sélectionner « Read » puis noter la valeur mesurée

- Faire de même avec le standard AF647 puis le blanc (200µl wash buffer)

- Reporter les valeurs dans le tableau Excel « QubitEasyCalculator » (copier-coller le tableau présent dans le dossier du dernier run proton)

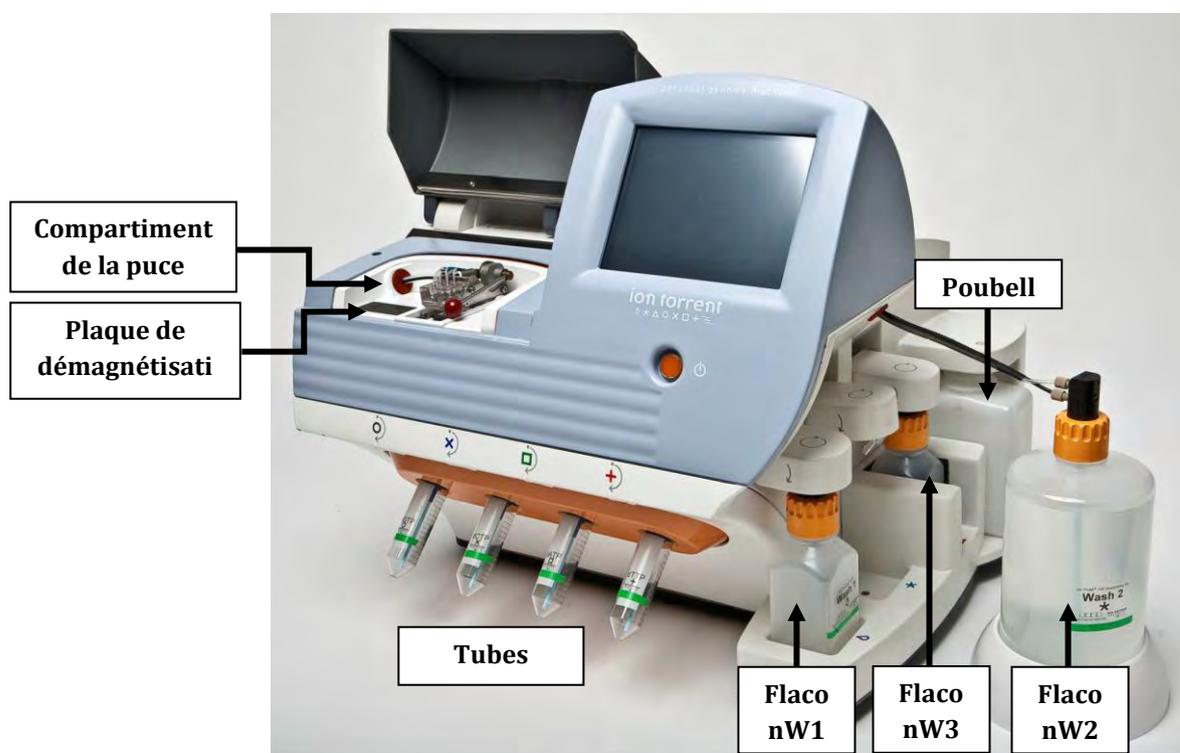
- Si l'on a utilisé un kit OneTouch d'un nouveau lot, se procurer le nouveau facteur de correction sur l'Ion community grâce au nouveau numéro de lot.

- Les conditions optimales pour le run de séquençage et un bon remplissage de la puce sont :

- % template ISP : **10% à 25%**, en dessous trop d'ISP vides, au-dessus trop d'ISP polyclonales

- Valeurs absolues **AF488>3000** (ISP totales) et **AF647>1000** (ISP chargées), en dessous pas assez d'ISP pour remplir la puce de façon optimale

Maintenance du PGM



- Lavage à l'eau

Le lavage à l'eau se pratique avant chaque initialisation, après un run, si le séquenceur sera inutilisé dans les 72h, ou après un lavage à la javel.

- Insérer une cleaning chip (puce déjà utilisée), dédiée au lavage à l'eau, dans l'instrument
- Sélectionner « Cleaning » sur l'instrument puis « 18MΩ water cleaning »
- Retirer tous les flacons et les tubes de l'appareil, laisser les sippers en place
- Ajouter **250ml d'eau 18MΩ** dans un flacon de lavage de 250ml et le placer en position W1 après avoir préalablement nettoyer le sipper en position W1 à l'aide d'une pissette remplie d'eau 18MΩ
- En suivant les instructions sur l'écran, placer un flacon de 2L vide (de lavage) en position W2 et un flacon de 250ml vide (de lavage) en position W3.
- Vider la poubelle et la remettre en place
- Placer des bacs collecteurs sous les sippers des dNTPs, « next » pour lancer le cleaning
- A la fin du cleaning, enlever les flacons W1, W2 et W3, placer les sippers dans des tubes 50ml
- Sélectionner « next » pour revenir au menu principal et procéder si besoin à l'initialisation

- Lavage à la javel

Le lavage à la javel est à effectuer une fois par semaine ou si les réactifs ont été laissés plus de 48h sur l'appareil.

- Remplir une bouteille en verre avec 1L d'eau milliQ
- Ajouter une pastille javel « Ion PGM cleaning tablet » (température ambiante, kit séquençage)
- Attendre 10mn la dissolution complète de la tablette
- Après dissolution, ajouter **1ml de NaOH 1M**
- Filtrer la solution avec un filtre 0.22 ou 0.45µm (solution à utiliser dans les 2 à 3 heures)
- Sélectionner « Clean » sur le l'appareil, puis « chlorite cleaning », insérer une puce dédiée javel
- Enlever et mettre de côté les flacons W1, W2, W3
- Rincer les flacons W1, W2, W3 dédiés au chlorite cleaning deux fois à l'eau milliQ ainsi que le sipper W1 à l'aide d'une pissette d'eau 18MΩ
- Remplir le flacon W1 avec **250ml de javel** préalablement filtrée et le fixer en position W1
- Fixer les flacons W2 et W3 vides dans leurs emplacements respectifs
- Placer des bacs collecteurs sous les sippers des dNTPs
- Vider la poubelle et la remettre en place
- Sélectionner « Next » sur l'appareil et suivre les instructions pour lancer le cleaning
- A la fin du lavage, presser « Next » pour revenir au menu principal
- Remplacer les flacons dédiés javel par les flacons pour lavage à l'eau après les avoir préalablement rincés (ainsi que les sippers).
- Faire un lavage à l'eau (cf ci-dessus)

Initialisation du PGM

- Important

- Pour toute initialisation, le premier run doit être lancé dans l'heure après la fin de l'initialisation et le dernier run doit être lancé dans les 27h.
- Pour éviter les contaminations, changer de gants entre chaque dNTP
- Les flacons W1, W2, W3 sont réutilisables 3 fois pour un total de 4 initialisations mais pourront resservir pour les lavages eau ou javel

- Avant de commencer

- Décongeler sur la glace les tubes de dNTPs (-20°C, tubes bouchons noir, bleu, vert, rouge)
- Vérifier que la pression de la bouteille d'azote est au minimum à **500psi**, le cas échéant la changer
- A l'aide d'un marqueur, surligner la ligne des 2L sur le flacon W2 (anneau)
- Préparer 500µl de NaOH 100mM

