UNIVERSITE DES ANTILLES

THESE DE SCIENCES EXPERIMENTALES

Pour obtenir le grade de DOCTEUR

Spécialité : Aspects moléculaires et cellulaires de la biologie

Préparée avec l'équipe de recherche BIOSPHERES, et le Centre Hospitalier

Universitaire de Martinique

Présentée par Anna-Gaëlle Giguet-Valard

Étude démographique et analyse CoDESeq des maladies neurodégénératives à expression motrice en Martinique

Soutenance prévue le 17 décembre 2020 devant le jury composé de :

Pr Moustapha DRAME	Président du jury	Université des Antilles
Pr Gaëtan LESCA	Rapporteur	Université de Lyon
Pr Mathieu ANHEIM	Rapporteur	Université de Strasbourg
Pr Juliette SMITH-RAVIN	Directeur de thèse	Université des Antilles
Pr Cyril GOIZET	Co-directeur de thèse	Université de Bordeaux
Dr Rémi BELLANCE	Membre invité	CHU de Martinique

UNIVERSITE DES ANTILLES

THESE DE SCIENCES EXPERIMENTALES

Pour obtenir le grade de DOCTEUR

Spécialité : Aspects moléculaires et cellulaires de la biologie



Étude démographique et analyse CoDEseq des maladies neurodégénératives à expression motrice en Martinique

Préparée avec l'équipe de recherche BIOSPHERES, et le Centre Hospitalier Universitaire de Martinique

Présentée par Anna-Gaëlle Giguet-Valard

Soutenance prévue le 17 décembre 2020 devant le jury composé de :

Pr Moustapha DRAME	Président du jury	Université des Antilles
Pr Gaëtan LESCA	Rapporteur	Université de Lyon
Pr Mathieu ANHEIM	Rapporteur	Université de Strasbourg
Pr Juliette SMITH-RAVIN	Directeur de thèse	Université des Antilles
Pr Cyril GOIZET	Co-directeur de thèse	Université de Bordeaux
Dr Rémi BELLANCE	Membre invité	CHU de Martinique

Résumé

Les Maladies Neurodégénératives Rares à Expression Motrice (MNDREM) sont un groupe très hétérogène de pathologies neurodégénératives de physiopathologie acquise ou innée, pouvant se déclarer à tout âge. Elles impactent de façon ciblée, les structures neurologiques centrales ou périphériques impliquées dans l'activité motrice. Ces structures sont fonctionnellement et topographiquement proches des centres de cognition. C'est ainsi que dans certaines conditions, la neurodégénérescence des structures motrices impacte les fonctions cognitives. Elles se traduisent par des troubles de la marche et /ou de l'équilibre, une diminution ou une absence de mouvement, une atteinte cognitive variée pouvant aller jusqu'à la démence.

Les MNDREM sont sporadiques ou familiales. Ces dernières sont donc à composante héréditaire et représentent des modèles d'étude intéressants car elles offrent la possibilité d'analyser des ADN d'apparentés et d'augmenter de ce fait la puissance de recherche de facteurs génétiques impliqués dans une symptomatologie familiale. La littérature scientifique indique un nombre important de variants génétiques pathogènes responsables des MNDREM, ce qui aide la classification mais la complexifie également puisqu'on constate qu'un même gène peut être impliqué dans plusieurs pathologies et une pathologie peut être causée par différents gènes. Par conséquent, dans les MNDREM, on observe des symptomatologies atypiques, des chevauchements ou « overlap » cliniques des maladies alléliques.

Aux Antilles françaises, les données de la littérature en matière de MNDREM sont limitées mais riches de certaines observations confortant les atypies comme par exemple le syndrome parkinsonien atypique des Caraïbes. Sur le plan génétique, les investigations diagnostiques sont restreintes aux filières médicales basées en métropole et sur la comparaison de données génomiques de populations le plus souvent caucasiennes.

Le travail de thèse présenté s'articule autour de 2 aspects : le premier, clinique et génétique, a permis de dresser un état des lieux des MNDREM en Martinique en les caractérisant sur ces plans, le deuxième sur la recherche de variants causaux par une technique de Séquençage Haut Débit (NGS) de 2e génération nommée CoDEseq (Copy number variation Detection and Exome sequencing). L'état des lieux des MNDREM en Martinique, et le travail expérimental d'analyse de données de NGS, sont les premiers travaux de ce type dans le domaine et dans notre région.

La méthode CoDEseq est innovante car elle autorise l'analyse exome entier à la recherche de Variants Simples Nuclotidiques (SNVs) ou Variants de Nombre de Copies (CNVs). Elle n'avait presque jamais été utilisée dans l'analyse des MNDREM. Nous l'avons employée car c'est une méthode de choix de recherche à la fois de variants simples et de variants de structures nouveaux. Les résultats montrent une rentabilité diagnostic de 58% puisque nous avons identifié un variant probablement pathogène chez 7 patients sur 12 testés. Nous avons trouvé ces variants dans des gènes connus de MNDREM parfois décrits dans des populations asiatiques, mais aussi, d'ascendance africaine ou caucasienne. Certains de ces gènes peuvent être impliqués dans plusieurs symptomatologies, confortant le constat de chevauchement clinique dans ces maladies.

Ce travail jette les bases épidémiologiques des MNDREM aux Antilles et propose un socle de registre en la matière. Sur le plan expérimental, il a permis de proposer une étiologie moleculaire impliquant des variants préalablement décrit ou nouveaux. Il est également une preuve de concept quant aux moyens bioinformatiques d'analyse de données de NGS en Martinique. Il ouvre la voie vers d'autres travaux du même genre, susceptibles de dresser les spécificités géniques des MNDREM de notre région et étoffer les données de la littérature scientifiques et médicales.

Abstract

Rare Neurodegenerative Diseases with Motor Expression (RNDME) are a very heterogeneous group of neurodegenerative pathologies of acquired or innate physiopathology that can occur at any age. They have a targeted impact on the central or peripheral neurological structures involved in motor activity. These structures are functionally and topographically close to the centers of cognition. Thus, under certain conditions, the neurodegeneration of motor structures impacts cognitive functions. They result in gait and/or balance disorders, a decrease or absence of movement, a varied cognitive impairment that can go as far as dementia.

RNDME are sporadic or family based. The latter have a hereditary component and therefore represent interesting study models because they offer the possibility of analyzing the DNA of relatives and thus increase the research power of genetic factors involved in a family symptomatology.The scientific literature indicates an important number of genetic pathogenic variation responsible for RNDMEs, which allows classification but also makes it more complex because it can be observed that the same gene can be involved in several pathologies and a pathology can be caused by different genes. Consequently, in RNDMEs, atypical symptoms, clinical overlaps or allelic diseases are observed.

In the French West Indies, the data in the literature on RNDMEs are limited but rich in certain observations confirming the atypical RNDMEs seach as the atypical Caribbean Parkinson's syndrome. On the genetic level, diagnostic investigations are limited to medical fields based in metropolitan France and most often about the comparison of genomic data of Caucasian populations.

The thesis work is about 2 aspects: the first, clinical and genetic, has allowed to draw up an inventory of the RNDMEs in Martinique by characterizing them on these plans**,** the second on the search for causal variants by a Next Generation Sequencing (NGS) technique called CoDEseq (Copy number variation Detection and Exome sequencing). The inventory of RNDMEs in Martinique and the experimental work of data analysis of NGS, are the first works of this type, in the field and in our region.

The CoDEseq method is innovative because it allows whole exome analysis in the detection of Single Nucleodis Variants (SNVs) and Copy Number Variants (CNVs). It had almost never been used in the analysis of RNDMEs. We used it because it is a method of choice for searching for both simple and structural variants. The results show a diagnostic cost-effectiveness of 58% since we identified a variant probably pathogenic in 7 patients out of 12 tested. We found these variants in known RNDMEs genes sometimes described in Asian populations, but also of African or Caucasian descent. Some of these genes may be involved in several symptomatologies, confirming the finding of overlap in these diseases.

This work lays the epidemiological basis for RNDMEs in the West Indies and provides a basis for a registry in this area. On the experimental level, it allow to propose a molecular cause associated with previously described or new variations. It is also a proof of concept regarding the bioinformatics means of analyzing NGS data in Martinique. It paves the way for other work of the same kind, susceptible to draw up the gene specificities of the RNDMEs in our region and to expand the data in the scientific and medical literature.

A mon indéfectible famille…
Mon dévoué mari, mes remarquables enfants,
Ma précieuse mère et ma talentueuse sœur,
À chacun de mes nombreux et uniques proches.

SOMMAIRE

Sommaire

Résumé **1**

Remerciements **10**

Préambule **13**

CHAPITRE I - État de l'art **23**

1. Dégénérescence neurologique à expression motrice *25*
1.1. Principaux mécanismes moléculaires de dégénérescence cellulaire 25
1.2. Organisation anatomique et cellulaire du système locomoteur 35
1.3. Manifestations cliniques de dégénérescence neurologique 42
2. Diagnostic des MND Rares à Expression Motrice (MNDREM) *49*
2.1. Caractérisation phénotypique 49
2.2. Analyses fonctionnelles et imageries de référence 51
2.3. Dosages enzymo-biologiques récurrents 52
2.4. Tests génétiques 54
3. MND Rares à Expression Motrice, MNDREM d'intérêt *57*
3.1. Les ataxies héréditaires 57
3.2. Les maladies de Huntington et Huntington-like 64
3.3. La Sclérose Latérale Amyotrophique 69
3.4. Les Syndromes parkinsoniens 76
4. Chevauchement clinique et moléculaire *84*
4.1. Atteinte de la corne antérieure et troubles cognitifs 84
4.2. Troubles cognitifs, syndrome parkinsonien et *ATXN2* 85
4.3. Accumulations protéiques et atteintes mitochondriales 86
4.4. Autres chevauchements de MNDREM 87
5. Quid du continuum clinique et moléculaire dans les MNDREM *88*

CHAPITRE II – Conditions de mise en œuvre **95**

1. Collecte de données cliniques, biologiques et génétiques *97*
1.1. Critères d'inclusion et d'exclusion 97
1.2. Cadre législatif et éthique 99
1.3. Système d'encodage des données collectées 100
2. Généralités sur le séquençage 2e génération *103*
2.1. La technique 103
2.2. La bioinformatique 112
3. Rappels sur l'interprétation de variants *119*
3.1. Bases de données 119
3.2. Logiciels de prédiction 121
3.3. Étude de la ségrégation familiale 125
3.4. Classification des variants 126
4. Détection de variants SNVs et CNVs par CoDE-seq *129*
4.1. Technique de séquençage 129
4.2. Dispositif informatique 131
5. Vers la recherche de SNVs et CNVs dans les MNDREM *138*

CHAPITRE III - Étude démographique et génomique des MNDREM ... 144

1. Population générale de MNDREM ... 146
1.1. Données descriptives ... 146
1.2. MNDREM génétiques : présentations cliniques ... 150
1.3. MNDREM sans étiologie génétique ... 161
2. Présentation clinique de l'échantillon sélectionné ... 165
2.1. Cas de Sclérose Latérale Amyotrophique (SLA) ... 165
2.2. Cas d'ataxies spinocérébelleuses (CA) ... 168
2.3. Cas de syndromes Parkinsoniens (PARK/SPARK) ... 170
3. Analyse génomique de l'échantillon ... 174
3.1. Analyse qualité technique ... 174
3.2. Gènes candidats dans les phénotypes PARK/SPARK ... 174
3.3. Gènes candidats dans les Ataxies Cérébelleuses ... 180
3.4. Gènes candidats dans la SLA ... 183
4. Données démographiques et variants génomiques nouveaux ... 186

CHAPITRE IV – Etude computationnelle préliminaire des interactions géniques ... 212

1. Rappels sur les Interactions Protéine-Protéine (IPPs) ... 214
1.1. Définitions ... 214
1.2. Détection expérimentale ... 215
1.3. Prédiction d'IPPs ... 216
2. Modélisation et analyse de réseaux ... 220
2.1. Outils de détection et de modélisation ... 220
2.2. Paramètres de base ... 222
2.3. Élimination des Faux Positifs ... 222
3. Étude computationnelle ... 223
3.1. Rationnel de l'étude ... 223
3.2. Design ... 224
3.3. Résultats ... 230
4. Interactions géniques et chevauchement clinique ? ... 239

CHAPITRE V - Discussion générale ... 242

Conclusion et perspectives ... 257

BIBLIOGRAPHIE ... 274

ANNEXES ... 286

LISTES DES ILLUSTRATIONS

LISTE DES TABLEAUX

Tableau 1. Organisation des cellules de l'encéphale..... 40
Tableau 2. Principaux neurotransmetteurs du SNC..... 41
Tableau 3. Classification des troubles de la marche 43
Tableau 4. Examen des réflexes ostéotendineux et cutanés plantaires..... 46
Tableau 5. Génétique des ataxies spinocérébelleuse DOMINANTES 62
Tableau 6. Génétique des ataxies spinocérébelleuses RECESSIVES..... 63
Tableau 7. Expansion de motifs répétés dans les maladies de Huntington..... 68
Tableau 8. Principaux gènes étudiés dans le SLA 75
Tableau 9. Sémiologie des SPARK 77
Tableau 10. Gènes impliqués dans le diagnostic de MPI et SPARK..... 83
Tableau 11. Outils de prédictions utilisés..... 137
Tableau 12. Clinique de la famille SLA anticipation..... 167
Tableau 13. Caratéristiques cliniques des PARK/SPARK testés 179
Tableau 14. Caracteristiques cliniques et moleculaires des CA testés 182
Tableau 15. Caractéristiques cliniques et moléculaires des SLA testés..... 185
Tableau 16. Liste et fonctions des gènes de notre population..... 226
Tableau 17. Processus cellulaires associés aux gènes de MNDREM 234
Tableau 18. Liste des interactants potentiels avec *JPH3*..... 236
Tableau 19. Nature des interactions et base de données..... 238

LISTE DES FIGURES

Figure 1. Types de Singles Nucleotids Polymorphisms du génome humain 26
Figure 2. Voie de signalisation et d'activation des caspases 29
Figure 3. Processus d'Ubiquitinylation/dégradation 31
Figure 4. Schéma de la production d'espèces réactives mitochondriales 35
Figure 5. Principales structures de la marche et du mouvement 36
Figure 6. Organisation de la transmission du signal moteur 36
Figure 7. Territorialisation des vertèbres du rachis 39
Figure 8. Schéma représentant l'activité mitochondriale 53
Figure 9. Schéma du principe de CGH 56
Figure 10. Représentation en éventail de l'overlap 89
Figure 11. Présentation de la méthodologie 97
Figure 12. Schéma représentatif de l'échantillonnage 98
Figure 13. Encodage de la base de données des MNDREM 100
Figure 14. Pedigree type en fonction des modes de transmission 101
Figure 15. Pedigree type transmission mitochondriale 102
Figure 16. Méthodes d'enrichissement 104
Figure 17. Méthodes d'amplification clonale 106
Figure 18. Comparaison des techniques d'enrichissement 107
Figure 19. Comparaison des séquenceurs 110
Figure 20. Représentation visuelle des notions de SHD 111
Figure 21. Principe de détection des CNVs par XHMM 116
Figure 22. Schéma du procédé de détection des SNVs 118
Figure 23. Classifications des variants ACMG 127
Figure 24, Algorithme de classification ACMG 128
Figure 25. Schéma de la méthode d'extraction d'ADN employée 129
Figure 26. Répartition des MNDREM 146
Figure 27. Antécédents familiaux dans les MNDREM 147
Figure 28. Répartition des modes de transmission 148
Figure 29. Répartition des modes de transmission par pathologies 149
Figure 30. Proportion de variant causal identifiée par phénotype 150

Figure 31. Répartition des phénotypes avec ou sans étiologie 163
Figure 32. Pedigree du cas 17S353 165
Figure 33. Pedigree SLA familiale avec anticipation 167
Figure 34. Pedigree des cas 18S231 et 17S162..... 168
Figure 35. Pedigree du cas 14S121 170
Figure 36. Pedigree du cas 15S310 171
Figure 37. Représentation schématique des IPPs..... 214
Figure 38. Représentation de la courbe ROC 218
Figure 39. Différentes méthodes de détection d'IPPs 220
Figure 40. Réseau d'IPPs des gènes de notre population 230
Figure 41. Cluster d'IPPs entre HTT, FXN, ATXN1, ATXN2 et ATXN3 231
Figure 42. Cluster de gènes MNDREM..... 233

REMERCIEMENTS

Remerciements

J'ai passé 6 ans à travailler le sujet que je présente dans ce manuscrit. Six années de labeur de recherche mais également de labeur professionnel et personnel. Six années d'expériences réjouissantes mais parfois aussi douloureuses : un mariage, deux enterrements, une naissance et un enfant malade, un diplôme universitaire, des problèmes au travail et même la pandémie mondiale du COVID19 !...

Il était temps que je conclue ce travail et que je le présente enfin à tous ceux qui ont assisté, collaboré, supporté ou favorisé ces travaux. Ils sont bien entendus nombreux et j'aimerais n'oublier personne. Je vais tenter d'être exhaustive mais j'oublierai probablement d'écrire les noms de ceux qui sont, immanquablement, inscrits dans ma mémoire.

Je voudrais remercier en premier lieu le docteur Rémi BELLANCE, chef de service et coordonnateur du centre de référence CERCA, président de la Société Caribéenne de myologie, sans qui ce travail n'aurait tout simplement jamais pu voir le jour. Je voudrais te remercier pour ton soutien infaillible depuis notre rencontre en 2007. « Faites votre trou !.. » m'avais-tu dis ! Merci pour la bétonneuse que tu m'as mise entre les mains.

Dans le registre des collaborateurs, je tiens à adresser des remerciements chaleureux au docteur Élisabeth SARRAZIN, neuropédiatre du CERCA. Discrète mais efficace, petite mais costaud…merci pour tes épaules.

Enfin, puisque ce travail à des couleurs d'accomplissement, je voudrais aussi remercier tous les seniors de mon entourage professionnel qui, bien que médecin et donc de profession différente de la mienne, sont pour moi de véritables modèles et des moteurs de persévérance : Michèle GUENERET, Bruno SCHAUB, Jocelyn INAMO…je saisi l'opportunité de vous présenter mes sincères remerciements pour les professionnels que vous êtes et pour le privilège que vous m'accordé de pouvoir travailler avec vous.

En parlant de privilège de collaboration, je souhaiterais remercier bien sûr tous mes collègues du CERCA sans qui six ans auraient pu être une éternité. Mais ce serait sans compter sur nos discussions et rires des pauses déjeuner. Alida, Maëlle,

Oriane, Sophie, Micheline et Fred…merci ! Et je n'oublie pas celles qui ont quittés le service, Éliette, Régine et Fabienne, vous avez compté, merci !

Merci aussi aux « drôles de dames » du CPDPN : Valérie, Soraya, Corinne, Carole, Céline et Agnès vous ignorez à quel point vous m'avez fait du bien. Merci ! Là non, plus, je n'oublie pas celles qui sont parties avant que je ne termine ces travaux : Jessica et Gladia, bel passaj' !

Et que serait cet accomplissement sans la poésie d'une vie pleine d'humour qui m'a ramené de façon improbable mon premier chef et formateur, le Professeur Patrice BOUVAGNET. Quelle ironie, quelle surprise mais quelle chance que tu aies apparu à ce moment, retour vers 2003… Je te dois un grand soutien dans la réalisation des travaux bioinformatiques. Je te remercie profondément et espère que nous aurons encore beaucoup d'autres opportunités de travailler ensemble.

En parlant de premier chef, je ne voudrais pas oublier de souligner ta contribution, Jacqueline BAUDIN-VERONIQUE. Tu m'avais encouragé dès notre première rencontre, 2002… Tu m'avais promis de toujours être là. Tu avais placé en moi beaucoup d'espoirs et tu as tenu promesse, tu as répondu présente à chaque fois que je t'ai sollicitée. J'espère que tu es fière de moi parce que je suis fière de toi. Merci d'avoir été là.

En parlant d'humour, je ne peux pas oublier les amis du Diplôme Universitaire (DU) de Séquençage Haut Débit de Dijon, promotion 2017-2018 !… Je voudrais adresser un remerciement spécial à Monsieur Yannis DUFFOURD, responsable et formateur de ce DU. Merci pour ton énorme sympathie, ta patience, tes conseils. J'ai envie de dire : big up !… Quant à Mélissa, Bénédicte, Christelle et Samira, quel plaisir de vous avoir rencontré. Je voudrais avoir un mot tout particulier pour toi Samira, pour ton immense cœur, ta sympathie et ton soutien. Merci !

J'ai eu d'excellent contact avec la plate-forme LIGAN-GENOSCREEN. Je remercie tout particulièrement Mehdi DERHOURHI et Véronique DHENNIN pour leur réactivité et leur professionnalisme.

Et sur cet aspect bioinformatique, je tiens à tirer mon chapeau à Kevin RAGUETTE, bioinformaticien fraichement diplomé, tu as participé à la réalisation du travail bioinformatique et je voudrais t'adresser des remerciements particuliers pour ta justesse et ta rigueur de travail. Je te souhaite longue vie et épanouissement professionnels.

Comment alors ne pas remercier le monument, monsieur le Professeur Cyril GOIZET, coordonateur du site constitutif de neurogénétique au CHU de Bordeaux et co-directeur de cette thèse, dont je retiens surtout la capacité à remplir l'espace sonore de son rire indomptable à la façon antillaise, j'adore !...humour et talent, voilà de bonnes inspirations.

Et le Professeur Didier LACOMBE, directeur de l'unité INSERM U1211 à Bordeaux, merci de ton soutien : Arabie Saoudite, Afrique du Sud, Bordeaux, Martinique…tu étais avec moi ! Je te remercie de ta confiance, elle est source d'espoir et de confiance en moi.

Et madame LE Professeur Juliette SMITH-RAVIN, directrice de cette thèse et du groupe de recherche BIOSPHERES, présidente de l'association AREBio qui a soutenu les publications de ces travaux. Je n'aurai pas assez de place pour te dire toute ma reconnaissance. Tu m'as ouvert les portes de la recherche en Martinique. Tu m'as fait une confiance presqu'aveugle depuis le début et même bien avant cette idée folle de doctorat. Tu as été d'un soutien à toutes épreuves et sur tous les plans : professionnels, personnels et spirituels. Tu as donné bien plus que n'importe quel autre directeur de thèse et je n'ai pas de mots pour te dire toute ma reconnaissance.

Et puis, un merci solennel et officiel mais bien pauvre par rapport à ce qu'il est réellement pour toi, invétérée mamie Jeanne, MA maman ! Et pour toi, MA sœur…

Enfin, inutile de m'adresser ici aux plus importants d'entre tous puisqu'ils ne liront sans doute pas ce manuscrit (en tout cas probablement pas tout de suite). Mais pour ceux qui en auront le courage, sachez qu'ils s'appellent Mehdi, Isaac, Emmanuel et Joachim…et, eux, une vie ne suffirait pas à les remercier pour ce qu'ils sont pour moi…

« Le vieillissement est comme le temps : phénomène insaisissable et tantôt vue de l'esprit !?... Ils sont intriqués, et font depuis toujours, l'objet de questions énigmatiques »

Anna-Gaëlle

PREAMBULE

Préambule

Tan fè tan, tan kité tan…
Littéralement du créole : *le temps a fait son temps, le temps a quitté son temps…*
Signifie que beaucoup de temps s'est écoulé

Avant de débuter la présentation des travaux, il me semble important de poser le socle théorique des pathologies d'intérêt et de rappeler en préambule les deux concepts que sont : la neurodégénérescence et le chevauchement clinique. Car, ils justifient l'intrigue d'un sujet fantasmatique de recherche mettant en jeu certains aspects des processus induits par le temps qui passe.

Le terme de « dégénérescence » dérive du concept de « dégénération ».

Au XVIII^{eme} siècle, le naturaliste GL Leclerc de Buffon (1707-1788), explique la variété des espèces par la dégénération. Cette dernière devient au cours du XIX^{eme} siècle la marque des ratures du genre humain. On en étudie alors les signes. Des centaines d'ouvrages portent sur l'analyse de dégénération physique ou morale, l'altération des sens et autres aliénations comme la perte de fécondité, les prostitués, les criminels etc…

C'est avec W Gowers (1845-1915) que commence la conception moderne de la dégénérescence. Celle-ci propose d'opposer deux processus vitaux :

- un processus global, entretenu par le sang propulsé par le cœur, oxygéné par les poumons ;
- et un processus local, géré au niveau des organes et susceptible d'altération ne compromettant pas le processus global.

La dégénération devient « *a defect of vital endurance* » c'est à dire « *une mort lente d'une partie du tout* ». Il tire ses exemples des modifications du système pileux et des pathologies nerveuses décrites à l'époque : l'ataxie de Friedreich, d'apparition précoce ou encore la paralysie agitante à laquelle Charcot conféra le nom de Parkinson, qui est d'apparition tardive.

La dégénérescence est donc d'abord un phénomène spontané et naturel, lié au temps qui passe : tout organisme vivant dégénère, c'est le vieillissement !

Par la suite, on comprendra que la dégénérescence prématurée et pathologique, fait intervenir les mêmes circuits physiologiques que ceux qui impliquent le vieillissement. Vieillissement et pathologies se rencontrent…Mais à cette époque, le concept de cellule n'existait pas !…[1]

De façon plus contemporaine, et sur un plan strictement biologique, la dégénérescence s'entend de l'altération progressive des fonctions. Cette altération prend source dans un dysfonctionnement cellulaire dont le retentissement est graduel : cellule, tissus, organe, système.

Il faut donc considérer qu'une maladie dégénérative est une pathologie provoquant le déclin anormalement prématuré des fonctions de certains organes. Communément, les maladies dégénératives s'expliquent soit, par l'accumulation de toxines ou de produits biologiques ; soit par l'absence prolongée d'un produit biologique, provoquant le dysfonctionnement voire la dégradation de ou des organes concernés.

La neurodégénérescence, est donc la dégénérescence qui concerne le système nerveux. Les Maladies Neurodégénératives (MND) constituent un sous-groupe de maladies dégénératives. Il s'agit de pathologies chroniques, le plus souvent lentement progressives provoquant la dégradation des cellules nerveuses. L'atteinte tissulaire peut être diffuse ou limitée. De la localisation de l'atteinte, et du groupe cellulaire concerné, dépend la symptomatologie.

La neurodégénérescence est un des « défis majeurs » de société.

Depuis plus de dix ans, la recherche sur les MND est une priorité Européenne. Et dès 2005, le gouvernement français publie des plans de santé pour lutter contre les maladies rares. Les axes définis pour la recherche dans les MND se sont étendus à toutes les maladies neurodégénératives invalidantes, en particulier la maladie de Huntington, les maladies des motoneurones comme la sclérose latérale amyotrophique, mais aussi, les ataxies spinocérébelleuses ou les maladies à prions. La prévalence de ces-dernières est plus faible mais ces maladies comportent souvent une composante familiale.

En France, l'organisation de la santé a permis de programmer pour la période 2014-2019 des axes prioritaires de recherche et d'amélioration diagnostic/qualité de vie des patients atteints de la maladie d'Alzheimer, de la maladie de Parkinson et de la sclérose en plaques. Rappelons que, dans ce cadre, le gouvernement a aussi encouragé la création de centres de référence pour le diagnostic clinique et la prise en charge des maladies rares.

Depuis 2006 aux Antilles françaises, le CEntre de Référence CAribéen (CERCA) des maladies rares est agréé pour le diagnostic des troubles neuromusculaires et neurologiques. Il est situé au Centre Hospitalier Universitaire de Martinique (CHUM), à Fort-de-France. Le CERCA coordonne les consultations médicales en Guadeloupe et en Guyane française. Les consultations ont lieu régulièrement dans les îles de Sainte-Lucie et de la Dominique. Le CERCA reçoit également à la demande, des patients d'autres régions des Caraïbes. [2]

Complexité diagnostic et étiologique des MND

Les premières classifications cliniques des pathologies neurodégénératives étaient fondamentalement basées sur l'observation des symptômes. Mais très rapidement, des études anatomopathologiques, en particulier du cerveau de sujets symptomatiques, ont permis de distinguer des sous-groupes cliniques en fonction du type de lésions observées. Ainsi, on a pu corréler la neurodégérescence soit à une atrophie, c'est-à-dire une perte neuronale et une altération structurale plus ou moins spécifique ; soit à un dépôt de matériel cérébral appelé gliose. Dans ces dernières, l'identification des protéines spécifiquement accumulées, ainsi que

l'identification de certains processus biochimiques et moléculaires, ouvrent des orientations diagnostiques à la fois plus fines mais également plus complexes.

Dorénavant, le diagnostic génétique s'ajoute aux démarches étiologiques et « classifiantes ». Or, il existe probablement plus de 1500 gènes responsables de pathologies neurodégénératives. Les classifications explosent et sont revues régulièrement à la lumière des avancées moléculaires et des observations cliniques.

Les MND sont sporadiques ou familiales. Dans ce cas, on peut voir tous les modes de transmission : mitochondriale, mendélienne ou poly-factorielle.

Et sur le plan génomique, on observe des variants pathogènes de toutes natures : ponctuelles, insertion/délétion, amplification de motifs répétés, ou Variations de Nombres de Copies (CNVs).

Actuellement, on considère que les formes familiales/héréditaires et les formes sporadiques/tardives ne répondent pas exactement au même processus de dégénérescence : hyperproduction pour les premières, accumulation pour les deuxièmes.

Les MND d'origine génétique sont souvent d'expression clinique plus précoce et/ou plus sévère. Les formes Héréditaires (MNDH) représentent des modèles d'étude offrant la possibilité d'analyser des ADN apparentés et d'augmenter de ce fait la puissance de recherche de facteurs génétiques impliqués dans une symptomatologie familiale.

Les MNDH sont rares. Une revue extensive de la littérature permet de penser qu'il existe des processus physiopathologiques communs à la plupart des MNDH. Ces mécanismes semblent proches de ceux que nous connaissons du vieillissement, impliquant notamment les voies de l'apoptose, des perturbations du système Ubiquitine-protéasome, un dysfonctionnement mitochondrial et/ou métabolique, des perturbations cellulaires par les petits ARN…[3]–[5]

On trouve de plus en plus de variants génétiques pathogènes responsables des MNDH. Ce qui conforte l'idée que les atteintes neurodégénératives seraient une forme d'anticipation programmée du vieillissement.

Hétérogénéité génique et spécificités cliniques

Les processus neurodégénératifs peuvent générer une détérioration des capacités cognitives associée ou non à une atteinte motrice, et vice-et-versa. Mais les classifications cliniques et génétiques ne sont toujours pas assez discriminantes. On se retrouve alors face à des symptomatologies atypiques, des chevauchements cliniques, des maladies alléliques.

Comme dans les variants pathogènes du gène *C9ORF72*, responsables tantôt de Sclérose Latérale Amyotrophique, tantôt de Démence Fronto-Temporale ; ou quand l'altération moléculaire du gène *ATXN2* habituellement associée à une ataxie spinocérébelleuse, est responsable d'une maladie de Parkinson typique d'évolution lentement progressive sans atrophie du cervelet ni troubles moteurs ; Ou encore, dans l'observation de ces quelques cas d'atteinte de la corne antérieure associée à des troubles ataxiques.

Une meilleure connaissance des processus physio-moléculaires de dégénérescence pourrait vraisemblablement aider à la compréhension de l'origine et du pronostic de ces expressions phénotypiques.

La population caribéenne présente des spécificités dans l'expression de MND. Certaines sont publiées, comme le Parkinsonisme atypique en Guadeloupe/Martinique [6]–[8]. Le brassage ethnique d'où est issu la population Caribéenne, en particulier martiniquaise, laisse présumer de variétés cliniques et moléculaires à explorer.