- Préparation du flacon W2

- Lors de sa première utilisation, conditionner la W2 en la remplissant d'eau 18MΩ jusqu'à la ligne des 2L. Ajouter la bouteille entière de Wash 2 Conditioning Solution. Mélanger 5X par retournement. Laisser la bouteille à température ambiante pendant au minimum 8h ou la nuit. Cette bouteille peut servir pour 40 initialisations.
- Rincer 3X la bouteille **W2** avec 200ml d'eau 18MΩ et la dégazer
- Remplir la bouteille jusqu'à la marque des **2L** avec de l'eau **18MΩ** issue directement de l'appareil
- Ajouter une **bouteille entière** de « **Ion PGM Hi-Q Sequencing W2 Solution** » dans le flacon W2
- Ajouter **70µl** de NaOH 100mM fraîchement préparé dans le flacon W2
- Mettre le bouchon du flacon et mélanger 5X par inversion puis procéder immédiatement à la suite

- Préparation des flacons W1 et W3

- Rincer 3X les flacons W1 et W3 avec 50ml d'eau 18MΩ
- Ajouter **350µl** de NaOH 100mM fraîchement préparé dans le **flacon W1** puis le boucher
- Ajouter du « **Ion PGM™ Sequencing W3 Solution** » jusqu'à la ligne de **50ml** du flacon **W3** puis le boucher

- Commencer l'initialisation

- Laisser les sippers des dNTPs en place, laisser la puce de lavage en place et sélectionner « Initialize »
- Scanner ou saisir le barcode de la bouteille W2 du kit. Le cas échéant sélectionner dans le menu déroulant « Ion PGM Hi-Q Sequencing Kit ». Sélectionner « Next »
- Valider que la puce de lavage, les sippers et les bacs collecteurs sont en place. Sélectionner « Next »
- Le système vérifie la pression de gaz, si la pression est suffisante, sélectionner « next » pour commencer l'initialisation. Si la pression est basse, presser « yes » pour relancer la vérification. Si elle demeure basse, contacter le support
- Avec des gants propres, suivre les instructions à l'écran pour remplacer le sipper W2 (long). Mettre le plus rapidement possible le flacon W2 en place, visser fermement le bouchon et presser « next »
- Changer de gants pour placer des sippers (gris) propres et les flacons W1 et W3. Presser « next »
- L'appareil vérifie l'absence de fuite, remplit le flacon W1 puis règle le pH de la solution W2 (~30mn)

- Préparation et installation des dNTPs

- Manipuler les tubes de dNTP avec précaution pour éviter les contaminations, ne pas les inverser
- Une fois les tubes stocks de dNTP décongelés, les vortexer, centrifuger et laisser sur la glace
- Identifier 4 nouveaux tubes de 50ml pour les solutions dNTP avec les étiquettes fournies dans le kit
- En utilisant des pointes à filtre, transférer délicatement **20µl de chaque dNTP** dans le tube correspondant. Changer de gant entre chaque dNTP
- Une fois l'initialisation passée, suivre les instructions à l'écran pour enlever sippers et bac collecteur
- Avec de nouveaux gants, placer de nouveaux sippers (bleus) pour les dNTPs
- Fixer à l'appareil et serrer chaque tube de dNTP en respectant l'ordre **G, C, A, T** de gauche à droite
- Suivre les instructions à l'écran pour terminer l'initialisation : l'appareil vérifie l'absence de fuite et remplit les tubes de dNTP avec la solution W2. (Pendant ce temps, s'occuper des ISPs et planifier le run)
- A la fin de l'initialisation, l'appareil vérifie les pH des solutions et affiche un « **passed** » vert si il affiche un écran rouge d'échec d'initialisation, lire l'appendice B « troubleshooting » sur le protocole officiel Life technologies
- Sélectionner « next » pour terminer le processus d'initialisation et retourner au menu principal

Protocole de séquençage Puce 316 ou 318

- Avant de commencer

Sortir les primers de séquençage (-20°C, bouchon blanc), les ISPs contrôles (-20°C, bouchon transparent) laisser décongeler et mettre sur la glace.

Penser à planifier le run, il est nécessaire d'avoir la version 4.2. Sélectionner Ion PGM HI-Q Sequencing Kit.

- Préparation des ISPs positives et annealing des primers de séquençage

- Vortexer les **ISPs contrôles** 2 sec, centrifuger puis transférer **5µl** directement dans notre tube d'ISP

- Bien mélanger les ISPs par va et vient à la pipette puis centrifuger 15500g, 2mn

- Retirer délicatement tout le surnageant **sauf 15µl** (comparer avec un tube témoin)

- Ajouter **12µl** de « sequencing primer » et confirmer que le **volume total** est de **27µl**, compléter si nécessaire avec de « l'annealing buffer » (Flacon à +4°C)

- Bien mélanger par va et vient à la pipette pour dissoudre le culot d'ISP (~40X)

- Placer le tube dans un thermocycleur pour hybrider les amorces de séquençage : **95°-2mn**, **37°-2mn** à la fin du programme le tube d'ISP peut rester à température ambiante dans le thermocycleur

- Vérification de la puce

- Pour éviter d'endommager la puce avec des décharges électrostatiques,

- ne pas poser la puce directement sur la paillasse (ou autre surface).
- toujours placer la puce soit sur la plaque de démagnétisation du PGM soit sur la nacelle de la centrifugeuse à puce
- ne pas porter de gants lors des transferts de puce depuis ou vers le PGM

- Retirer une puce neuve de son emballage la poser sur la plaque de démagnétisation du PGM, l'annoter et garder l'emballage pour le scanner plus tard

- Sélectionner « run » sur le PGM et suivre les instructions à l'écran, lorsque demandé :

> Introduire une puce de lavage (la même que pour l'initialisation) cliquer sur next

> Se démagnétiser sur la plaque du PGM, retirer la puce de lavage et loger la nouvelle puce

> Scanner le barcode sur l'emballage de la puce ou sélectionner «change» pour le faire manuellement

- Sélectionner « chip check » sur l'écran et vérifier l'absence de fuite sur la puce (si fuite, « abort »)

- A la fin du « chip check », si la puce est passée, sélectionner « next ».

- Si la puce n'est pas passée, enlever et replacer la puce, sélectionner « calibrate » pour relancer la procédure. Si la puce passe sélectionner « next » le cas échéant sélectionner « Main menu » et recommencer avec une nouvelle puce

- Remettre une puce de lavage, vider la poubelle du PGM et passer à la suite

- *Liaison de la polymérase aux ISPs*

- Après l'hybridation des primers aux ISPs, retirer le tube du thermocycleur et ajouter **3µl** de « Ion PGM Hi-Q Sequencing Polymerase » (-20°C) pour un volume total de 30µl

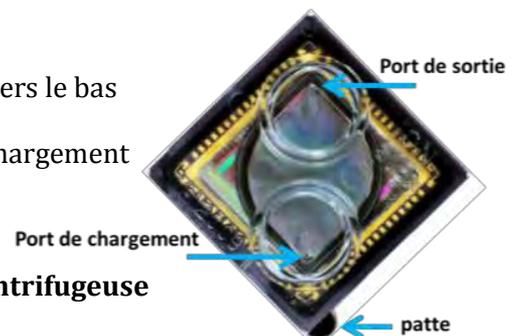
- Mélanger par va et vient à la pipette, incuber **5mn** à température ambiante