C'est dans ce cadre que s'inscrivent les travaux présentés dans ce manuscrit autour de la caractérisation d'une population de MNDH Rares à Expression motrice (MNDREM) et l'identification de variants géniques impliqués dans un échantillon de MNDREM probablement prédisposé.

Ce travail de recherche effectué à l'université des Antilles, s'inscrit dans un paysage de recherche expérimentale et clinique relativement pauvre. Innovant tant sur le plan clinique que moléculaire, il a été supporté par la Société Caribéenne de Myologie (CSM) et AREBio, association locale de soutien à la recherche. Il a été possible dans le cadre de l'exercice de mes fonctions d'ingénieur en biologie moléculaire et conseillère génétique au Centre de Référence Caribéen des Maladies Rares neurologiques et neuromusculaires (CERCA), basé au CHU de Martinique (CHUM), à Fort-de-France.

Ces travaux ont fait l'objet de déclaration CNIL et obtenus avis favorable de l'Institutional Review Board (IRB) du CHUM. Toutes les dispositions d'informations et consentement des participants ont été prises pour la réalisation du travail dans le respect des réglementations éthiques de recherche et d'étude.

La réalisation de ces travaux revêt donc une dimension novatrice d'ores et déjà publié dans le Journal of Global Health [9].

Le rapport de thèse s'articule en 4 chapitres…

Le premier chapitre dresse un état des lieux clinique et moléculaire autour des pathologies d'intérêt : Maladies Neuro-Dégénératives Rares à Expression Motrice (MNDREM) associées ou non à un profil cognitif et/ou démentiel rencontrées au centre de référence de Martinique, le CERCA. Ce chapitre révèle la complexité moléculaire de ces atteintes. Il documente les chevauchements symptomatiques décrits dans la littérature et les limites des classifications actuelles pourtant en constante évolution. Il amène à la justification de ce travail.

Le second chapitre rapporte la stratégie d'étude démographique des MNDREM et la méthode expérimentale mise en œuvre pour la caractérisation moléculaire d'un échantillon de cette population. En particulier le Séquençage Nouvelle Génération (Next Generation Sequencing, NGS) qui est un outil d'analyse puissant, rapide et sensible. Cet outil est employé dans la recherche étiologique de maladies de diverses spécialités médicales, comme en cardiologie dans le déterminisme des malformations cardiaques congénitales où il a permis de détecter un variant pathogène causal chez 15% des fœtus testés [10]. Pour ces travaux, nous avons employé une méthode de NGS originale permettant de détecter à la fois les variants génomiques simples, SNVs et variants de nombre de copies, les CNVs.

Le troisième chapitre est le rapport des résultats. Il présente d'une part, l'étude démographique et la caractérisation moléculaire des patients atteints de MNDREM suivis au CERCA, d'autre part, les variants détectés par le travail expérimental. Nous permettant ainsi de définir la liste des gènes en cause dans les MNDREM vues au CERCA.

Le quatrième chapitre est une étude préliminaire bioinformatique des Interactions Protéines-Protéines (IPPs) issues des gènes associés aux MNDREM de notre population. Il est une démonstration de l'intérêt de l'approche bioinformatique sur la question du chevauchement clinique. Il ouvre des perspectives de travaux innovants en la matière.

Valard-Giguet A-G, Véronique-Baudin J, Duffourd Y, et al. Capacity building in clinical and molecular characterization of rare neurodegenerative diseases through the development of a local register and improved use of bioinformatics data analysis in Martinique. *Journal of Global Health Reports*. 2020;4:e2020003.
doi:10.29392/001c.11877

Research Articles

Capacity building in clinical and molecular characterization of rare neurodegenerative diseases through the development of a local register and improved use of bioinformatics data analysis in Martinique

Anna-Gaelle Valard-Giguet[1], Jacqueline Véronique-Baudin[1], Yannis Duffourd[2], Jocelyn Inamo[1], Cyril Goizet[3], Rémi Bellance[1], Juliette Smith-Ravin[4]

[1] Caribbean Reference Center for Rare Neuromuscular and Neurological Disorders (CERCA) – Fort-de-France University Hospital Center, Martinique, [2] Computing Center, University of Burgundy, France, [3] Neurogenetics Department - Bordeaux University Hospital Center, France, [4] Biospheres Group, University of Antilles, Martinique

Keywords: martinique, bioinformatics, neurodegenerative diseases, capacity building

https://doi.org/10.29392/001c.11877

Journal of Global Health Reports

Vol. 4, 2020

Neurodegenerative diseases (NDs) are very polymorphic and affect people of all ages. They concern both rare diseases and more common diseases. Improving knowledge of NDs in the French Caribbean region requires special attention because of the unique environmental and genetic background of its populations. In Martinique, there is a huge need for scientific research on rare NDs. While epidemiological and clinical research has increased over the past decade in the French West Indies, there is still little or no data on the genetics of NDs in these regions. The advent of Next Generation Sequencing (NGS) on a global scale has the potential to make a difference in this field, provided that technological and analytical knowledge, based on bioinformatics, is available in Martinique. Caribbean Reference Center for rare neuromuscular and neurological disorders (CERCA), is a reference center for diagnosis, health care and treatment of rare NDs. It supports our capacity building project in the clinical characterization of rare NDs with motor impairment, with or without dementia. In collaboration with CERCA, we designed a two-step pilot study, consisting of: 1) the creation of a clinical and biological database to select informative cases, and 2) to screen them by whole exome sequencing. This innovative approach involves: 1) mobilization of medical and clinical knowledge for the characterization of rare ND, with the support of a register of experts, and 2) expertise in molecular biology, molecular pathways and bioinformatics. This preliminary study confirms the need to consider our French West Indian population in these specificities. It reveals the effectiveness of a well-tune database for identifying pathogenic variants in a cohort of French West Indian patients with presumed genetic NDs, associated with motor impairment with or without dementia. It is a proof of concept that the creation of a register of NDs and the mastery of NGS technique, can provide additional expertise in research and patient management in the Caribbean. The continuation of this capacity building project should increase CERCA's skills and outreach. Such an initiative is clearly innovative for the region and would bring the Caribbean, Latin America and North America communities together, around the subject.

INTERNATIONAL AND NATIONAL CONTEXT AND ISSUES

As neurodegeneration is one of society's "major challenges", research on neurodegenerative diseases (NDs) has been a European priority for more than 10 years. The organization called "Join Project Neurodegenerative Diseases Research" has ruled on the necessity to organize European research on neurodegenerative diseases, and to develop a partnership with Latin America and the Caribbean countries, where populations have different environmental and genetic backgrounds that could be a model study.[1]

The first French health plan for neurodegenerative diseases was drafted in 2014 for a period of five years. The directions and actions of the Neurodegenerative Diseases Plan (in French, "PNMD") are part of the National Health Strategy and the National Research Strategy. It has 3 main priorities:

- Improve patient diagnosis and management.
- Ensure the quality of life of patients and their caregivers.
- Develop and coordinate research.

The plan for the period of 2014-2019 concerns all patients with Alzheimer's disease, Parkinson's disease and multiple sclerosis, and is extends to all neurodegenerative diseases. The plan includes other disabling neuropathic diseases, such as Huntington's disease, motor neuron diseases (such as amyotrophic lateral sclerosis), spinocerebellar ataxias or prion diseases. Their prevalences are low but they often include a family component. Neurodegenerative diseases are very polymorphic and affect all ages of life. They concern both rare diseases and more common diseases.[2] The French government has been publishing health plans to fight rare diseases in particular, since 2005. For the clinical diagnosis and management of rare diseases, the French government encourages the creation of reference centers.

In the French West Indies (FWI), the Caribbean reference center for rare diseases has been specialized in the diagnosis of neuromuscular and neurological disorders since 2006. It is called the "Caribbean Reference Center" (CERCA) and

is located at the "Centre Hospitalier Universitaire" (CHUM), in Fort-de-France, Martinique. CERCA coordinates medical consultations in Guadeloupe and French Guiana. Consultations are held regularly in the islands of Saint Lucia and Dominica. CERCA also receives patients from other Caribbean areas, according to their requests.[3] The FWIs are particular French territories: Martinique and Guadeloupe are Caribbean islands, and French Guyana is located on the Latin American continent. The French Caribbean countries were inhabited by Caribbean Amerindians and were later populated mainly by African and European migrations, as well as Indian. The inhabitants of these French-speaking countries are therefore atypical French from an environmental and genetic point of view, but typical West Indians resulting from a unique genetic mix, just like other independent Caribbean peoples.

LOCAL AND REGIONAL CONTEXT AND ISSUES

Since 2015, Martinique has been a member of the Organization of Eastern Caribbean States (OECS). It is an international intergovernmental organization dedicated to economic harmonization and integration, to the protection of human and legal rights, and the promotion of good governance between independent and non-independent countries in the Eastern Caribbean. Our island is therefore involved in the development of centers of excellence in public health for the OECS, in the achievement of economies of scale, the dissemination of knowledge and experience, and in the research synergy to improve the quality of care and health care delivery. Strengthening the capacity of CERCA, which is already involved in the diagnosis and treatment of Caribbean patients with ND, would reinforce its excellence and its outreach, while helping to improve the skills of neighbouring islands.

Capacity building through the creation of a registry is the first argument of our approach. It is a process of individual and institutional development that leads to higher levels of competence and a greater capacity to conduct useful research. Population-based registries provide public health surveillance and research. Health indicators generated by population registers are useful, not only for the population, but also for:

- health facilities and local authorities
- public health decision-makers
- clinicians and researchers.

To develop sustainable epidemiological surveillance and evaluation, as well as effective research programs on rare ND in the Caribbean, the entry of clinical and molecular data into a registry will be crucial for: the production of reliable epidemiological data, improving management, reduction of mortality and so on.

Considering the Caribbean population origins, genetic determinism must be explored, in order to explain certain specificities of Caribbean NDs. Some local specificities remain clinically determined and published, such as atypical Parkinsonism in Guadeloupe/Martinique,[4] but there is no local or specific epidemiological database on NDs, whereas the epidemiological register of cancers and fetal malformations in the FWI has already existed for more than 20 years for the first and 10 years for the second. As the literature on genetics of NDs in the Caribbean is rather poor, collaborations consisting of a transfer of know-how and experience from these highly qualified registers to a NDs register will be considered, both at the local and Caribbean level.

AN INNOVATIVE PROJECT IN EVERY RESPECT THAT OPENS UP NEW PERSPECTIVES

The main objective of this study is to highlight causal mutations in targeted rare diseases, presumed genetic, for which the classical molecular diagnostic pathway has not been able to identify a causal mutation. There are many secondary objectives:

- describe the clinical characteristics of the targeted pathologies in Martinique
- describe the modes of transmission of family pathologies
- provide diagnostic assistance
- allow genetic counselling
- test a new experimental device
- highlight local specificities
- initiate the establishment of a register of clinical data specific to targeted diseases in Martinique.

The first step of our capacity building project is to develop a clinical and biological database on NDs associated with motor disability, with or without dementia, in Martinique. It is a cornerstone and a stepping stone for our pilot study. The second step of our capacity building project consists in highlighting the potential of the register in the selection of cases on which to carry out the molecular research program. More than 80% of rare diseases are genetically determined, but many rare diseases remain unexplained at the molecular level. With the advent of next-generation sequencing (NGS), the world of molecular diagnostics and genetic investigation strategies is changing on 2 main levels:

1. We are moving from a genetic medicine based mainly on monogeny to a genetic medicine open to polygenism.
2. Clinical limitations can be overcome by interrogating gene panels.

THE NGS, A TOOL OF CHOICE FOR OUR INVESTIGATIONS

Gene panels are currently used for the diagnosis of NDs according to their expression and whole exome sequencing is used for the molecular diagnosis of NDs with a motor impairment, such as spinocerebellar ataxias. A national policy attempts to harmonize practices and costs of research on the whole exome and the whole genome. But the lack of molecular diagnostic technology in Martinique is leading the medical and scientific community to less effective research. For instance, we identified only 5 studies published on Pubmed since 1996, which used DNA samples from the French West Indies. None of them report DNA analysis on a local platform.

We have built for the first time in Martinique a NGS project in which all data analyses were carried out locally. Using clinical data entered into the clinical database we have created, we selected patients in whom a complete exome sequencing could be performed. We have explored presumed genetic cases, for which the usual molecular diagnostic pathways have not revealed any mutations, and through this work, we have found causal mutations for some patients, and highlighted molecular originalities.

This NGS project that stems from our embryonic register of rare NDs, is a proof of concept that we have new possibilities, with a well-adjusted register and bioinformatics technologies, to provide answers in terms of genetic characterization of our patients, genetic counselling and ultimately, patient management and treatment. The number of people

with NDs tends to increase, particularly for diseases affecting cognitive functions. Our overall capacity building project to create a register and improve the ability to use data to conduct research projects, such as metabolomics, epigenetics, and so on, is fully in line with European and international policies, as well as national and Caribbean policies.

Our project was entirely financed by private funds raised by the Caribbean Society of Myology. This is the first NGS project of its kind in Martinique and the FWI Islands. This is very innovative for the region and could be a call for cooperation. Such collaborative approaches could help to bridge disciplines and reduce differences in the management of rare NDs between Caribbean countries by assessing care models for patients with rare NDs. An additional file concerning the "Standards for Reporting Implementation Studies" shows completion of our pitch in the field.

METHODS

DESCRIPTIVE DATA COLLECTION AND SELECTION OF STUDY PARTICIPANTS

The first step of this project is the creation of a synthetic register to collect local cases and their clinical specificities. National databases already exist but are not specific to our local NDs cases and do not really meet our needs. The management of these national databases is still a continental matter and collaboration on research projects may be difficult due to political and geographical distance or differences in organizational methods between France and French West Indian islands. A local register should be objectively more effective, just as those that already exist locally, such as the cancer register, for more than 20 years or the fetal malformation register (REMALAN ©), for 10 years.

Using Filemaker © software (Claris Inc, Santa Clara, CA, USA), we design a computer database to collect rare cases of NDs, with or without dementia, observed at CERCA. This database is located at Martinique University Hospital and is supported by the Caribbean Society of Myology. It lists:

- patients identifications
- clinical data, particularly age of onset
- inaugural signs, and associated medical history
- the result of systematic exploration in diagnostic practice
- molecular research carried out
- family history, with particular attention to family history of other neurodegenerative diseases. For example, a parent with dementia may be related to a case of amyotrophic lateral sclerosis.

This step forms the basis for a monocentric retrospective clinical study of 300 adult patients with rare neurodegenerative diseases associated with motor disability with or without dementia. Molecular analysis is based on the selection of a cohort of patients who are highly susceptible to genetic damage. We therefore apply the following selection criteria: patients with a family history or early age of onset, without mutation identified by the conventional diagnostic pathways. We carry out a study by NGS analysis of approximately 14 cases out of 300, i. e. 5% of our listed population, according to the literature.

TECHNICAL INFORMATIONS

As part of routine diagnosis, panels have been set up to search for mutations in cases of unresolved NDs. But exome sequencing is widely used today and gives excellent results to improve molecular diagnosis in certain neuromuscular diseases, especially in spinocerebellar ataxia.[6]

For the research and diagnosis of our cases in the French West Indies, patients' DNA is currently sent to specific laboratories based mainly in France. The cost of genetic testing per NGS is approximately €2206 (US$ 2452.43). Specific envelopes are allocated by the French government to cover this cost and allow payment to approved laboratories to carry out analyses. Given the technicality and level of expertise required, however, there will be fewer and fewer accredited laboratories and more and more queues. Qualifying our hospital and our university for this technology would represent a definite economic and scientific advantage, and a springboard towards regional collaborations.

There is no technological platform dedicated to these genetic tests in Martinique. For our project, exome experiments were subcontracted to LIGAN. We studied cases of amyotrophic lateral sclerosis, spinocerebellar ataxia and early multisystemic atrophy. Fourteen cases were tested for a total of €1806 (US$ 2007.75). The raw data was sent to BIOSPHERES by an external hard disk. Bioinformatics analyses of the selected cases were carried out on the computer centers of University of Bourgogne, University of Lyon and University of Antilles.