- *Chargement de la puce*

- Tourner la puce à 45° afin que le port de chargement soit vers le bas

- Insérer fermement la pointe de la pipette dans le port de chargement

- Aspirer et jeter le plus de liquide possible



Dans les étapes suivantes penser à bien équilibrer la centrifugeuse

- Placer la puce dans la nacelle de la centrifugeuse **face vers le bas** avec la **patte de la puce pointant vers l'intérieur** de la centrifugeuse

- Centrifuger 5s pour vider complètement la puce, retirer la puce et essuyer toute trace de liquide sur la nacelle - Placer la puce dans la nacelle à plat sur la paillasse

- Après les 5mn d'incubation des ISPs, transférer la totalité des ISPs (~30µl) dans la pointe d'une pipette Rainin SR-L200F puis insérer fermement la pointe dans le port de chargement de la puce

- Vider lentement le contenu de la pointe dans la puce par un mouvement de rotation régulier du poussoir de la pipette, laisser un petit volume mort dans la pointe pour éviter d'introduire une bulle

- Aspirer et jeter tout liquide qui ressortirait par le port de sortie de la puce

- Placer la puce dans la centrifugeuse avec la patte de la puce **pointant vers l'intérieur**

- Centrifuger 30s puis retirer la puce de la centrifugeuse

- Mélanger l'échantillon dans la puce :

- régler la pipette sur **25µl** et presser le piston

- Incliner la puce à 45° et introduire le cône de la pipette dans le port de chargement

- Sans retirer le cône faire 3X un lent mouvement de va et vient en évitant de faire des bulles

- Centrifuger la puce 30s avec la patte de la puce **pointant vers l'extérieur**

- Re-mélanger l'échantillon dans la puce puis centrifuger 30s, la patte **pointant vers l'intérieur**

- pour le séquençage 400pb uniquement, re-mélanger l'échantillon dans la puce 5x

- Tourner la puce à 45° et, à l'aide d'une pipette, aspirer et jeter le plus de liquide possible par le port de chargement. Si nécessaire faire une centrifugation de 5s, patte vers l'extérieur et re-aspirer
- S'il reste encore du liquide, taper plusieurs fois et doucement la patte de la puce sur la paillasse puis aspirer et jeter le liquide. Ne pas injecter d'air dans la puce
- Sélectionner « Next » à l'écran et procéder immédiatement à la sélection et au lancement du run

Sélection et lancement du Run de séquençage Puce 314, 316 et 318

- Sélectionner le run, préalablement planifié sur le seueur du PGM, en sélectionnant « browse » dans le menu déroulant « plannedrun » puis sélectionner « Next »
- Toutes les caractéristiques du run planifié s'affichent alors à l'écran. Vérifier qu'il n'y a aucune erreur, le cas échéant apporter les modifications nécessaires à l'aide des menus déroulants
- Sélectionner « Next » puis « OK » pour confirmer les caractéristiques ou « Cancel » pour remodifier les caractéristiques
- Suivre les instructions à l'écran pour mettre en place la puce dans l'appareil puis cliquer « Next »
- L'appareil va laver la puce des ISPs non fixées puis la calibrer. Vérifier durant ce processus qu'il y a aucune fuite puis fermer le couvercle
- A la fin de la calibration (~1mn) l'écran indique le statut de la calibration
 - Si la puce a passé la calibration, cliquer « Next » pour lancer le run
 - Si la calibration a échoué, cliquer « Abort », remettre la puce en place et represser « Calibrate »
 - Si la deuxième calibration échoue lancer tout de même le run et informer le support
- Après 90s le run démarre automatiquement ou sélectionner « Next » pour le lancer immédiatement
- Éviter de toucher l'appareil, les tubes ou les flacons pendant le run, ça peut impacter sur la qualité des mesures.
- A la fin du run, l'écran reviendra au menu d'accueil, on peut alors lancer un deuxième run ou un lavage/initialisation

BIBLIOGRAPHIE

- ¹ Huson SM, Hughes RAC. The neurofibromatoses : a pathogenetic and clinical overview. Chapman and Hall. 1994
- ² Huson SM, Compston DAS, Harper PS. A genetic study of von Recklinghausen neurofibromatosis in South East Wales. II : Guidelines for genetic counselling. *J Med Genet.* 1989;26:712-21
- ³ Nordling C. A new theory on cancer-inducing mechanism. *Br J Cancer.* 1953. 7 (1): 68-72
- ⁴ Marte B. Milestone 9: (1953) Two-hit hypothesis - It takes (at least) two to tango. *Nature Milestones Cancer.* 2007
- ⁵ McCubrey JA, Steelman LS, Chappell WH, et al. Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochim Biophys Acta.* 2007. 1773 : 1263-84
- ⁶ Chardin, P.; Camonis, J.; Gale, N.; Van Aelst, L.; Schlessinger, J.; Wigler, M.; et al. Human Sos1: A guanine nucleotide exchange factor for Ras that binds to GRB2. *Science.* 1993. 260 (5112): 1338-1343
- ⁷ Lièvre A, Laurent-Puig P. La voie de signalisation RAS/MAPK. *Cancéro dig.* 2010. 2 (1) : 38-42
- ⁸ Schubbert S, Shannon K, Bollag G. Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer.* 2007. 7:295-308
- ⁹ Rajalingam K, Schreck R, Rapp UR, et al. Ras oncogenes and their downstream targets. *Biochim Biophys Acta.* 2007. 1773 : 1177-95
- ¹⁰ Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000. 100(1):57-70
- ¹¹ Blog science. <http://www.blogscience.com/199/the-hallmarks-of-cancer-les-mecanismes-fondamentaux-du-cancer> (consulté au 15.01.2016)
- ¹² Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011. 144(5):646-74
- ¹³ Denayer E, de Ravel T, Legius E. Clinical and molecular aspects of RAS related disorders. *J Med Genet.* 2008. 45(11):695-703
- ¹⁴ Roberts PJ, Der CJ. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene.* 2007. 26: 3291-3310
- ¹⁵ Galabova-Kovacs G, Matzen D, Piazzolla D, Meissl K, Plyushch T, Chen AP, et al. Essential role of B-Raf in ERK activation during extraembryonic development. *Proc Natl Acad Sci U S A.* 2006. 103:1325-1330.
- ¹⁶ Jiang H, Grenley MO, Bravo MJ, Blumhagen RZ, Edgar BA. EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in *Drosophila*. *Cell Stem Cell.* 2011. 8(1):84-95
- ¹⁷ Bentires-Alj M, Kontaridis MI, Neel BG. Stops along the RAS pathway in human genetic disease. *Nat Med.* 2006. 12:283-5
- ¹⁸ D. Bessis. Syndromes neuro-cardio-facio-cutané et RASopathies. *Images en Dermatologie.* 2010. 4
- ¹⁹ Rauen K. A. The RASopathies. *Annu Rev Genomics Hum Genet.* 2013. 14:355-369