A standard pipeline aligned with the current version of the Genome Reference Consortium of Human Build 38 (Grch38), has been developed. The GnomAD variable frequency database and the usual in silico prediction tools were used. The recommendations of the American College of Medical Genetics and Genomics (ACMG) were used for filtration of the variants.

STATISTICS

We did not use any significant statistical analyses. But we simply compared our results to those proposed in the literature, where this was useful and possible.

RESULTS

THE FIRST STEP CONFIRMS THE ORIGINALITY OF CERTAIN PHENOTYPES OR HEREDITARY MODELS

For the first time in Martinique, we are collecting rare cases of NDs, particularly associated with motor impairment, with or without dementia. The first step of the study allows us to retrospectively collect approximately 300 cases of rare NDs observed at CERCA. This step confirms the originality of certain phenotypes or hereditary models. It promotes an epidemiological database on amyotrophic lateral sclerosis, spinocerebellar ataxia, Parkinson's disease, Parkinson's syndromes and Huntington's disease. This retrospective study reveals some atypical characteristics:

- a high prevalence of Huntington's disease (HDL2)
- a phenomenon of anticipation rarely described in a large family of amyotrophic lateral sclerosis (ALS)
- a few cases of dominant or recessive spinocerebellar ataxia (SCA) for which conventional diagnostic pathways have not been able to determine mutations.
- the original forms of multisystemic atrophy (MSA).

THE SECOND STEP OF THE STUDY ALLOWS THE UNIVERSITY OF ANTILLES TO IMPROVE ITS LOCAL CAPACITY IN BIOINFORMATICS ANALYSIS AND FIND MUTATIONS

The second phase of the project involves applying New Generation Sequencing (NGS) to non-genetically determined cases of rare NDs with motor impairment, with or without dementia. The bioinformatics analysis was mainly carried out in collaboration with the BIOSPHERES research group in Martinique and the bioinformatics support of the computer center of University of Antilles and University of Bourgogne, with the kind contribution of the University Lyon I. This phase of the project allows the University of Antilles to improve its local capacity in bioinformatics analysis. Fourteen cases were analyzed by exome sequencing. At least 2 cases out of 15 (13%) have been elucidated, which match the expected effectiveness rate for the NGS strategy.[7]

The preliminary results of our study are encouraging in terms of resolving cases without identified mutations and highlighting the described rare mutations. This capacity building project is a proof of concept because it allowed us to:

- identify very rare mutations in dominant and recessive SCA families
- exclude digenism in the family case of ALS with anticipation
- propose candidate genes in the case of AMS.

It is also proof of our ability to solve clinically, diagnostically and molecularly important problems ourselves, especially at much lower costs than existing networks.

DISCUSSION

THE PROJECT CONFIRMS THE NEED TO CONSIDER OUR WEST INDIAN POPULATION IN ITS SPECIFICITIES

The capacity building project is based on an innovative process. This process combines the construction of a database, and the deployment of a new NGS data analysis capability. On the one hand, the database should form the basis for a register of rare NDs, which implies the mobilization of medical and clinical knowledge for the characterization of rare neurodegenerative diseases. On the other hand, increasing the capacity to analyse and interpret NGS data will improve skills in molecular biology, molecular pathways and bioinformatics.

The project objective is to clarify local specificities in terms of surveillance and research on rare NDs with motor impairment with or without dementia. The 1st step consists in accurately collecting clinical, biological and genetic data; the 2nd step consists in developing local capacities in innovative technologies such as NGS. Through this project, we were able to identify:

- a particular local phenotype in Huntington' disease
- a phenomenon rarely described in amyotrophic lateral sclerosis
- mutations that could not have been identified from the routine diagnostic process.
- particular mutations.

This pilot study demonstrates the effectiveness of the process in improving the clinical and molecular characterization of rare ND diseases and promoting research in this area. It confirms the need to consider our West Indian population in its specificities. We demonstrate the incredible potential for research and clinical surveillance of rare diseases associated with motor impairment with or without dementia, developed from the establishment of a local register, and the improvement of our expertise in genomic analysis and bioinformatics.

The search for a genomic mutation is becoming less expensive and less constraining. Bioinformatics plays a central role in this development. Henceforth, the deployment of a genomics service dedicated to the search for genetic mutations in Martinique should be achievable to take advantage of the benefits of clinical and fundamental research in the Caribbean population, particularly in the NDs.

LIMITATIONS OF THE STUDY

The only limit of this project remains the funding. It is essential to sustain our capacities and implement our project whose objectives are crucial and are already enshrined in national and European health policies.

CONCLUSIONS

THE PROJECT OPENS UP OPPORTUNITIES FOR COLLABORATION

The collaborations set up should make it possible to perpetuate our project and develop research partnerships on the subject in the Caribbean. Such a dynamic will enable our center to become one of the centers of excellence in research and surveillance of rare diseases, a vector of attractiveness for medical and scientific cooperation.

Beyond the essential project of strengthening the characterization capacities of West Indian patients suffering from NDs in Martinique, the challenges of this project also concern:

- the sharing of experiences and data
- access to technological innovation
- in fine, the avenues of therapeutic research.

This project opens up the possibility of many related works. In this instance, we were able to develop the NGS, but we could imagine other biological study schemes (such as metabolomics, epigenetics, and so on). This project is an opportunity to bring health professionals together around an electronic collection of biological and medical data. This allows us to consider scientific studies on a larger scale, in the inter-region. As with all challenges that affect multiple communities, a collaborative project with the University of Antilles and others, should be required to improve the characterization of local specificities and diversities in rare NDs.

In France, scientists and medical communities are trying to regulate access to genome-wide technologies in order to offer the same level of diagnostic quality throughout the country. Given our geographical location and our various local needs, it would be interesting to develop our expertise in genomics and bioinformatics.

THE PROJECT PROMOTES COOPERATION

Due to the geographical and ethnic proximity of the Caribbean populations, the needs for scientific research on rare NDs associated with motor impairment with or without dementia, are similar throughout the region. This project should make it possible to initiate collaborations with existing or emerging research programmes in the Caribbean. It meets the expectations of the "Join Project Neurodegenerative Diseases Research", in terms of investigations of the

Latin American and Caribbean populations.

As a general rule, Martinique' health professionals are trained in medical cooperation for the exchange and identification of needs for additional health training. This is an important issue in our outermost regions, which are involved in the Organization of Eastern Caribbean States (OECS).

In the field of rare ND diseases, collaborative efforts will map and model health data for this sector within the Caribbean region. These efforts should also help to highlight the importance of structuring the sometimes complex management of patients with rare NDs, with or without dementia.

CERCA is already a Caribbean reference centre, but this clinical and molecular capacity building project is an additional means of expertise. It strengthens medical research prospects and contributes to the emergence of innovative clinical, diagnostic or therapeutic tools in terms of patient service.

Our project was entirely financed by private funds raised by the Caribbean Society of Myology. We now need more local collaborative partners to bring research capacity in the Caribbean region to its optimal potential for the diagnosis and treatment of rare ND associated with motor impairment with or without dementia. While the Caribbean is only now trying to build its capacity in genetic testing,[8] Trinidad and Tobago has already been doing so since 2015,[9] as have our other parents from sub-Saharan Africa.[10]

Acknowledgments: I would like to warmly thank Mr Kevin Raguette and Dr Patrice Bouvagnet for their contribution to bioinformatics analyzes.

Funding: The Caribbean Society of Myology has met the financial needs of research, especially the subcontract of NGS technology.

Authorship contributions: The authors declare that the opinions expressed in this article are theirs and not subject to an institutional dictate.

Competing interests: The authors completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available upon request from the corresponding author), and declare no conflicts of interest.

Correspondence to:
Anna-Gaelle Valard-Giguet, Msc and genetic counselor
CERCA CHUM
PZQ1 niveau -1 97200
Fort-de-France
Martinique
anna-gaelle.giguet@chu-martinique.fr

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

REFERENCES

1. Joining forces to fight neurodegenerative diseases. *Lancet Neurol*. 2013;12:119-216. doi:10.1016/S1474-4422(13)70004-8

2. Le plan maladies neuro-dégénératives 2014-2019. https://www.gouvernement.fr/action/le-plan-maladies-neuro-degeneratives-2014-2019. Accessed November 29, 2018.

3. Bellance R. Le Centre de Référence Caribéen des maladies neuromusculaires et neurologiques rares. *médecine/sciences*. 2015;31:45-47. doi:10.1051/medsci/201531s313

4. Lannuzel A, Hoglinger GU, Verhaeghe S, Gire L, Belson S, Escobar-Khondiker M, et al. Atypical parkinsonism in Guadeloupe: A common risk factor for two closely related phenotypes? *Brain*. 2007;130:816-827. doi:10.1093/brain/awl347

5. Joachim C, Veronique-Baudin J, Ulric-Gervaise S, Pomier A, Pierre-Louis A, Vestris M, et al. Cancer burden in the Caribbean: An overview of the Martinique Cancer Registry profile. *BMC Cancer*. 2019;19(1):239. doi:10.1186/s12885-019-5434-6

6. Marelli C, Guissart C, Hubsch C, Renaud M, Villemin J-P, Larrieu L, et al. Mini-exome coupled to read-depth based copy number variation analysis in patients with inherited ataxias. *Hum Mutat*. 2016;37:1340-1353. doi:10.1002/humu.23063

7. Montaut S, Tranchant C, Drouot N, Rudolf G, Guissart C, Tarabeux J, et al. Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders. *JAMA Neurology*. 2018;75:1234. doi:10.1001/jamaneurol.2018.1478

8. Pan American Health Organization. *Report on Strengthening Research Capacities for Health in the Caribbean, 2007-2017*. Washington, DC: PAHO; 2017.

9. Roach A, Warner WA, Llanos AAM. Building capacity for human genetics and genomics research in Trinidad and Tobago. *Rev Panam Salud Publica*. 2015;38:425-430.

10. Wonkam A, Muna W, Ramesar R, Rotimi CN, Newport MJ. Capacity-Building in Human Genetics for Developing Countries: Initiatives and Perspectives in Sub-Saharan Africa. *Public Health Genomics*. 2010. doi:10.1159/000294171

CHAPITRE I

État de l'art

CHAPITRE I - État de l'art

Ce premier chapitre présente quelques généralités sur la dégénérescence neurologique à expression motrice avec ou sans démence ainsi que les Maladies Neuro-Dégénératives Rares à Expression Motrice (MNDREM) qui ont fait l'objet de ces travaux. Nous aborderons leurs moyens diagnostics, leurs caractéristiques clinique et moléculaire, et nous soulignerons quelques particularités révélant des chevauchements clinique et moléculaire ou des phénotypes atypiques.

Ce qui nous mènera à approcher la complexité de ces pathologies singulières mais très hétérogènes sur le plan clinique et génétique, et justifier le déploiement du NGS pour en étudier les caractéristiques génomiques non identifiés par les filières de diagnostic médical.

1. Dégénérescence neurologique à expression motrice

1.1. Principaux mécanismes moléculaires de dégénérescence cellulaire

Les principaux mécanismes moléculaires et cellulaires physiopathologiques impliqués dans la dégénérescence rejoignent les processus de mort cellulaire programmée. Les processus de vieillissement sont inscrits dans le patrimoine génétique de notre espèce, tandis qu'on pourrait dire que les processus neurodégénératifs seraient plutôt individuels. Il est également communément acquis que ces processus de dégénérescence, peuvent être édités par le temps et l'environnement.

1.1.1. Variants du génome et dégénérescence

Le génome, composé d'acide Désoxyribonucléique (ADN), porte l'information génétique. L'ADN est nucléaire ou mitochondrial. L'ADN nucléaire est responsable de transmission de caractères mendéliens, tandis que l'ADN mitochondrial, de transmission du même nom, porte des caractères se transmettant par le biais des ovules maternels. L'information portée par l'ADN est soumise à des variations structurelles de grande taille impliquant parfois des protéines support de l'information ; ou des variations nucléotidiques, c'est-à-dire d'échelle comparable au nucléotide lui-même.

Dès le XXe siècle, des processus moléculaires et cellulaires impliqués dans le vieillissement ont été identifiés. Au moins quatre ont été particulièrement étudiés dans divers modèles expérimentaux [11] :

- L'inhibition de l'axe insuline/IGF-1
- La production des espèces réactives de l'oxygène
- Le raccourcissement des télomères
- L'autophagie dans les lysosomes.

Le raccourcissement des télomères implique des mécanismes de variations génomiques de grande taille, type variant de structure. Tandis que, inhibition de l'axe insuline/IGF-1, autophagie dans les lysosomes et production des espèces réactives de l'oxygène sont associés majoritairement à des variants nucléotidiques.

Les variants du génome sont impliqués dans des perturbations de la synthèse, l'épissage, la maturation, le transport ou le fonctionnement de protéines et peuvent être responsables de MNDREM.

✓ Les Single Nucleotide Variants, SNVs

Le génome humain possèderait 3.10 [9] de paires de base permettant l'encodage de tout au plus 30 000 gènes dont les fonctions et fonctionnement sont loin d'être tous identifiés. A l'échelle d'un génome entier, on retrouve plus de trois millions de variations nucléotidiques réparties comme sur le FIGURE 1 ci-contre.

Ils sont pour la moitié situés dans des régions génomiques non-géniques. L'autre moitié affecte les gènes et se répartit tout le long de la structure de celui-ci. Ainsi, ces variations se situent dans les régions promotrices, introniques (affectant ou non les sites d'épissage ou les sites de régulation de l'épissage) ou 3'/5'-untranslated region (UTR).

Les SNVs, anciennement SNPs (Single Nucleotid Polymorphism) sont des variants de séquence concernant un seul nucléotide, les impacts fonctionnels sont les suivants :

SNP Type	No. of SNPs
Nongene	2,255,102
Gene	1,165,204
Intron	1,064,655
Promoter	60,075
3′ UTR	16,350
5′ UTR	3,517
Splice regulatory site	2,089
Splice site	112
Synonymous	9,337
Stop→stop	17
Nonsynonymous	9,069
Stop→gain	121
Stop→loss	27
Total	3,420,306

FIGURE 1, *types de Single Nucleotids Polymorphisms du génome humain*

Missense, qui transforment l'acide aminé.

Synonymes, qui ne changent pas l'acide aminé. Mais des variants synonymes peuvent être pathogènes.

Frameshift, qui impacte le cadre de lecture et peut être « disruptive » ou « conservative » du cadre, c'est-à-dire interrompre ou non la phase de traduction.

Intronique, situé dans l'intron en 3' ou 5' UTR, affectant ou non l'épissage.

Indel, insertion ou délétion de 1, 2 ou plus de nucléotides. Dont les conséquences directes peuvent être le décalage du cadre de lecture.

L'American College of Medical Genetics and Genomics (ACMG) rappelle dans ses recommandations de 2015 de ne plus utiliser les termes « mutation » ou « polymorphisme » mais seulement le terme « variant » qui sera interprété selon le niveau de pathogénicité. Un variant est donc une modification par rapport à la séquence de référence et les variants sont repartis en cinq classes différentes, allant de « pathogène » à « bénin ».

✓ Les Variants Structurels (SV)

Sont des variants de nombre de copies de plus de 50 bp dans les séquences génomiques. Parmi lesquels :

Les déséquilibres chromosomiques

Translocation et inversion péri- ou paracentrique déséquilibrées, aneuploïdies, sont autant d'anomalies mises en évidence par les techniques de cytogénétique constitutionnelle et pouvant être impliqués dans des syndromes neurodéveloppementaux associés ou non à des phénomènes de dégénérescence/neurodégénérescence.

Les Copy Number Variants, CNVs

Sont des variants structurels pouvant s'étendre de quelques kilo- à plusieurs méga-bases. Ce sont des délétions ou des duplications de fragment de séquence pouvant être orientés (inversés ou non). Leur présence peut être silencieuse ou pathogène. Traditionnellement, ils sont détectés par les techniques de cytogénétique moléculaire. En particulier, l'Analyse Chromosomique sur Puce à ADN (ACPA ou Comparative Genomic Hybridization, CGH array). La recherche de CNVs par le NGS demande une adaptation particulière des pipelines bioinformatiques et judicieuse de la technique.