-
- ²⁰ Yamamoto GL, Aguena M, Gos M, Hung C, Pilch J, Fahiminiya S, et al. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. *J Med Genet.* 2015. 52:413–421
- ²¹ Nemcikova M, Vejvalkova S, Fencel F, Sukova M, Krepelova A. A novel heterozygous RIT1 mutation in a patient with Noonan syndrome, leukopenia, and transient myeloproliferation—a review of the literature. *Eur J Pediatr.* 2015. 1-6
- ²² Antônio JR, Goloni-Bertollo EM, Trídico LA. Neurofibromatosis: chronological history and current issues. *Anais Brasileiros de Dermatologia.* 2013. 88(3):329-343.
- ²³ Crump T. Translation of case reports in Ueber die multiplen Fibrome der Haut und ihre Beziehung zu den multiplen Neuromen by F. v. Recklinghausen. *Adv Neurol.* 1981. 29:259–75
- ²⁴ J Cox. Quest for Quasimodo (V. Hugo). *Br Med J.* 1985. 291(6511): 1801–1803
- ²⁵ Tibbles JA, Cohen MM. The Proteus syndrome: The Elephant Man diagnosed. *Br Med J.* 1986. 293:683–5
- ²⁶ Huson SM, Compston DAS, Clark P et al. A genetic study of von Recklinghausen neurofibromatosis in south east Wales. 1. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J Med Genet.* 1989. 26: 704–711
- ²⁷ Ferner RE, Huson SM, Thomas N, et al. Guidelines for the diagnosis and management of individuals with neurofibromatosis 1 (NF1). *J Med Genet.* 2007. 44:81–88
- ²⁸ Diwakar G, Hornyak TJ. Café-au-lait patches and senile plaques: How APpt the connection? *J Invest Dermatol.* 2006. 126(3):547-50
- ²⁹ Kopf AW, Levine LJ, Rigel DS, Friedman RJ, Levenstein M. Prevalence of congenital-nevus-like nevi, nevi spili, and cafe au lait spots. *Arch Dermatol.* 1985. 121: 766-9
- ³⁰ Obringer AC, Meadows AT, Zackai EH. The diagnosis of neurofibromatosis 1 in the child under the age of 6 years. *Am J Dis Child.* 1989. 143:717-9
- ³¹ S. Pinson, A. Créange, S. Barbarot, J.-F. Stalder, Y. Chaix, D. Rodriguez, et al. Recommandations pour la prise en charge de la neurofibromatose 1. *J Fr. Ophthalmol.* 2002. 25(4) : 423-433
- ³² Huson S, Jones D, Beck L. Ophthalmic manifestations of neurofibromatosis. *Br J Ophthalmol.* 1987. 71:235–238
- ³³ Listernick R, Louis DN, Packer RJ, Gutmann DH. Optic pathway gliomas in children with neurofibromatosis 1 : consensus statement from the NF1 Optic Pathway Glioma Task Force. *Ann Neurol.* 1997. 41:143-9
- ³⁴ Crawford AH, Jr, Bagamery N. Osseous manifestations of neurofibromatosis in childhood. *J Pediatr Orthop.* 1986. 6:72–88
- ³⁵ K. Jett and J. M. Friedman. Clinical and genetic aspects of neurofibromatosis 1. *Genetics in Medicine.* vol. 2010. 12(1): 1–11
- ³⁶ Hyman SL, Shores A, North KN. The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology.* 2005. 65:1037–1044

-
- 37 Virdis R, Sigorini M, Laiolo A, et al. Neurofibromatosis type 1 and precocious puberty. *J Pediatr Endocrinol Metab.* 2000. 13:841–844
- 38 Friedman JM, Arbiser J, Epstein JA, et al. Cardiovascular disease in neurofibromatosis 1: report of the NF1 Cardiovascular Task Force. *Genet Med* 2002. 4:105–111
- 39 McCaughan JA, Holloway SM, Davidson R, Lam WW. Further evidence of the increased risk for malignant peripheral nerve sheath tumour from a Scottish cohort of patients with neurofibromatosis type 1. *J Med Genet.* 2007. 44:463–466
- 40 Sharif S, Moran A, Huson SM, et al. Women with neurofibromatosis 1 are at a moderately increased risk of developing breast cancer and should be considered for early screening. *J Med Genet.* 2007. 44:481–484
- 41 R.E. Ferner. Neurofibromatosis 1. *Eur. J. Hum. Genet.* 2007. 15: 131–138
- 42 Riccardi VM. Neurofibromatosis: Clinical heterogeneity. *Curr Probl Cancer.* 1982. 7:1–34
- 43 Oguzkan S, Cinbis M, Ayter S, Anlar B, Aysun S. Familial segmental neurofibromatosis. *J Child Neurol.* 2004. 19:392–394
- 44 Victor FC. Segmental neurofibromatosis. *Dermatol Online J.* 2005. 11:20–1
- 45 Lazaro C, Ravella A, Gaona A, Volpini V, Estivill X. Neurofibromatosis type 1 due to germ-line mosaicism in a clinically normal father. *N Engl J Med.* 1994. 331:1403–1407
- 46 Rasmussen SA, Yang Q, Friedman JM. Mortality in neurofibromatosis 1: an analysis using U.S. death certificates. *Am J Hum Genet.* 2001. 68:1110–1118
- 47 Zöller ME, Rembeck B, Odén A, Samuelsson M and Angervall L. Malignant and benign tumors in patients with neurofibromatosis type 1 in a defined Swedish population. *Cancer.* 1997. 79(11):2125–2131
- 48 Lazaro C, Gaona A, Lynch M, Kruyer H, Ravella A, Estivill X. Molecular characterization of the breakpoints of a 12-kb deletion in the NF1 gene in a family showing germ-line mosaicism. *Am J Hum Genet.* 1995. 57:1044–1049
- 49 Verlinsky Y, Rechitsky S, Verlinsky O et al: Preimplantation diagnosis for neurofibromatosis. *Reprod Biomed Online.* 2002. 4:218–222
- 50 Wolkenstein P, Durand-Zaleski I, Moreno JC, Zeller J, Hemery F, Revuz J. Cost evaluation of the medical management of neurofibromatosis 1: a prospective study on 201 patients. *Br J Dermatol.* 2000. 142:1166–1170
- 51 Serletis D, Parkin P, Bouffet E, Shroff M, Drake JM, Rutka JT. Massive plexiform neurofibromas in childhood: natural history and management issues. *J Neurosurg.* 2007.106: 363–367
- 52 Lee AG. Neuroophthalmological management of optic pathway gliomas. *Neurosurg Focus.* 2007. 23:E1
- 53 Kleinerman RA. Radiation-sensitive genetically susceptible pediatric subpopulations. *Pediatr Radiol.* 2009. 39:27–S31
- 54 Sharif S, Ferner R, Birch JM, et al. Second primary tumors in neurofibromatosis 1 patients treated for optic glioma: substantial risks after radiotherapy. *J Clin Oncol.* 2006. 24:2570–2575

-
- ⁵⁵ Brems H, Chmara M, Sahbatou M, Denayer E, Taniguchi K, Kato R, Somers R, Messiaen L, De Schepper S, Fryns JP, Cools J, Marynen P, Thomas G, Yoshimura A, Legius E. Germline loss-of-function mutations in SPRED1 cause a neurofibromatosis 1-like phenotype. *Nat Genet.* 2007. 39:1120–6
- ⁵⁶ Messiaen L, Yao S, Brems H, Callens T, Sathienkijkanchai A, Denayer E, Spencer E, et al, Legius E. Clinical and mutational spectrum of Neurofibromatosis type 1-like syndrome. *JAMA.* 2009. 302:2111–8
- ⁵⁷ Denayer E, Chmara M, Brems H, Kievit AM, van Bever Y, Van den Ouweland AM et al. Legius syndrome in fourteen families. *Hum Mutat.* 2011. 32:1985–98
- ⁵⁸ Laycock-van Spyk S, Jim HP, Thomas L, Spurlock G, Fares L, Palmer-Smith S, Kini U, Saggari A, Patton M, Mautner V, Pilz DT, Upadhyaya M. Identification of five novel SPRED1 germline mutations in Legius syndrome. *Clin Genet.* 2011. 80:93–6
- ⁵⁹ Pasmant E, Ballerini P, Lapillonne H, Perot C, Vidaud D, Leverger G, Landman-Parker J. SPRED1 disorder and predisposition to leukemia in children. *Blood.* 2009. 114:1131
- ⁶⁰ Pasmant E, Gilbert-Dussardier B, Petit A, de Laval B, Luscan A, Gruber et al. SPRED1, a RAS MAPK pathway inhibitor that causes Legius syndrome, is a tumor suppressor downregulated in paediatric acute myeloblastic leukaemia. *Oncogene.* 2014. 34(5):631-8
- ⁶¹ Muram-Zborovski TM, Stevenson DA, Viskochil DH, Dries DC, Wilson AR, Mao Rong. SPRED 1 mutations in a neurofibromatosis clinic. *J Child Neurol.* 2010. 25:1203–9
- ⁶² Barker, D., Wright, E., Nguyen, K., Cannon, L., Fain, P., Goldgar, D. et al, M. Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. *Science.* 1987. 36: 1100-1102
- ⁶³ Viskochil D, Buchberg AM, Xu G, Cawthon RM, Stevens J, Wolff RK, et al. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell.* 1990. 62:187–92
- ⁶⁴ Viskochil, D., Cawthon, R., O'Connell, P., Xu, G. F., Stevens, J., Culver, M. et al. The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene. *Molec. Cell. Biol.* 11: 906-912, 1991.
- ⁶⁵ Atwal, J. K., Pinkston-Gosse, J., Syken, J., Stawicki, S., Wu, J., Shatz, C. et al, M. PirB is a functional receptor for myelin inhibitors of axonal regeneration. *Science.* 2008. 322: 967-970
- ⁶⁶ Abramowicz A, Gos M. Neurofibromin in neurofibromatosis type 1 - mutations in NF1 gene as a cause of disease. *Dev Period Med.* 2014. 18(3):297-306
- ⁶⁷ Ohno, S. so much "junk" DNA in our genome. *Brookhaven symposia in Biology.* 1972. 23, 366-370
- ⁶⁸ Kamalika S., Tapash C.G. Pseudogenes and their composers: delving in the 'debris' of human genome. *Briefings in Functional Genomics.* 2013. 12(6): 536-547
- ⁶⁹ Pei B et al. The GENCODE pseudogene resource. *Genome Biology.* 2012. 13:R51

-
- 70 Marchuk DA, Saulino AM, Tavakkol R, Swaroop M, Wallace MR, Andersen LB et al. Genomics. cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product. 1991. 11(4):931-40
- 71 Danglot G, Régnier V, Fauvet D, Vassal G, Kujas M, Bernheim A. Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5' part of the gene. Hum Mol Genet. 1995. 4(5):915-20
- 72 Kaufmann D, Müller R, Kenner O, Leistner W, Hein C, Vogel W et al. The N-terminal splice product NF1-10a-2 of the NF1 gene codes for a transmembrane segment. Biochem Biophys Res Commun. 2002. 294(2):496-503
- 73 Barron V.A., Lou H. Alternative splicing of the neurofibromatosis type I pre-mRNA. Biosci. Rep. 2012. 32:131-138
- 74 Gutmann DH, Cole JL, Collins FS. Modulation of neurofibromatosis type 1 gene expression during in vitro myoblast differentiation. J Neurosci Res. 1994. 37(3):398-405
- 75 Uchida T, Matozaki T, Suzuki T, Matsuda K, Wada K, Nakano O, et al. Expression of two types of neurofibromatosis type 1 gene transcripts in gastric cancers and comparison of GAP activities. Biochem Biophys Res Commun. 1992. 187(1):332-9
- 76 Park VM, Kenwright KA, Sturtevant DB, Pivnick EK. Alternative splicing of exons 29 and 30 in the neurofibromatosis type 1 gene. Hum Genet. 1998. 103(4):382-5
- 77 Skuse GR, Cappione AJ, Sowden M, Metheny LJ, Smith HC. The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. Nucleic Acids Res. 1996. 24:478-85
- 78 Cappione A. J., French B. L., Skuse G. R. A potential role for NF1 mRNA editing in the pathogenesis of NF1 tumors. Am. J. Hum. Genet. 1997. 60: 305-312
- 79 Nordlund M, Gu X, Shipley MT and Ratner N. Neurofibromin is enriched in the endoplasmic reticulum of CNS neurons. J Neurosci. 1993. 13(4):1588-1600
- 80 Gutmann DH, Geist RT, Rose K and Wright DE. Expression of two new protein isoforms of the neurofibromatosis type 1 gene product, neurofibromin, in muscle tissues. Dev Dyn. 1995. 202(3):302-311
- 81 Mangoura D, Sun Y, Li C, Singh D, Gutmann DH, Flores A, et al Phosphorylation of neurofibromin by PKC is a possible molecular switch in EGF receptor signaling in neural cells. Oncogene. 2006. 25(5):735-745
- 82 Mattocks C, Baralle D, Tarpey P, French-Constant C, Bobrow M and Whittaker J. Automated comparative sequence analysis identifies mutations in 89% of NF1 patients and confirms a mutation cluster in exons 11-17 distinct from the GAP related domain. J Med Genet. 2004. 41(4)
- 83 Boyanapalli M, Lahoud OB, Messiaen L, Kim B, Anderle de Saylor MS, Duckett SJ, et al. Neurofibromin binds to caveolin-1 and regulates ras, FAK, and Akt. Biochem Biophys Res Commun. 2006. 340(4):1200-1208
- 84 Dasgupta B, Dugan LL and Gutmann DH. The neurofibromatosis 1 gene product neurofibromin regulates pituitary adenylate cyclase-activating polypeptide-mediated signaling in astrocytes. J Neurosci. 2003. 23(26):8949-8954