Motifs répétés ou tandem repeats

Caractéristiques communes aux génomes eucaryotes et procaryotes sont des mots de 3 à 6 nucléotides susceptibles de s'étendre par glissements de l'enzyme de réplication à chaque cycle cellulaire. En principe, la pathogénicité varie avec la taille de l'expansion. On définit graduellement une taille d'expansion normale - sans expresssion phénotypique, une zone grise - pour laquelle les perturbations biologiques sont observables de façon peu fréquente ou paucisymptomatique, et une expansion de taille pathologique – responsable de la maladie.

Les motifs répétés peuvent se trouver dans les régions codantes ou non codantes du génome. On imagine facilement que les tandems repeats des séquences codantes génèrent des protéines toxiques ou défectueuses par l'ajout de tracts d'acide aminés supplémentaires susceptibles de compromettre les structures et fonctions du gène et ou de son expression.

Mais on sait aujourd'hui, que les répétitions non codantes peuvent également avoir des effets significatifs en :

- Générant de la fragilité chromosomique,
- Inhibant la fonction du gène dans lesquels ils se trouvent,
- Perturbant la transcription et la traduction,
- Et en provoquant la séquestration des protéines impliquées dans des processus tels que l'épissage ou l'architecture cellulaire. [12]

Les variants pathogènes par expansion de motifs répétés sont étonnement fréquents dans les maladies neurodégénératives. En 2001, Lebre et Brice rapportaient que 9 pathologies humaines neurodégénératives sont causées par ce

type de variants pathogènes [13] : maladie de Huntington (HD), six formes d'ataxies cérébelleuses autosomiques dominantes (ADCA), l'atrophie dentato-rubro-pallido-luysienne (DRPLA) et la maladie de Kennedy ou amyotrophie spino-bulbaire (SBMA).

En se basant sur les conclusions expérimentales de modèles animaux, ainsi que sur l'observation phénotypique de patients présentant des délétions ou des variants pathogènes ponctuels de gènes d'intérêt, ils formulent alors diverses hypothèses sur une physiopathologie commune à ce type de variants pathogènes et énoncent des étapes physiopathologiques : « *les mutations par expansion de triplets seraient des mutations gain de fonction conduisant à un recrutement de partenaires protéiques inadaptés, générant des inclusions cellulaires nucléaires ou cytoplasmiques jouant un rôle dans la pathogénicité* ».

Cette dernière a plusieurs caractéristiques : mort neuronale, perturbation des fonctions nucléaires, implication des protéines chaperon, implication de la voie de dégradation ubiquitine-protéasome.

1.1.2. Apoptose

L'apopotose, du grec : apo, « au loin » et ptosis, « chute », est le processus par lequel les cellules déclenchent leur autodestruction. Les caspases, de l'anglais Cysteinyl-Aspartate-Cleaving Proteases, sont un groupe de protéases à cystéine effectrices de l'apoptose.

Actuellement, 3 voies de signalisation schématisées sur la FIGURE 1, sont connues pour mettre en jeu l'activation des caspases :

- La voie intrinsèque ou mitochondriale
- La voie extrinsèque
- La voie de l'Apoptose Inducing Factor (AIF)

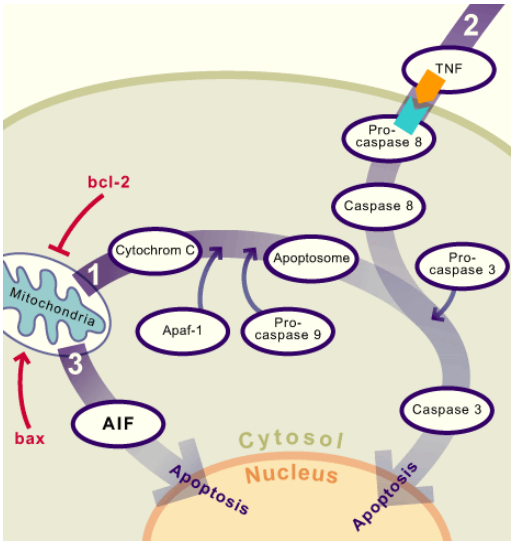


FIGURE 1, *voie de signalisation et d'activation des caspases responsables de l'apoptose* [4]

Les processus apoptotiques sélectifs se traduisent par une atrophie locale. La perte d'une certaine densité de population cellulaire entraîne des symptômes dépendant de la fonction de cette population. Lorsque l'apoptose atteint des populations cellulaires fonctionnellement impliquées dans l'activité neurologique, elle signe la neurodégénérescence. L'atrophie corticale est ainsi observée dans la maladie d'Alzheimer par exemple.

Il existe une balance entre des facteurs pro-apoptotiques et des facteurs anti-apoptotiques. Ainsi, le déclenchement de certaines maladies pourrait être le reflet du dérèglement de cette balance apoptotique.

Dans l'état actuel des connaissances, les résultats des études montrent que la mort des neurones dans les MND se fait par apoptose, sans que l'on sache si une dérégulation de l'apoptose est à l'origine des maladies, ou si l'apoptose n'est qu'une conséquence d'autres facteurs qui la déclenche.

1.1.3. Système Ubiquitine-protéasome : accumulation protéique

Le système ubiquitine/protéasome est un système multi-enzymatique sophistiqué qui joue un rôle majeur dans la vie de la cellule : élimination des protéines anormales ou surnuméraires, maturation de précurseurs en protéines biologiquement actives et approvisionnement en peptides antigéniques des molécules du complexe majeur d'histocompatibilité, destruction nécessaire et réglée de nombreuses protéines contrôlant la prolifération, la différenciation, l'apoptose et la réponse aux stimuli extracellulaires.

Le système fonctionne en 2 grandes étapes. La 1^{e} consiste au marquage de la molécule ciblée par conjugaison covalente d'Ubiquitine (Ub) enchaînée les unes aux autres. L'Ub est une petite protéine ubiquitaire très conservée de 76 Acides Aminés (AA). Ces chaînes permettent ensuite la reconnaissance et la dégradation des molécules polyubiquitinylées par un complexe protéolytique de 2000 KDa, le protéasome 26S. Cette chaîne poly-Ub est une simplification remarquable du problème de l'identification de la grande diversité des protéines à dégrader. Il s'agit d'un processus inédit développé par les eucaryotes. La 2nde consiste en la dégradation en propre.

La FIGURE 3, ci-après schématise le processus d'ubiquitynilation médié par un ensemble de protéine de conjugaison répartie en 3 familles :

- E1 pour les « Ubiquitin-activating enzymes »,
- E2 pour les « Ubiquitine-conjugating enzymes » ou Ubc
- et E3 pour les « Ubiquitine-protein ligase ».

Un dernier composant, dit E4, jouerait un rôle dans l'allongement des chaînes de poly-Ub de certains substrats.[14]

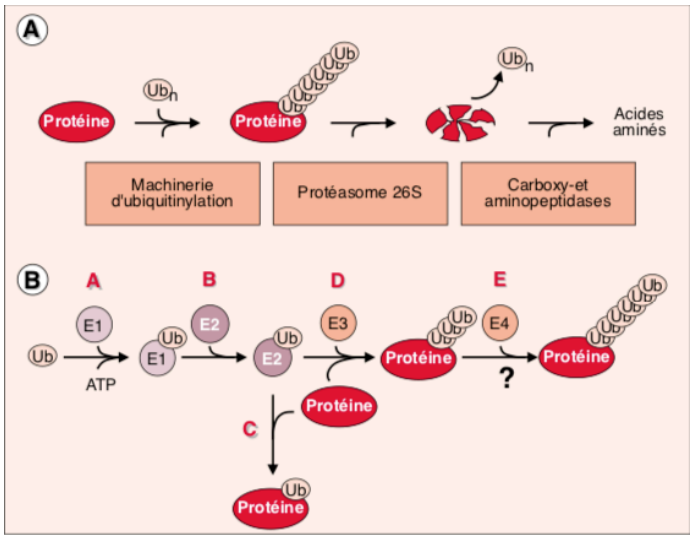


FIGURE 3 - *Schéma tiré de la littérature, montrant le processus d'Ubiquitinylation/ dégradation*

LEGENDE

A. Schéma des étapes de dégradation médiées par la machinerie d'ubiquitinylation détaillée en B., et le protéasome 26S capable d'hydrolyser les protéinés marquées en peptides de 3 à 20 AA. Ils sont ensuite dégradés en AA par des carboxypeptidases et aminopeptidases.

B. Schématise l'ubiquitinylation médiée par E1, ubiquitin-activating enzyme qui se fixe à l'Ub par liaison thioester (A) pour la transférer à E2, ubiquitin-conjugating enzyme ou Ubc, liée à l'Ub par liaison thioesther également (B). E2 ainsi conjugué paut assurer seul le marquage de la protéine par transfert de Ub (C), mais en général, un 3e facteur E3, ubiquitin-protein ligase, est indispensable à l'ubiquitinylation efficace (D). E4 jouerait un rôle dans l'allongement des chaînes poly-Ub.

Lorsque, sous certaines contraintes physiologiques, des protéines adoptent des configurations inadaptées, elles peuvent être la cible du système Ubiquitine-protéasome. Cependant, certains facteurs peuvent perturber le système et provoquer l'accumulation protéique. Ce sont : l'âge, les variants génétiques pathogènes, des modifications de l'environnement intracellulaire comme les variations de pH, le stress oxydant ou encore les métaux lourds.

Les protéines accumulées ont généralement tendance à s'agréger pour former des oligomères, qui polymérisent eux-mêmes en fibrilles insolubles. Ces fibrilles forment des inclusions intranucléaires ou intracytoplasmiques.

Dans la maladie d'Alzheimer, elles forment des agrégats extracellulaires appelés plaques séniles [15] ; De même, il a été largement démontré que la présence de ces agrégats de protéines mal repliées sont corrélés avec la progression des Maladies Neuro-Dégénératives (MND) : protéines Tau dans la dégénérescence fronto-temporale, corps de Lewy dans la maladie de Parkinson, protéine à prions dans la maladie de Creutzfeld-Jacob.[16].

Les mécanismes par lesquels la toxicité des agrégats cellulaires est véhiculée restent encore obscurs. Trois hypothèses peuvent expliquer cette toxicité [17]:

- L'hypothèse de la perte de fonction : la mort neuronale serait causée par la perte de l'activité des protéines mal repliées et agrégées.
- L'hypothèse du gain de fonction : les protéines mal repliées et agrégées gagnent une fonction neurotoxique. Par exemple, leur accumulation extracellulaire pourrait engendrer l'interaction avec des récepteurs et déclencher une mort cellulaire programmée.
- L'hypothèse de l'inflammation : l'agrégation des protéines provoquerait une réaction inflammatoire au sein du système nerveux qui serait, à la longue, délétère pour les neurones.

Il semblerait que dans les MND, les agrégats extracellulaires pourraient servir de signal pour déclencher l'apoptose des neurones.

Actuellement, ni la sélectivité des neurones touchés, ni la sélectivité des zones du système nerveux endommagées ne sont expliquées.

1.1.4. Micro-ARN

Les microARNs (miRNA) sont une catégorie de petits acides ribonucléiques, simple-brin, non-codants et propres aux cellules eucaryotes. Le premier miARN, lin-4, a été caractérisé au début des années 1990 chez le nématode *Caenorhabditis elegans*. Ils sont appelés ainsi à cause de leur longueur : 21 à 24 nucléotides en général.

Les miARNs sont des régulateurs post-transcriptionnels capables d'éteindre l'expression d'un gène par appariement à l'ARN messager (ARNm) cible. Cet appariement compétitif entraîne l'inhibition de la traduction par la dégradation de l'ARNm. Cela dépend du degré de complémentarité entre la séquence du miARN et celle de son ARNm cible.

Parce qu'ils affectent l'expression de nombreux gènes, les miARNs interviennent dans la plupart des processus biologiques. En particulier, ils participent à la croissance et la différenciation cellulaire, l'apoptose et le métabolisme. En plus de la méthylation de l'ADN, les modifications des histones et le remodelage chromatinien, les miARNs sont des facteurs épigénétiques capable de moduler l'expression des gènes.

Au cours du développement, comme dans les pathologies humaines, des miARN circulent dans le milieu extracellulaire et représentent des marqueurs d'évolution d'une condition physiologique donnée [18]. Les miARN circulants sont principalement médiés par quatre types de structures différentes associés ou non à des complexes lipidiques :

- Les exosomes
- Les microvésicules, plus grandes que les exosomes et issues du bourgeonnement de la membrane plasmique
- Les corps apoptotiques qui se forment lors des stades précoces de l'apoptose
- Les lipoprotéines (HDL, high density lipoprotein et LDL, low density lipoprotein).

L'association étroite entre contextes pathologiques et taux de certains miARN dans la circulation fait de ces derniers d'excellents candidats biomarqueurs pour la clinique, et les études se multiplient depuis 2008 pour valider de telles associations.

Des dérégulations dans leur biogenèse ou leurs fonctions sont impliquées dans le vieillissement et dans de nombreuses pathologies humaines, en particulier les MND. [19]

1.1.5. Atteintes mitochondriales : métabolisme et stress oxydatif

La mitochondrie est un organite essentiel à la survie cellulaire. Elle possède son propre ADN, l'ADNmt qui est double brin et circulaire. Les mitochondries sont d'origine maternelle car absentes des spermatozoïdes.

Leur répartition dans l'espace cellulaire post-mitotique est aléatoire. On parle d'hétéroplasmie quand il existe plusieurs types de génomes mitochondriaux dans une cellule ou un tissu.

La mitochondrie se compose de molécules codées à la fois par des gènes nucléaires, à la fois par des gènes mitochondriaux. Elle est le lieu du métabolisme des molécules organiques. Elle permet la production d'Adénosine Triphosphate (ATP) grâce à la chaîne respiratoire et à partir du dioxygène. Elle fournit également divers cofacteurs métaboliques comme le Nicotinamide Adénine Hydrogénase (NADH) et le Flavine Adénine Dinucléotide Hydrogénase (FADH) à partir du cycle de Krebs. (FIGURE 4).

La mitochondrie est donc impliquée dans différents processus tels que la communication cellulaire, la différenciation, l'apoptose et la régulation du cycle cellulaire. Les mitochondries jouent un rôle important dans le processus de vieillissement.

Des dysfonctionnements mitochondriaux sont associés à certaines maladies humaines telles que des retards mentaux, des atteintes cardiaques ou encore des myopathies.

Les produits du métabolisme du dioxygène, appelés Espèces Réactives Oxygénées (ROS en anglais), sont très toxiques pour la cellule. Ces composés ont un potentiel redox très important et peuvent donc facilement oxyder les molécules avoisinantes comme les protéines, les lipides ou l'ADN, ce qui peut être délétère. Il existe normalement un système de protection qui permet d'éliminer ces composés toxiques.

Dans les MND, le système de protection est débordé et n'arrive plus à contrer la production de ROS qui s'accumulent dans les neurones et les conduisent lentement vers leur mort, notamment par le mécanisme d'apoptose. Plusieurs marqueurs du stress oxydant ont notamment été retrouvés dans certaines MND. Là aussi, de l'origine ou la conséquence, le stress oxydant n'est pas encore classé. [20]

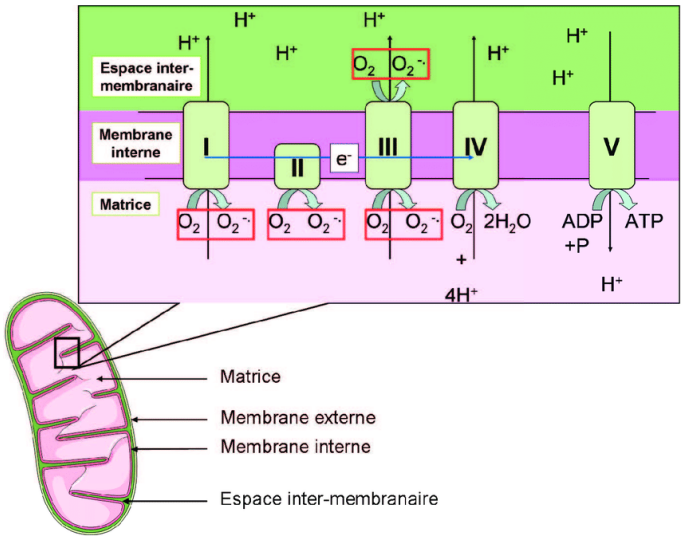


FIGURE 4, *schéma de la production d'espèces réactives mitochondriale par les complexes I à III*

1.2. Organisation anatomique et cellulaire du système locomoteur

L'activité motrice est la manifestation de l'ensemble des fonctions neurologiques et musculaires permettant le déplacement ou le mouvement de tout ou partie du corps. L'activité motrice du corps peut être volontaire ou autonome.

La marche est une activité motrice volontaire, automatique et réflexe impliquant de nombreuses structures anatomiques du système nerveux central et périphérique et nécessitant l'intégrité de l'appareil ostéoarticulaire.

La mise en jeu de l'appareil musculosquelettique, principal effecteur de la marche, est dépendante de l'activation de structures spinales, supraspinales et sous-cortico-corticales du Système Nerveux Central (SNC).

1.2.1. Structures du système nerveux impliquées

Les principales structures impliquées dans l'organisation de la motricité, en particulier la marche, sont représentées dans la FIGURE 5 suivante, et elles sont décrites ci-après. Tandis que la FIGURE 6 montre le schéma anatomique de l'organisation de la transmission du signal moteur, du cortex à la moelle épinière en passant par le tronc cérébral [21].

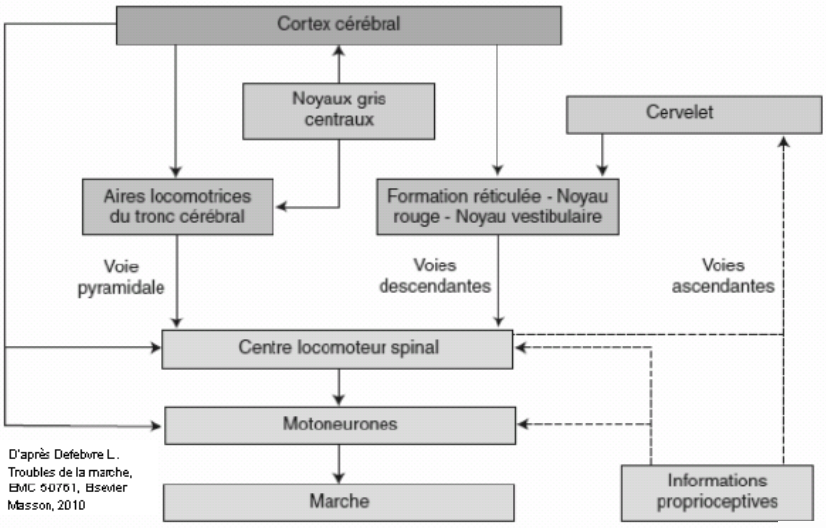


FIGURE 5, *Principales structures impliquées dans l'organisation de la marche et du mouvement* [21]

Le cortex moteur désigne l'ensemble des aires du cortex cérébral qui participent à la planification, au contrôle et à l'exécution des mouvements volontaires des muscles du corps.

Les noyaux gris centraux sont un ensemble de structures sous-corticales constitué par des noyaux pairs, interconnectés. Leur structure peut varier selon qu'on les définit de manière anatomique ou fonctionnelle.

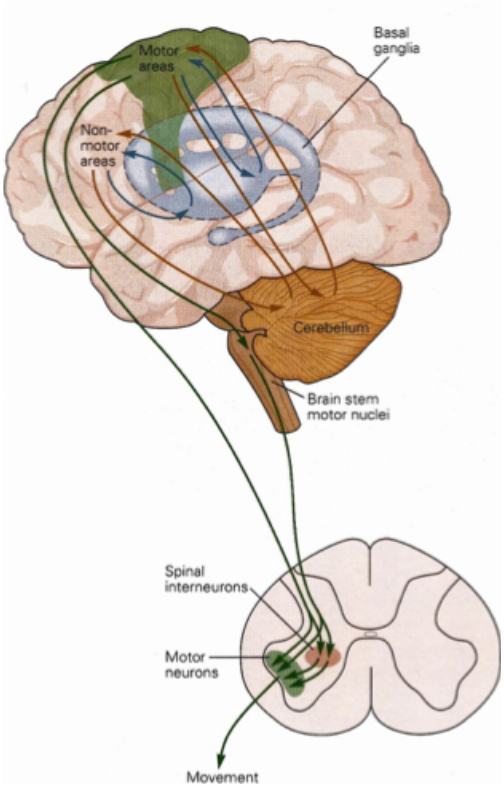


FIGURE 6 - *Schéma anatomique de l'organisation de la transmission du signal moteur, du cortex à la moelle épinière en passant par le tronc cérébral* [141]

Au sens purement anatomique ils comprennent quatre noyaux pairs : le noyau caudé, le noyau lenticulaire correspondant au putamen et au pallidum, le noyau sous-thalamique, la substance noire.

Au sens fonctionnel on compte : le striatum composé du noyau caudé et du putamen (correspondant à la partie latérale du noyau lenticulaire), le globus pallidus interne et le globus pallidus externe, correspondant au pallidum (partie médiane du noyau lenticulaire), le noyau sous-thalamique, la substance noire compacte, et la substance noire réticulée.

Ces noyaux participent à des réseaux neuronaux disposés en boucles parallèles motrices, oculomotrices, cognitives et limbiques.

La formation réticulée est une structure du tronc cérébral à l'interface des systèmes autonome, moteur et sensitif. Elle intervient dans la régulation de grandes fonctions vitales comme les cycles veille-sommeil, le contrôle d'activités motrices réflexes ou stéréotypées, comme la marche ou le tonus postural et dans des fonctions cognitives telles que l'attention. D'un point de vue phylogénétique, la formation réticulée est l'une des plus anciennes structures nerveuses du tronc cérébral, lui-même étant ce qu'on peut considérer comme un des centres les plus primitifs du système nerveux central. La formation réticulée est située sur toute la longueur du tronc cérébral entre les grands faisceaux ascendants et descendants.

Le noyau rouge est une structure paire située dans le tegmentum mésencéphalique, réseau de neurones multi-synaptiques du tronc cérébral. Il reçoit des afférences en provenance du cervelet et du cortex cérébral sensorimoteur. Il est à l'origine du faisceau rubro-spinal, lui-même impliqué dans le contrôle du tonus de posture et dans la coordination des muscles distaux des membres supérieurs et proximaux des membres inférieurs. Chez l'homme, le faisceau rubro-spinal reste vestigial, et le noyau rouge se projette essentiellement sur l'olive bulbaire principale. L'olive à son tour émet des axones vers le noyau dentelé et le cortex cérébelleux.

Les noyaux vestibulaires sont des structures de projection des fibres vestibulaires dans le tronc cérébral. Ils reçoivent des informations, non seulement du système vestibulaire, mais également des récepteurs visuels, des extérocepteurs cutanés, des propriocepteurs ostéoarticulaires et musculo-tendineux, et du cervelet. En réponse à ces influx divers, ils transmettent des informations aux motoneurones somatiques par la voie vestibulo-spinale, aux motoneurones oculo-moteurs par la voie vestibulo-oculaire, au cortex vestibulaire par les voies vestibulo-corticales et aux neurones du système neuro-végétatif. L'ensemble de ces connexions et des réponses musculaires qui en résultent assure ainsi le maintien de la posture et de la vision nette durant les déplacements du corps et de la tête.

Formation réticulée, noyau rouge et noyaux vestibulaires sont à l'origine des voies informatives descendantes.

Le faisceau pyramidal - Les aires locomotrices du tronc cérébral sont à l'origine du faisceau pyramidal. Celui-ci désigne les axones moteurs cortico-spinaux qui transmettent la commande motrice du cortex cérébral jusqu'aux motoneurones et interneurones de la corne antérieure de la moelle épinière. Il intervient préférentiellement dans le contrôle de la motricité volontaire.

Le cervelet, du latin cerebellum, petit cerveau, est la structure responsable du contrôle de l'équilibre, coordonne le tonus postural et les mouvements volontaires en intégrant les informations sensorielles essentiellement visuelles et auditives.

Le centre locomoteur spinal se dit du réseau de neurones localisés dans la moelle spinale et responsable de la locomotion. Il peut fonctionner de manière autonome, indépendamment des commandes descendantes et des retours sensoriels. Après avoir été activé par le cortex moteur ou d'autres régions supraspinales, ce réseau peut générer à lui seul l'activité locomotrice.

Les motoneurones constituent la voie finale de tout acte moteur. Les corps cellulaires des motoneurones sont situés soit dans le tronc cérébral, soit dans la corne ventrale de la substance grise de la moelle épinière. Chaque motoneurone possède un axone qui part du système nerveux central pour innerver les fibres musculaires d'un muscle. L'ensemble constitué par un motoneurone et les fibres musculaires qu'il innerve constitue une unité motrice.

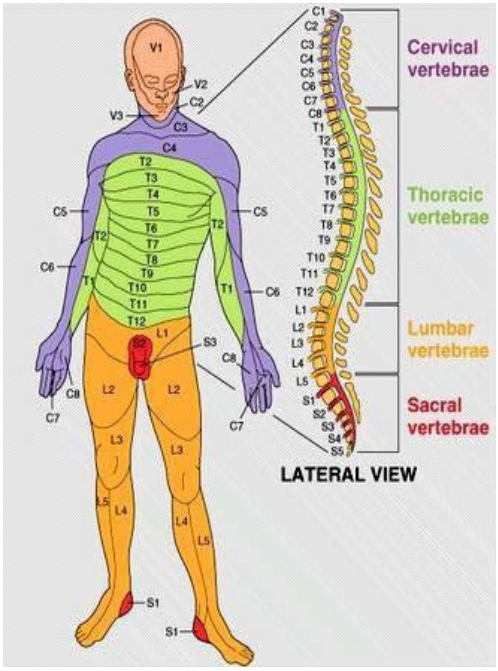


FIGURE 7 - *Territorialisation des vertèbres du rachis*

L'appareil ostéoarticulaire est l'effecteur du mouvement. L'intégrité des muscles et des articulations qui le compose est également indispensable. Dans ce registre, le rachis participe au maintien et la posture. Organisé en 4 zones : cervicale, thoracique, lombaire ou pelvienne. Cette organisation traduit également une innervation géographiquement ordonnée. Les ramifications nerveuses qui commandent les membres supérieurs sont partagées entre les vertèbres cervicales et thoraciques ; les lombaires sont responsables de la communication aux membres inférieurs.

La FIGURE 7 ci-contre montre la territorialisation des vertèbres du rachis [22].

1.2.2. Organisation cellulaire

L'encéphale est un organe hautement hiérarchisé et parfaitement structuré. Sur le plan macroscopique, on peut distinguer 3 structures : le cerveau, le cervelet et le tronc cérébral.

Un 1e niveau de structuration en lien avec l'aspect de cet organe. Il est indiqué par la substance grise, qui contient les corps cellulaires et l'arbre dendritique des neurones ainsi que de certaines cellules gliales ; et la substance blanche, principalement composée des axones myélinisés des neurones.

Le TABLEAU 1, ci-après, résume la nature et la fonction des types cellulaires distingables sur le plan histologique, au niveau de l'encéphale et de ces structures.

CELLULES	FONCTION	NATURE
Les **neurones**	Véhiculent l'information	9 types : Les cellules horizontales, de Cajal-Retzius, Cellule de Martinotti, Cellule pyramidale, Grande cellule pyramidale, Cellule à double bouquet, Cellule fusiforme, Cellule étoilée, Cellule névroglique Neurones spécifiques des structures cérébrales : Cellules granulaires du CERVELET[1] Cellules du PLEXUS CHOROÏDES[2]
Cellules de la **névroglie** ou cellules gliales interstitielles	Assurent le soutien du tissu nerveux	Astrocytes, support et protection des neurones, défense immunitaire, réparation cicatrisation cérébrale ; Microglie, macrophages résidants du cerveau.
Cellules **d'isolement électrique**	Jouent un rôle dans la myélinisation et l'isolement du tissu nerveux	Oligodendrocytes Cellules de Schwann, système périphérique uniquement)

***TABLEAU 1**, organisation des cellules de l'encéphale*

([1] les plus petits neurones du cerveau réunis en densité neuronale la plus marquée dans le cerveau ; [2] sont des villosités qui font saillie dans les cavités ventriculaires)

Le cortex sensorimoteur est lui-même organisé en couches neuronales successives. Chaque couche étant composée de nombre et nature différentes de neurones. Comme dans tout système, le fonctionnement du système nerveux impliqué dans la motricité et dans la marche, est rendu possible par les mécanismes de communication cellulaire.

Elle se traduit d'une part par la conduction électrique, d'autre part par la transmission chimique : La conduction électrique est possible par propagation de dépolarisation cellulaire suite aux échanges ioniques dans l'axone. La transmission chimique a lieu dans la fente synaptique et est médiée par les neurotransmetteurs.

Le TABLEAU 2, qui suit, résume les principaux neurotransmetteurs du système nerveux central sans tenir compte de l'ensemble des neuropeptides qui participent également à la communication inter-neuronale ; comme la substance P (excitatrice) ou l'endorphine (généralement inhibitrice).

NEUROTRANSMETTEURS	NATURE	FONCTION	ROLES
Acétylcholine	Excitateur très répandu	Déclenche la contraction musculaire et stimule l'excrétion de certaines hormones	EVEIL, ATTENTION, COLERE, AGRESSION, SEXUALITE ET SOIF.
Dopamine	Excitateur/ inhibiteur	Contrôle du mouvement et de la posture.	Module aussi L'HUMEUR et joue un rôle central dans le RENFORCEMENT POSITIF et la DEPENDANCE comportementale
Noradrénaline	Excitateur/ inhibiteur	Libérée comme une hormone dans le sang, elle contracte les vaisseaux sanguins et augmente la fréquence cardiaque.	ATTENTION, EMOTIONS, SOMMEIL, REVE ET APPRENTISSAGE
Sérotonine	Généralement inhibiteur	Libéré également dans le tube digestif	Régule la TEMPERATURE, le SOMMEIL, L'HUMEUR, L'APPETIT et la DOULEUR.
Acide gamma-aminobutyrique (GABA)	Inhibiteur très répandu au niveau de l'encéphale	Contribue au contrôle moteur, à la vision et à plusieurs autres fonctions corticales	Régule aussi L'ANXIETE
Acide Aspartique	Excitateur	Active les récepteurs au Glutamate	Impliqué dans les processus MEMOIRE
Glutamate	Excitateur majeur	Le plus important du SNC	Associé à L'APPRENTISSAGE et la MEMOIRE.
Glycine	Inhibiteur	Ouvre les canaux à Cl^- ligands-dépendants, et permet la relaxation des muscles antagonistes	Contrôle de L'ADAPATATION

***TABLEAU 2**, principaux neurotransmetteurs du SNC*

1.3. Manifestations cliniques de dégénérescence neurologique

La perte fonctionnelle induite par un dysfonctionnement cellulaire ou moléculaire au niveau d'un territoire du système nerveux impliqué dans la fonction motrice, induit des symptômes spécifiques. L'examen neurologique permet de les détecter. Nous listons ci-après les signes de la neurodégénérescence décelable à l'examen clinique.

1.3.1. Force et motricités

✓ Atteinte bulbaire et atteinte pyramidale

Le syndrome bulbaire est un ensemble de signes qui traduit une atteinte du bulbe rachidien, c'est-à-dire la partie inférieure du tronc cérébral qui est en continuité avec la moelle épinière. Les symptômes sont le témoin d'une lésion des différentes structures nerveuses contenues à l'intérieur du bulbe rachidien.