-
- ⁸⁵ Ozawa T, Araki N, Yunoue S, Tokuo H, Feng L, Patrakitkomjorn S et al. The neurofibromatosis type 1 gene product neurofibromin enhances cell motility by regulating actin filament dynamics via the Rho-ROCK-LIMK2-cofilin pathway. *J Biol Chem.* 2005. 280(47):39524-39533
- ⁸⁶ Kweh F, Zheng M, Kurenova E, Wallace M, Golubovskaya V and Cance WG. Neurofibromin physically interacts with the N-terminal domain of focal adhesion kinase. *Mol Carcinog.* 2009. 48(11):1005-1017
- ⁸⁷ Arima Y, Hayashi H, Kamata K, Goto TM, Sasaki M, Kuramochi A et al. Decreased expression of neurofibromin contributes to epithelial-mesenchymal transition in neurofibromatosis type 1. *Exp Dermatol.* 2010. 19(8):136-141
- ⁸⁸ Peltonen J, Peltonen S: Composition of neurofibromas, NF1 expression, and comparison of normal and NF1 haploinsufficient cells. *Neurofibromatoses.* 2008. 129 – 142
- ⁸⁹ D. P. Bartel, *MicroRNAs: Target Recognition and Regulatory Functions.* Cell. 2009. 136(2): 215–233
- ⁹⁰ Chai G, Liu N, Ma J, Li H, Oblinger J, Prahalad A et al. MicroRNA-10b regulates tumorigenesis in neurofibromatosis type 1. *Cancer Sci.* 2010. 101:1997–2004
- ⁹¹ Ji Y, Wei Y, Wang J, Gong K, Zhang Y, Zuo H. Correlation of microRNA-10b upregulation and poor prognosis in human gliomas. *Tumour Biol.* 2015. 36:6249–6254
- ⁹² Lenarduzzi M, Hui ABY, Alajez NM, Shi W, Williams J, Yue S et al. MicroRNA-193b enhances tumor progression via down regulation of neurofibromin 1. *PLoS ONE.* 2013. 8(1)
- ⁹³ Phillips, T. The role of methylation in gene expression. *Nature Education.* 2008. 1(1):116
- ⁹⁴ Fishbein L, Eady B, Sanek N, Muir D, Wallace M: Analysis of somatic NF1 promoter methylation in plexiform neurofibromas and Schwann cells. *Cancer Genet Cytogenet.* 2005. 157:181–186
- ⁹⁵ Gutmann DH, McLellan MD, Hussain I, Wallis JW, Fulton LL, Fulton RS et al. Somatic neurofibromatosis type 1 (NF1) inactivation characterizes NF1-associated pilocytic astrocytoma. *Genome Res.* 2013. 23(3):431–439
- ⁹⁶ Hollstein PE, Cichowski K. Identifying the Ubiquitin Ligase complex that regulates the NF1 tumor suppressor and Ras. *Cancer Discov.* 2013. 3(8):880–893
- ⁹⁷ McGillicuddy LT, Fromm JA, Hollstein PE, Kubek S, Beroukhi R, De Raedt T et al. Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer cell.* 2009. 16(1):44–54
- ⁹⁸ Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet.* 2007. 8(8):619-31
- ⁹⁹ Huson SM, Compston DA, Clark P, Harper PS. A genetic study of von Recklinghausen neurofibromatosis in south east Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J Med Genet.* 1989. 26:704-711
- ¹⁰⁰ Jadayel D, Fain P, Upadhyaya M, Ponder M, Huson S, Carey J, et al. Paternal origin of new mutations in von Recklinghausen neurofibromatosis. *Nature.* 1990. 343:558 –559

-
- ¹⁰¹ Upadhyaya M, Ruggieri M, Maynard J, Osborn M, Hartog C, Mudd S et al: Gross deletions of the neurofibromatosis type 1 (NF1) gene are predominantly of maternal origin and commonly associated with a learning disability, dysmorphic features and developmental delay. *Hum Genet.* 1998. 102:591–597
- ¹⁰² Klose, A. et al. Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1. *Hum. Mol. Genet.* 1998. 7, 1261–1268
- ¹⁰³ Kluwe L, Siebert R, Gesk S, Friedrich RE, Tinschert S, Kehrer-Sawatzki H et al. Screening 500 unselected neurofibromatosis 1 patients for deletions of the NF1 gene. *Hum Mutat.* 2004. 23:111–116
- ¹⁰⁴ De Raedt T, Brems H, Wolkenstein P, Vidaud D, Pilotti S, Perrone F et al. Elevated risk for MPNST in NF1 microdeletion patients. *Am J Hum Genet.* 2003. 72:1288–1292
- ¹⁰⁵ Pasmant E, Sabbagh A, Spurlock G, Laurendeau I, Grillo E, Hamel MJ et al. NF1 microdeletions in neurofibromatosis type 1: from genotype to phenotype. *Hum Mutat.* 2010. 31:1506–1518
- ¹⁰⁶ Vogt J, Nguyen R, Kluwe L, et al. Delineation of the clinical phenotype associated with non-mosaic type-2 NF1 deletions: two case reports. *Journal of Medical Case Reports.* 2011. 5:577
- ¹⁰⁷ Messiaen L, Vogt J, Bengesser K, Fu C, Mikhail F, Serra R et al. Mosaic type-1 NF1 microdeletions as a cause of both generalized and segmental neurofibromatosis type-1 (NF1). *Hum Mutat.* 2011. 32:213–219
- ¹⁰⁸ Ko, J. M., Sohn, Y. B., Jeong, S. Y., Kim, H. J. & Messiaen, L. M. Mutation spectrum of NF1 and clinical characteristics in 78 Korean patients with neurofibromatosis type 1. *Pediatr. Neurol.* 2013. 48, 447–453
- ¹⁰⁹ Alkindy, A., Chuzhanova, N., Kini, U., Cooper, D. N. & Upadhyaya, M. Genotype–phenotype associations in neurofibromatosis type 1 (NF1): an increased risk of tumor complications in patients with NF1 splice-site mutations? *Hum. Genomics.* 2012. 6, 12
- ¹¹⁰ De Raedt, T. et al. Elevated risk for MPNST in NF1 microdeletion patients. *Am. J. Hum. Genet.* 2003.72, 1288–1292
- ¹¹¹ Upadhyaya, M. et al. An absence of cutaneous neurofibromas associated with a 3-bp inframe deletion in exon 17 of the NF1 gene (c.2970–2972 delAAT): evidence of a clinically significant NF1 genotype–phenotype correlation. *Am. J. Hum. Genet.* 2007. 80, 140–151
- ¹¹² Santoro C, Maietta A, Giugliano T, Melis D, Perrotta S, Nigro V et al. Argsubstitution in neurofibromin: further evidence of a genotype-phenotype correlation in neurofibromatosis type 1. *Eur J Hum Genet.* 2015.
- ¹¹³ Rojnueangnit, K et al. High Incidence of Noonan Syndrome Features Including Short Stature and Pulmonic Stenosis in Patients carrying NF1 Missense Mutations Affecting p.Arg1809: Genotype–Phenotype Correlation. *Hum. Mutat.* 2015. 36: 1052–1063
- ¹¹⁴ Rieley, M. B. et al. Variable expression of neurofibromatosis 1 in monozygotic twins. *Am. J. Med. Genet.* 2011. 155: 478–485