Il s'agit notamment de :

- Une hémiplégie sans atteinte du visage.
- Une hémianesthésie c'est-à-dire une perte de la sensibilité de la moitié du corps qui, ici également, ne concerne pas le visage.
- Un syndrome cérébelleux qui se traduit par l'apparition de troubles de l'équilibre, des mouvements désordonnés, des tremblements, et des troubles de la coordination (ataxie).
- Des bourdonnements d'oreille, des vertiges en lien avec une atteinte du nerf cochléo-vestibulaire.
- Un syndrome de Claude Bernard Horner (rétrécissement de la pupille, un enfoncement du globe oculaire à l'intérieur de l'orbite, une chute de la paupière supérieure).
- Des troubles de la déglutition qui sont le résultat d'une paralysie du pharynx, à cause d'une atteinte du nerf glossopharyngien.
- Une paralysie de la moitié du voile du palais et du larynx, associée le plus souvent à une hémianesthésie du voile du palais.
- Une paralysie des muscles de la moitié de la langue.
- Une paralysie des muscles trapèze, et des muscles sterno-cléido-mastoïdiens.

L'atteinte pyramidale est une atteinte de la voie cortico-spinale (faisceau pyramidal), support de la commande motrice volontaire. Elle se caractérise notamment par l'association de signes déficitaires et de signes de spasticité :

- Déficit moteur pouvant intéresser l'hémicorps ou la face
- Un syndrome pseudobulbaire (dysarthrie paralytique).
- Trouble du tonus musculaire
- Anomalies des Réflexes Ostéotendineux (ROT) et réflexes cutanés

✓ Troubles de la marche, de l'équilibre et de la coordination

La variété des troubles de la marche et de l'équilibre observée dans les affections neurologiques et ostéoarticulaires, rend compte de la complexité et de la diversité des mécanismes impliqués dans cette fonction essentielle. Comme la marche et la station debout qui mobilisent plusieurs systèmes, entre autres : corticospinal, cérébelleux, vestibulaire, sensitif profond - proprioceptif.

L'orientation diagnostique devant un trouble de la marche ou de l'équilibre s'appuie essentiellement sur l'interrogatoire et l'examen clinique. On examine la vitesse, l'amplitude et la régularité du pas, l'orientation (rectiligne ou non), le demi-tour (harmonieux ou décomposé), le ballant automatique des bras. Les troubles de la coordination motrice sont évalués sur les membres inférieurs et supérieurs. Ils résultent d'une anomalie fonctionnelle du cervelet, la zone du cerveau qui coordonne les mouvements volontaires et contrôle l'équilibre. Les troubles de la marche sont malgré tout stéréotypés.

Le TABLEAU 3 suivant répertorie les types de marche en regard de leur étiologie ou contexte physiopathologique :

TYPE DE MARCHE	ETIOLOGIES
Douloureuse	Origine et circonstance variées
Ataxique	Cérébelleuse, vestibulaire, proprioceptive
Déficitaire	Centrale, périphérique
Hyperkinétique (mouvements anormaux)	Dystonie, chorée, tremblement (Essentiel ou orthostatique)
Hypokinétique (à petits pas)	Syndromes parkinsoniens, hydrocéphalie à pression normale, états lacunaires
Psychogène	Marche précautionneuse, phobie de la marche somatoforme et simulations

TABLEAU 3, *Classification des troubles de la marche* [23]

✓ Altération de la force et du tonus musculaire

La force musculaire est la capacité d'un muscle à résister à une force. Elle est évaluée cliniquement selon le British Medical Research Council System sur une échelle allant de 0 à 5 :

- 0 : aucune contraction musculaire
- 1 : contraction musculaire sans mouvement
- 2 : mouvement dans le plan du lit
- 3 : mouvement contre la gravité
- 4 : mouvement contre résistance
- 5 : force musculaire normale.

Le recrutement des cellules musculaires pour effectuer une action motrice nécessite la participation des aires cérébro-spinales locomotrices et la plaque neuromusculaire qui fait jonction entre système nerveux et muscle.

Le tonus musculaire est l'état de tension minimale permanente des muscles de maintien du corps. Il existe deux principales sortes d'hypertonie :

- Spastique traduisant un syndrome pyramidal
- Rigide de type syndrome parkinsonien.

Une hypotonie peut être présente dans les déficits moteurs périphériques, à la période initiale d'une lésion de la voie cortico spinale ou lors du syndrome cérébelleux.

✓ Motricité de la face et oculomotricité

L'oculomotricité peut être évaluée par un testing volontaire, automatique ou réflexe. La motricité de la face au repos et lors des mimiques peut renseigner sur la qualité des nerfs crâniens et des voies visuelles.

✓ Phonation, déglutition et motricité linguale

Phonation, déglutition sont perturbées dans les atteintes mixtes des nerfs IX, X et XI. Pour la motricité linguale, c'est le nerf XII.

1.3.2. Sensorialité et réflexes

✓ Perturbation des Réflexes Ostéo-Tendineux (ROT)

Le réflexe est une réponse musculaire involontaire et stéréotypée. Le réflexe ostéotendineux (ROT), aussi appelé réflexe d'extension, est la contraction d'un muscle d'étirement opérée par la voie monosynaptique excitatrice (réflexe myotatique) et la voie polysynaptique qui entraînent le relâchement du muscle antagoniste.

L'examen des ROT traduit l'état de la moelle spinale et des structures cérébrales supérieures. Les réflexes sont symétriques et de vivacités variables selon les sujets. Ils peuvent être abolis, pendulaires ou exagérés. Leur vivacité isolée peut se voir chez certains sujets dits « neurotoniques » (tendus, anxieux). Quand d'autres éléments s'ajoutent à cette vivacité, cela traduit un syndrome pyramidal. Une aréflexie ostéotendineuse peut relever d'une cause centrale ou d'une interruption de l'arc réflexe (lésion des cornes antérieures de la moelle, lésion du nerf spinal, lésion de la racine postérieure).

Certains sujets ont de façon constitutionnelle des réflexes très faibles, ou pas de réflexes du tout, sans que cela soit pathologique. Le TABLEAU 4 suivant résume l'examen des ROT.

✓ Troubles de la sensibilité

L'altération de la sensibilité peut impacter les capacités motrices. L'évaluation de la sensibilité permet de statuer sur l'état de fonctionnement des grosses fibres (cordons postérieurs de la moelle et lemnisque médian), des petites fibres sensitives (fibres amyéliniques/faisceau spino-thalamique), cortex (pariétal).

✓ Troubles sphinctériens

Sont détectés essentiellement à l'interrogatoire. Ils se traduisent par des mictions difficiles, trop nombreuses, des troubles de l'érection ou du contrôle des selles. Ils mettent en jeu les vertèbres sacrées S2 à S4.

Réflexe	Stimulus	Résultat	Territoire neurologique ou rachidien
MEMBRE SUPERIEUR			
Bicipital	En percutant le pli du coude	Flexion de l'avant-bras	C5/C6
Stylo-radial	En percutant le styloïde ulnaire, avant-bras fléchi et détendu	Contraction du brachio-radial et flexion du coude	C6
Cubito-pronateur	En percutant la styloïde ulnaire, avant-bras demi-fléchi et en légère supination	Entraîne la pronation de l'avant-bras	C8
Tricipital	En laissant pendre l'avant-bras et percutant le tendon du triceps	Contraction du triceps et extension du coude	C7
Palmaire	En percutant la face du poignet	fFexion des doigts	C8
MEMBRE INFERIEUR			
Rotulien [1]	En percutant le ligament rotulien, genou fléchi et muscles de la cuisse relâchés	L'extension de la jambe par contraction du quadriceps	L3-L4
Achiléen	à genou, on percute le tendon d'Achille	Contraction du triceps et la flexion plantaire	S1
CUTANES PLANTAIRES			
Signe de Babinski [2]	En promenant un objet non piquant du talon vers les orteils	Replier les orteils	Faisceau pyramidal
Cutané abdominaux	En stimulant la paroi abdominale dans le sens transversal	Entraine la contraction des muscles sous-jacents	L'arc réflexe supérieur (T6 - T8), Moyen (T8 - T10), Inférieur (T10 - T12)
Cornéen	Obtenu par l'excitation douce de la cornée	Clignement de la paupière	Nerf ophtalmique, le centre protubérantiel et le nerf facial
Voile du palais	En attouchant l'hémi-voile	Contraction et élévation de l'hémi-voile	Nerf glosso-pharyngien, recrutement du centre bulbaire et efférence par le nerf vague.
Crémastérien et anal	Stimulation du tiers supérieur et antéro-médial de la cuisse /étirement du sphincter interne par un doigt intra-rectal	Élévation du testicule ou de la grande lèvre / contraction sphincter	L1-L2 et S4

[1] Ce réflexe traduit l'état de la moëlle spinale et des structures cérébrales supérieures. Le réflexe est exacerbé quand ces composantes sont endommagées.

[2] l'extension lente et majestueuse du gros orteil, parfois associé à un écartement en éventail des autres orteils traduit l'atteinte pyramidale

TABLEAU 4, *Examen des réflexes ostéotendineux et cutanés plantaires*

1.3.3. Fonctions cognitives

Depuis très longtemps, la littérature rapporte une interconnexion entre troubles moteurs et démence. Dès 1999, Franssen et al [24] justifient que les troubles de l'équilibre et les déficits moteurs sont d'autant plus marqués que l'état démentiel altère les mécanismes cérébraux compensatoires.

Par ailleurs, les lésions cérébrales dégénératives compromettent le traitement des informations venant des entrées multi-sensorielles impliquées dans l'équilibre et allongent le délai d'activation des réactions d'adaptation posturale en réponse aux perturbations de la posture [24]–[26].

Ainsi, la démence impacte la motricité. De même, les troubles de la marche sont parfois associés à des démences.

✓ Définition de la démence

Les troubles démentiels sont définis par l'Organisation Mondiale de la Santé (OMS). La démence consiste en « l'altération acquise progressive de la mémoire et d'au moins une autre fonction supérieure, suffisamment marquée pour handicaper les activités de la vie quotidienne, apparue depuis au moins six mois ». Cette dégradation porte au moins sur deux fonctions cognitives parmi :

- Les praxies (ensemble des fonctions de coordination et d'adaptation des mouvements volontaires de base dans le but d'accomplir une tâche donnée)
- Le langage
- La gnosie (reconnaissance de ce qui est perçu par les sens)
- Le jugement et le raisonnement
- Les pensées abstraites
- La mémoire

Dans les démences évoluées, la désorientation temporo-spatiale est un trouble presque systématique.

On distingue les démences dégénératives, qui résultent d'une dégénérescence des cellules nerveuses cérébrales et les démences non dégénératives, qui proviennent d'un agent ou d'un groupe de facteurs pathogènes de type vasculaire, toxique, traumatique ou tumoral. Les démences d'origine vasculaires sont les plus fréquentes.

✓ Troubles du comportement

Les troubles comportementaux sont fréquents chez les patients atteints de démence. Les symptômes psychologiques et comportementaux de la démence découlent des modifications fonctionnelles et physiologiques liées à la démence : désinhibition, mauvaise interprétation de signaux visuels et auditifs, altération de la mémoire à court terme, capacité réduite ou incapacité à exprimer des besoins.

Le Cohen-Mansfield Agitation Inventory est l'outil le plus fréquemment utilisé pour caractériser les troubles du comportement, sur la base d'un caractère physique ou verbal, agressif ou non [27].

2. Diagnostic des MND Rares à Expression Motrice (MNDREM)

Le diagnostic des MND à expression motrice repose sur l'examen clinique et les analyses complémentaires biologiques, fonctionnelles ou d'imagerie.

2.1. Caractérisation phénotypique

2.1.1. Examen clinique

La caractérisation des signes cliniques comporte l'examen de :

- De la marche et de l'équilibre
- Neurologique à la recherche d'un déficit moteur par l'étude des ROT et réflexes cutanés plantaires, la recherche d'une amyotrophie, la recherche d'un déficit sensitif, d'un syndrome cérébelleux, d'un syndrome vestibulaire périphérique ou central, d'un syndrome parkinsonien ou de troubles du mouvement, de troubles vésico-sphinctériens, de troubles cognitifs.
- Général : dermatologique, ostéoarticulaire, cardio-vasculaire, ophtalmologique.

Les étiologies d'un trouble de la motricité peuvent être :

- Motrices, en rapport avec des affections ostéoarticulaires
- Centrales, secondaires à des affections neurologiques
- Musculaires
- Directement liées à la régression psychomotrice secondaire, au vieillissement ou aux différentes complications qui font suite à un traumatisme.

Le diagnostic étiologique est déterminé par les données de l'interrogatoire et les résultats des examens complémentaires. Ces-derniers sont indiqués selon les hypothèses diagnostiques retenues en fonction de la présentation clinique. [28]

2.1.2. Bilan neuropsychologique

Il permet de statuer sur les niveaux d'atteinte cognitive.

Les démences sont classifiées selon des critères diagnostic fournis par le Manuel diagnostique et statistique des troubles mentaux (également désigné par le sigle DSM, abréviation de l'anglais : Diagnostic and Statistical Manual of Mental Disorders). Il s'agit d'un ouvrage de référence publié par la Société américaine de psychiatrie (APA) pour la caractérisation de troubles mentaux spécifiques.

Ainsi, le diagnostic de démence est établi selon les critères suivants :

- Déficits cognitifs multiples (mémoire + langage ou praxies ou gnosies ou fonctions exécutives)
- Déclin cognitif (fonctionnement antérieur)
- Début progressif et déclin continu
- Pas d'atteinte du SNC (AVC, Parkinson, HSB, hydrocéphalie, tumeur…), pas d'affection générale (hypothyroïdie, carence B12 ou folates, hypercalcémie…), pas de toxicité médicamenteuse, d'OH ou de stupéfiant
- Pas de confusion mentale
- Pas de dépression ou de psychose

Le concept de démence proposé par le DSM exclut : débilité mentale, déficit cognitif limité à une fonction, syndrome confusionnel, syndrome dépressif.

Le diagnostic repose sur les arguments d'évaluation des fonctions basées sur le Mini Mental State Examination (MMSE, Folstein 1975), qui permet la mesure du fonctionnement cognitif global et sa graduation selon l'échelle suivante :

- 20< score < 24 : DEMENCE LEGERE ;
- 10< score < 20 : DEMENCE MODEREE ;
- Score< 10 : DEMENCE SEVERE.

D'autres échelles d'évaluation existent et sont basées sur des tests type l'épreuve des cinq mots ou l'épreuve de Grober et Buschke.

2.2. Analyses fonctionnelles et imageries de référence

2.2.1. Electroneuromyogramme, ENMG

Il permet l'étude de l'activité électrique des nerfs et des muscles. Le tracé est un ensemble de potentiels traduisant les phénomènes électriques de polarisation/dépolarisation. Le tracé de repos indique l'activité des unités battant à basses fréquences. La contraction musculaire se traduit par l'augmentation de la fréquence des battements. On y associe l'étude des Vitesses de Conduction Nerveuses (VCN) motrices et sensitives. Les valeurs habituelles sont 50 m/s en moyenne. En cas d'atteinte d'origine nerveuse, on observe une accélération du tracé de repos. Les VCN sont fortement abaissées dans les atteintes axonales. On parle de tracé neurogène. Dans les atteintes musculaires (myopathie, myosite), le tracé de repos est sans particularité, mais à la contraction, il s'enrichit de façon rapide et les potentiels sont de petites tailles. On parle de tracé myogène.

L'interprétation de l'ENMG réalisée par un expert, aide à la détection de pathologies neurologique et/ou musculaire.

2.2.2. Imagerie par Résonnance Magnétique cérébrale et/ou médullaire, IRM

L'Imagerie par Résonnance Magnétique cérébrale (IRM cérébrale), permet de mettre en évidence des particularités anatomiques cérébrales. On peut y reconnaître des territoires atrophiés, mal organisés ou anormalement envahis par du liquide, comme dans les accidents vasculaires par exemple. On parle d'hyper signaux face à un contraste de teinte entre deux territoires. Ils signent une variation dans la composition tissulaire et permet de distinguer du liquide comme le sang ou des calcifications. L'imagerie cérébrale peut être fonctionnelle et permettre d'évaluer des niveaux d'activités cérébrales.