-
- 115 Zhang, M. et al. Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors. *Nature Genet.* 2014. 46, 1170–1172
- 116 Saal, H. M. et al. Racial differences in the prevalence of optic nerve gliomas in neurofibromatosis type 1. *Am. J. Hum. Genet.* 1995. 57:54
- 117 Diggs-Andrews KA, Brown JA, Gianino SM, Rubin JB, Wozniak DF, et al. Sex Is a major determinant of neuronal dysfunction in neurofibromatosis type 1. *Ann Neurol.* 2014. 75: 309–316
- 118 Ling, J. Q. et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science.* 2006. 312: 269–272
- 119 Messiaen, L. M. et al. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.* 2000.15 : 541–555
- 120 Sabbagh, A., E. Pasmant, A. Imbard, A. Luscan, M. Soares, H. Blanche, et al. NF1 molecular characterization and neurofibromatosis type I genotype-phenotype correlation: the French experience. *Hum. Mutat.* 2013. 34:1510–1518
- 121 Schouten JP, McElgunn CJ, Waaijer R, Zwiijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research.* 2002. 30(12):57
- 122 De Luca A, Bottillo I, Dasdia MC, et al. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. *Journal of Medical Genetics.* 2007. 44(12):800-808
- 123 Sakharkar MK, Chow VTK, Kanguene P. Distributions of exons and introns in the human genome. *In Silico Biol.* 2004. 4:387–93
- 124 Yu B, Sawyer NA, Chiu C, Oefner PJ, Underhill PA. DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Curr Protoc Hum Genet.* 2006. 7:710
- 125 Han SS., Cooper DN., Upadhyaya MN. Evaluation of denaturing high performance liquid chromatography (DHPLC) for the mutational analysis of the neurofibromatosis type 1 (NF1) gene. *Hum Genet.* 2001. 109:487–497
- 126 Valero MC., Martín Y., Hernández-Imaz E, et al. A highly sensitive genetic protocol to detect NF1 mutations. *J Mol Diagn.* 2011. 13:113–122
- 127 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America.* 1977. 74(12):5463-5467
- 128 Prober, J., Trainor, G., Dam, R., Hobbs, F., Robertson, C., Zagursky, R. et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science.* 1987. 238: 336-341
- 129 Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R. et al. Fluorescence detection in automated DNA sequence analysis. *Nature.* 1986. 321:674–679
- 130 Ewing, B., Hillier, L., Wendl, M.C., and Green, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 1998. 8: 175-185

-
- 131 Ewing, B. and Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* 1998. 8: 186-194
- 132 Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mole Biol.* 1981. 147(1):195-7
- 133 Jamuar S. S., Walsh C. A. Somatic mutations in cerebral cortical malformations. *N. Engl. J. Med.* 2014. 371:2037-2038
- 134 Wakioka T, Sasaki A, Kato R, Shouda T, Matsumoto A, Miyoshi K et al. Spred is a Sprouty-related suppressor of Ras signalling. *Nature.* 2001. 412: 647-651
- 135 Yoshimura A. Regulation of cytokine signaling by the SOCS and Spred family proteins. *Keio J Med.* 2009. 58: 73-83
- 136 Phoenix TN, Temple S. Spred1, a negative regulator of RasMAPK-ERK, is enriched in CNS germinal zones, dampens NSC proliferation, and maintains ventricular zone structure. *Genes Dev.* 2010. 24: 45-56
- 137 Shmueli O, Horn-Saban S, Chalifa-Caspi V, Shmoish M, Ophir R, Benjamin-Rodrig H, GeneNote: whole genome expression profiles in normal human tissues. *C R Biol.* 2003. 326: 1067-1072
- 138 Johne C, Matenia D, Li XY, Timm T, Balusamy K, Mandelkow EM. Spred1 and TESK1 – two new interaction partners of the kinase MARKK/TAO1 that link the microtubule and actin cytoskeleton. *Mol Biol Cell.* 2008. 19:1391-1403
- 139 Zhuang L, Villiger P, Trueb B. Interaction of the receptor FGFR1 with the negative regulator Spred1. *Cell Signal.* 2011. 23: 1496-1504
- 140 Li D, Jackson RA, Yusoff P, Guy GR. Direct association of Sprouty-related protein with an EVH1 domain (SPRED) 1 or SPRED2 with DYRK1A modifies substrate/kinase interactions. *J Biol Chem.* 2010. 285: 35374-35385
- 141 Quintanar-Audelo M, Yusoff P, Sinniah S, Chandramouli S, Guy GR. Sprouty-related Ena/vasodilator-stimulated phosphoprotein homology 1-domain-containing protein (SPRED1), a tyrosineprotein phosphatase non-receptor type 11 (SHP2) substrate in the Ras/extracellular signal-regulated kinase (ERK) pathway. *J Biol Chem* 2011. 286: 23102-23112
- 142 Stowe IB, Mercado EL, Stowe TR, Bell EL, Oses-Prieto JA, Hernandez H et al. A shared molecular mechanism underlies the human rasopathies Legius syndrome and neurofibromatosis-1. *Genes Dev.* 2012. 26: 1421-1426
- 143 Hirata Y, Brems H, Suzuki M, Kanamori M, Okada M, Morita R et al. Interaction between a domain of a negative regulator of the RAS-ERK pathway, SPRED1, and the GTPase-Activating Protein-Related Domain of neurofibromin is implicated in Legius Syndrome and Neurofibromatosis Type 1. *J Biol Chem.* 2015.
- 144 Yoshida, T., Hisamoto, T., Akiba, J., Koga, H., Nakamura, K., Tokunaga, et al. Spreds, inhibitors of the Ras/ERK signal transduction, are dysregulated in human hepatocellular carcinoma and linked to the malignant phenotype of tumors. *Oncogene.* 2006. 25:6056-6066

-
- ¹⁴⁵ Miyoshi, K., Wakioka, T., Nishinakamura, H., Kamio, M., Yang, L., Inoue, M. et al. The Sprouty-related protein, Spred, inhibits cell motility, metastasis, and Rho-mediated actin reorganization. *Oncogene*. 2004. 23:5567- 5576
- ¹⁴⁶ Maertens O, De Schepper S, Vandesompele J, Brems H, Heyns I, Janssens S et al. Molecular dissection of isolated disease features in mosaic neurofibromatosis type 1. *Am J Hum Genet*. 2007. 81: 243–251
- ¹⁴⁷ Brems H and Legius E. Legius Syndrome, an Update and quantitative PCR. *Am J Med Genet A*. 2011. 155: 1352–1359
- ¹⁴⁸ Brems H, Pasmant E, Van Minkelen R, Wimmer K, Upadhyaya M, Legius E et al. Review and update of SPRED1 mutations causing Legius syndrome. *Hum Mutat*. 2012. 33: 1538–1546
- ¹⁴⁹ Messiaen L, Yao S, Brems H, Callens T, Sathienkijkanchai A, Denayer E et al. Clinical and mutational spectrum of neurofibromatosis type 1-like syndrome. *JAMA*. 2009. 302:2111–2118
- ¹⁵⁰ E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin et al. Initial sequencing and analysis of the human genome. *Nature*. 2001. 409:860–921
- ¹⁵¹ J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton et al. The sequence of the human genome. *Science*. 2001. 291:1304–1351
- ¹⁵² International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature*. 2004. 431:931–945
- ¹⁵³ M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005. 437: 376–380
- ¹⁵⁴ T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky et al. Single-molecule DNA sequencing of a viral genome. *Science*. 2008. 320: 106–109
- ¹⁵⁵ A. Valouev, J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res*. 2008. 18:1051–1063
- ¹⁵⁶ D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008. 456:53–59
- ¹⁵⁷ M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005. 437:376–380
- ¹⁵⁸ A. Mellmann, D. Harmsen, C.A. Cummings, E.B. Zentz, S.R. Leopold, A. Rico, K. et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE*. 2011. 6:22751
- ¹⁵⁹ J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011. 475:348–352

-
- ¹⁶⁰ L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong et al. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012
- ¹⁶¹ M.O. Carneiro, C. Russ, M.G. Ross, S.B. Gabriel, C. Nusbaum, M.A. DePristo. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics.* 2012. 13:375
- ¹⁶² E.W. Loomis, J.S. Eid, P. Peluso, J. Yin, L. Hickey, D. Rank, S et al. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome Res.* 2013. 23:121–128
- ¹⁶³ A.C. English, S. Richards, Y. Han, M. Wang, V. Vee, J. Qu et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE.* 2012. 7:47768
- ¹⁶⁴ M.J.P. Chaisson, J. Huddleston, M.Y. Dennis, P.H. Sudmant, M. Malig, F. Hormozdiari et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature.* 2015. 517: 608–611
- ¹⁶⁵ D. Sharon, H. Tilgner, F. Grubert, M. Snyder. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* 2013. 31:1009–1014
- ¹⁶⁶ B.A. Flusberg, D.R. Webster, J.H. Lee, K.J. Travers, E.C. Olivares, T.A. Clark, J et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods.* 2010. 7:461–465
- ¹⁶⁷ Wang Y, Yang Q, Wang Z. The evolution of nanopore sequencing. *Frontiers in Genetics.* 2014. 5:449
- ¹⁶⁸ Kasianowicz J. J., Brandin E., Branton D., Deamer D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci.* 1996. 93:13770–13773
- ¹⁶⁹ Clarke J, et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol.* 2009. 4(4):265–270
- ¹⁷⁰ T. Laver, J. Harrison, D.J. Studholme, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quant.* 2015. 3:1–8
- ¹⁷¹ Derek Stein. Nanopore Sequencing: Forcing Improved Resolution. *Biophysical journal.* 2015
- ¹⁷² M. Jain, I.T. Fiddes, K.H. Miga, H.E. Olsen, B. Paten, M. Akeson. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods.* 2015. 12:351–356
- ¹⁷³ Datta S, Budhaliya R, Das B, Chatterjee S, Vanlalhmuaaka, Veer V. Next-generation sequencing in clinical virology: Discovery of new viruses. *World Journal of Virology.* 2015. 4(3):265-276
- ¹⁷⁴ Sowmiya Moorthie PhD, Alison Hall MA, Caroline F. Wright PhD. Informatics and clinical genome sequencing: opening the black box. *Genetics in Medicine.* 2013. 15:165–171
- ¹⁷⁵ Li, H, Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009. 25(14):1754-60
- ¹⁷⁶ Li, H, Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010. 26(5):589-95

-
- ¹⁷⁷ Ning, Z, Cox, AJ, Mullikin, JC. SSAHA: a fast search method for large DNA databases. *Genome Res.* 2001. 11(10):1725-9
- ¹⁷⁸ Li, Heng . Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics.* 2012. 28(14):1838-1844
- ¹⁷⁹ Smith, T. F. and Waterman, M. S. Identification of common molecular subsequences. *Journal of Molecular Biology.* 1981. 147(1):195-197
- ¹⁸⁰ De Leeneer K, De Schrijver J, Clement L, et al. Practical Tools to Implement Massive Parallel Pyrosequencing of PCR Products in Next Generation Molecular Diagnostics. *PLoS ONE.* 2011. 6(9):25531
- ¹⁸¹ Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics.* 2011. 27(15):2156-2158
- ¹⁸² Kent WJ., Sugnet CW., Furey TS., Roskin KM., Pringle TH., Zahler AM et al. The human genome browser at UCSC. *Genome Research.* 2002. 12(6) : 996-1006
- ¹⁸³ A.Janin. Mise au point d'une technique de séquençage à haut débit pour l'étude des gènes du facteur VII et du facteur IX. Thèse 13-2015. Lyon 1. 2015
- ¹⁸⁴ Sherry ST, Ward M-H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research.* 2001. 29(1):308-311
- ¹⁸⁵ Ryo Maruoka, Toshiki Takenouchi, Chiharu Torii, Atsushi Shimizu, Kumiko Misu, Koichiro Higasa et al. The use of next-generation sequencing in molecular diagnosis of neurofibromatosis type 1: a validation study. *Genet Test Mol Biomarkers.* 2014. 18(11):722-35

L'ISPB-Faculté de Pharmacie de Lyon et l'Université Claude Bernard Lyon 1 n'entendent donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions sont considérées comme propres à leurs auteurs

L'ISPB-Faculté de Pharmacie de Lyon est engagé dans une démarche de lutte contre le plagiat. De ce fait une sensibilisation des étudiants et encadrants des thèses a été réalisée avec notamment l'incitation à l'utilisation de méthodes de recherche de similitudes

BUISSON Adrien

Mise au point d'une technique de séquençage haut débit par technologie Ion Torrent™ pour l'étude des gènes *NF1* et *SPRED1*.

Th. D. Pharm., Lyon 1, 2016, 300p.

RESUME

La neurofibromatose de type 1 (NF1), ou maladie de Von Recklinghausen est l'une des maladies génétiques les plus fréquentes. Le déficit en neurofibromine en est la conséquence et elle est due à une atteinte d'un gène suppresseur de tumeur : *NF1*.

Le syndrome de Legius, découvert en 2007, se caractérise principalement par la présence de signes cutanés similaires à ceux de la NF1. Le gène mis en cause est le gène *SPRED1*.

Actuellement, la stratégie du laboratoire repose sur une analyse séquentielle de ces 2 gènes. L'analyse du gène *NF1* est composée d'une recherche de grands réarrangements par technique M.L.P.A et d'un criblage par dHPLC des mutations, confirmées par séquençage Sanger. Pour le gène *SPRED1*, séquençage systématique des exons puis une recherche par M.L.P.A sont réalisés. Cette méthode reste longue et coûteuse, mais elle permet d'atteindre une sensibilité proche des 95%.

Notre laboratoire a mis au point deux designs de banque NGS grâce à l'outil Ion Ampliseq™ Designer. Le séquençage a été effectué par un automate Ion PGM™, qui utilise la technologie par semi-conducteur Ion Torrent™ (Life Technologies®). Cette méthode nous permet de réaliser l'analyse simultanée de tous les exons des gènes *NF1* et *SPRED1* ainsi que leurs régions 5' et 3' non traduites.

Afin de valider notre étude, les variants de patients précédemment génotypés ont été comparés aux données récoltées par NGS. L'analyse de séries de nouveaux patients ont été réalisées par séquençage haut débit et en parallèle par les techniques standards. Les résultats démontrent une bonne sensibilité de la méthode, avec la mise en évidence de la majorité des variants polymorphiques ou causals.

Cette nouvelle approche permet d'améliorer le temps de rendu du résultat et le coût de l'analyse. L'objectif de notre travail est de remplacer à terme la méthode de criblage actuelle par la méthode NGS.

MOTS CLES Neurofibromatose, Syndrome de Legius
Gène *NF1*, Gène *SPRED1*
Séquençage haut débit, NGS
Ampliseq, Ion Torrent

JURY Mme. VINCIGUERRA Christine, Professeur en Pharmacie
M. CALENDER Alain, Professeur en Médecine
Mme. BARDEL Claire, Maître de Conférences
M. PINSON Stéphane, Docteur en Médecine

DATE DE SOUTENANCE

Vendredi 12 Février 2016

ADRESSE DE L'AUTEUR

285, Rue Jacques Brel – 69400 Gleizé