On a recours généralement à l'IRM, mais les scanners Tomo-Densito-Métrique (TDM) et Positrons Émission Tomography (PET-scan) sont parfois prescrits. L'IRM médullaire permettra de détecter des anomalies rachidiennes pouvant justifier des anomalies dégénératives et/ou de conduction nerveuse.

2.2.3. Biopsie musculaire

Il s'agit d'un examen invasif qui consiste à prélever un fragment musculaire ou neuromusculaire (jonction neuromusculaire) afin d'en étudier les caractéristiques histopathologiques. Des méthodes chimiques, enzymatiques ou immunologiques permettent de préciser la nature des perturbations tissulaires. Certains aspects histologiques sont caractéristiques de pathologies neuromusculaires.

On détecte ainsi des anomalies organisationnelles : inégalités de tailles des fibres musculaires, centralisations nucléaires, noyaux en chaînette etc… des anomalies métaboliques : surcharges lipidiques, déficit d'activité enzymatique ciblées etc… des anomalies structurales : déficit ou absence de protéines cytoplasmiques ou membranaires.

En complément d'investigation, une méthode nommée « Western-Blot », issue du génie moléculaire, permet de quantifier les protéines du tissu prélevé. Cette méthode renseigne sur la composition quantitative en protéines musculaires et est un précieux examen diagnostic de myopathies à composantes structurales.

2.3. Dosages enzymo-biologiques récurrents

2.3.1. Ponction lombaire

Elle consiste à prélever du Liquide Cérébro-Spinal (LCS, autrement appelé Liquide Céphalo-rachidien, LCR) entre 2 vertèbres du bas du dos. Nécessairement stérile, la composition en protéines reflète l'état neuro-immunologique du sujet. On y recherche le lactate, les neurotransmetteurs, ou des peptides accumulés de façon inhabituels, pouvant être des marqueurs de certaines atteintes neurodégénératives, comme dans les Tau-pathies ou la SLA.

Dans les MNDREM, la ponction lombaire n'est pas un examen systématique de diagnostic mais peut-être sollicité dans une démarche étiologique.

2.3.2. Créatinine Phosphokynase (CPK), lactate et pyruvate

Pour vérifier le fonctionnement métabolique du système neuromusculaire, on peut évaluer les taux d'activité de certaines enzymes.

La créatine phosphokinase (CPK), est une enzyme ubiquitaire. Dans la mitochondrie, elle catalyse des réactions permettant la mise à disposition de molécules énergétique comme l'ATP. La valeur physiologique habituelle des CPK plasmatiques est de 15-130UI/L. Mais en cas d'effort musculaire intense ou de dégradation du tissu musculaire, cette valeur peut augmenter de plus de 40 fois.

Lactate et pyruvate sont des substrats de la chaîne respiratoire mitochondriale. Des taux inhabituels de ces substrats caractérisent une atteinte probablement métabolique. En particulier, un rapport lactate/pyruvate diminué peut signer une atteinte de la chaîne respiratoire de la mitochondrie.

De même, d'autres activités enzymatiques peuvent être mesurées pour renseigner sur une composante métabolique dans les MNDREM.

2.3.3. Chaîne respiratoire mitochondriale

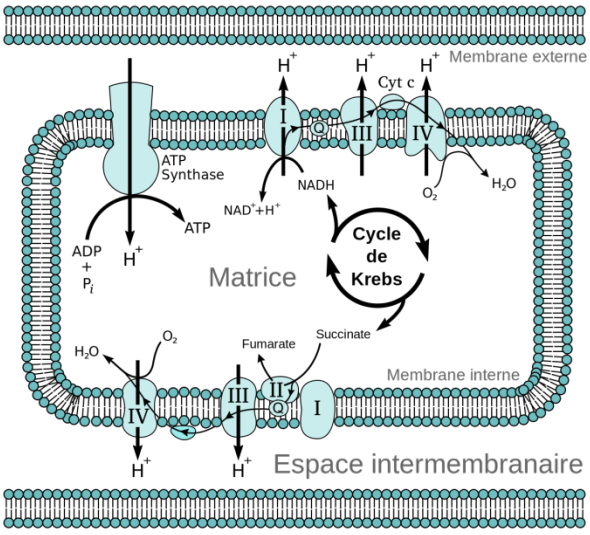


FIGURE 8, schéma représentant l'activité mitochondriale [141]

L'étude de l'activité des complexes de la chaîne respiratoire indique son niveau de perturbation. Le principe repose sur l'analyse des réactions enzyme-substrat. Les déficits en complexe des chaînes respiratoires sont impliqués dans des pathologies humaines à caractère moteur et/ou cognitif.

La FIGURE 8 ci-contre schématise le fonctionnement de la chaîne respiratoire et du cycle de Krebs.

2.4. Tests génétiques

2.4.1. Aspects éthiques de l'examen des caractéristiques génétiques

L'analyse des gènes d'un individu à des fins médicales, est possible en respectant un cadre légal de prescription fixé par l'arrêté du 27 mai 2013 définissant les règles de bonnes pratiques applicables à l'examen des caractéristiques génétiques d'une personne à des fins médicales.

En pratique, l'analyse du génome est accessible après consentement éclairé du patient obtenu dans le cadre d'une consultation singulière avec le spécialiste, clinicien ou conseiller génétique.

2.4.2. Recherche de SNVs et tandem repeats : Séquençages

Selon les orientations cliniques, l'analyse d'un ou plusieurs gènes peut être demandée à un laboratoire agréé par l'Agence de Biomédecine. D'un point de vue technologique, le séquençage (Sanger ou Haut Débit) est l'examen de choix pour la recherche de petites variations nucléotidiques.

- ✓ **Sanger**

Le séquençage dit de « Sanger », est la 1e méthode de séquençage mise au point au début des années 80 par Sanger et Gilbert. Elle consiste en la possibilité de cibler une région génomique à amplifier puis séquencer. On obtient une séquence complémentaire du brin cible que les moyens technologiques et informatiques permettent de lire. L'alignement de la séquence obtenue par séquençage Sanger, contre une séquence du génome humain de référence, permet, par comparaison, de mettre en évidence des anomalies génomiques de la région étudiée.

En pratique, les moyens technologiques ont permis une parallélisation des échantillons à tester. Mais ce séquençage est onéreux et ne permet pas le multiplexage des régions cibles. Il est supplanté par les méthodes dites de

séquençage massif en parallèle ou Séquençage Haut Débit (SHD), Next Generation Sequencing (NGS).

Le séquençage Sanger reste une méthode de référence et certain laboratoire l'utilise comme méthode de contrôle des résultats obtenus par d'autres méthodes de séquençage.

✓ Séquençage Haut Débit ou Next Generation Sequencing

Né du challenge du séquençage du génome humain au début des années 2000, cette méthode de séquençage massif en parallèle permet la détection de variants génomiques simultanés. Plusieurs régions peuvent être ciblées en même temps, et plusieurs échantillons sont testés en une seule manipulation. Cette méthode, moins chère et aussi plus sensible, requiert cependant un dispositif informatique robuste et adapté. Il s'agit d'une méthode hautement performante en termes de sensibilité et de rentabilité. Les prouesses technologiques ont amené la possibilité de tester plusieurs ADN sur plusieurs régions, dans la même expérience.

Sur le plan pratique, l'alignement contre les séquences du génome humain de référence est toujours le seul moyen de détecter des variants du génome testé. En particulier SNVs mais aussi parfois des CNVs, à condition d'adapter la méthode expérimentale et informatique.

Les analyses NGS proposées peuvent être :

- L'étude de panel de gènes impliqués dans une pathologie donnée,
- L'etude de l'exome (Whole Exome Sequencing, WES),
- Ou du génome (Whole Genome Sequencing, WGS).

En France, plusieurs multinationales se partagent le leadership de ces technologies, notamment Illumina, Life science ou Roche.

Le NGS est moins onéreux que le séquençage 1e génération, le Sanger. Mais de façon générale, les fragments séquencés sont moins longs : 250 à 300pb maximum. Le NGS autorise la détection de variant nucléotidique type Single Nucleotid Variants (SNVs) ou insertion-délétion (indel). La détection de variant d'expansion de motif est possible mais n'est pas encore optimal. Les progrès en matière bioinformatique sont attendus à ce niveau.

Les données informatiques générées par le NGS occupent de gros volumes. Mais ils sont décroissants selon l'analyse pratiquée : WGS, WES ou panel. Le développement du NGS permet de réaliser le séquençage d'un génome d'un individu à un coût de séquençage aux alentours des 1000$.

Aujourd'hui, les technologies de 3e génération sont déjà entrain de se mettre en place. Il s'agit de méthodes ne nécessitant pas l'amplification avant séquençage, et surtout, autorisant la lecture de longs fragments. Cette 3e génération offre de grand espoir dans l'analyse des variants par expansion de motifs répétés.

Actuellement, c'est la PCR long range qui a la particularité de pouvoir amplifier de longs fragments et est utilisée prioritairement en routine dans la détection des motifs répétés.

2.4.3. Recherche de CNVs : Analyse Comparative sur Puce à ADN, ACPA ou CGH array

Pour la recherche de remaniement génomique, en particulier délétion et Variation de Nombre de Copies (Copy Number Variations, CNVs), d'autres méthodes peuvent être employées, comme l'Analyse Comparative sur Puce à ADN (ACPA), ou CGH en anglais (Comparative Genomic Hybridization). Des méthodes moins récentes, comme la PCR en temps réelle ou le Southern-Blot sont utiles pour des confirmations ou quantification de fragment d'ADN.

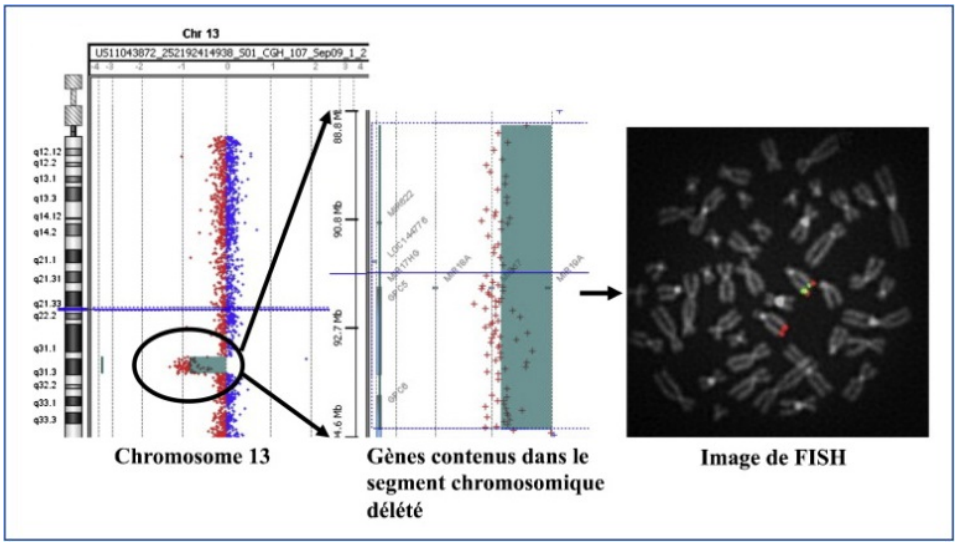


FIGURE 9, *schéma du principe de la CGH* [140]

3. MND Rares à Expression Motrice, MNDREM d'intérêt

Les travaux réalisés ont porté sur les MNDREM vues au CERCA, *a fortiori*, rencontrées en Martinique et dans les Antilles françaises.

3.1. Les ataxies héréditaires

3.1.1. Symptômes

Les Ataxies Spinocérébelleuses (par la suite dénommées CA pour Cerebellar Ataxia) sont un groupe de MND caractérisé par une perte de l'équilibre et un trouble de la coordination des mouvements. Elles relèvent du dysfonctionnement du cervelet et/ou des pédoncules cérébelleux et/ou des voies spino-cérebelleuses. On distingue :

- Le syndrome cérébelleux statique, le plus souvent en lien avec une atteinte du vermis et caractérisé par une instabilité à la marche, un élargissement du polygone de sustentation et une hypotonie avec réflexes ostéotendineux pendulaires
- Le syndrome cérébelleux cinétique caractérisé par une atteinte segmentaire avec dysmétrie, tremblement d'action et poursuite saccadée.

3.1.2. Diagnostic, suivi de l'évolution, diagnostics différentiels

Le diagnostic repose sur l'examen clinique et neurologique, et la caractérisation stéréotypique de l'atteinte motrice ataxiante. La composante neurodégénérative est confirmée par l'IRM cérébrale qui objective l'atrophie localisée au cervelet et/ou la moelle épinière et/ou les nerfs périphériques. L'ataxie spinocérébelleuse peut s'inscrire dans un cadre syndromique ou isolé. Le diagnostic différentiel est très large mais doit permettre d'exclure nécessairement les causes acquises.

L'échelle SARA (Scale for the Assessment and Rating of Ataxia) permet de donner un score de sévérité aux ataxies entre 0 et 42 [29]. Plus le score est élevé, plus la pathologie est invalidante.

L'échelle SDFS (Spinocerebellar Degeneration Functional Score) permet de juger du degré d'autonomie d'un patient avec un score allant de 0 à 7. [30]

3.1.3. Classification

Les CA sont très hétérogènes sur le plan génétique comme clinique : de nombreux gènes et variants pathogènes impliqués, phénotype très variable, mécanisme physiopathologique également. Il existe une bonne cinquantaine de CA. On les classe selon leur mode de transmission. Tous les modes de transmission existent dans les CA :

- Autosomique dominant (ADCA). On dénombre 38 types (SCA1-SCA38). SCA3 est le plus fréquent. Un peu plus d'1/3 est dû à une variation délétère par expansion de motifs répétés. 40-48% des ADCA n'ont pas d'étiologie moléculaire [31].
- Récessif (ARCA). Fogel and Perlman [32] proposent une classification en regard du phénotype de la maladie de Friedreich (AF), la plus fréquente des ARCA :
    - Friedreich's ataxia-like
    - Friedreich's ataxia-like with cerebellar atrophy
    - Early onset ataxia with cerebellar atrophy
- Lié à l'X. La plus courante liée à la prémutation du gène FMR1. L'histoire familiale retrouve des cas de retards mentaux. Les variants pathogènes du gène FMR1 sont associés au syndrome d'X-fragile
- Mitochondriale. Hors mis la maladie de Friedreich dont les variants pathogènes nucléaires sont responsables d'un dysfonctionnement mitochondrial, il existe des maladies mitochondriales associées à une atteinte du SNC se traduisant entre autres, par une ataxie [33].

D'autres classifications existent, basées sur l'âge de début ou le type de neuropathie.

3.1.4. Épidémiologie

Les CA isolées d'origine génétique sont rares : environ 1 cas sur 10 000 en France. Les CA sont hétérogènes et complexes. Tous les modes de transmission existent. L'âge de début est variable. Les âges de début et les délais de survie dépendent du type de CA et en général, son mode de transmission. Une bonne soixantaine de gènes sont maintenant impliqués dans les CA, associés à plus de 200 variants pathogènes.

Il n'y a pas de données épidémiologiques caribéennes ou Antillaises. Mais en 1995-1996, notre équipe publiait les premiers cas de SCA2, ADCA liée à une variation délétère par expansion de triplets dans le gène *ATXN2*. [34], [35]. Depuis, de grandes familles ont été identifiées.

3.1.5. Génétique et physiopathologie

La plupart des CA dominantes est associée à des variants pathogènes par expansion de motifs répétés. On observe alors des formes familiales avec anticipation. Les CA récessives se rapportent à des variants pathogènes de toute nature. La revue de la littérature permet de dresser les TABLEAUX 5 (CA dominants) et 6 (CA récessifs) ci-après, qui répertorient les phénotypes, fonctions et variants des gènes, ainsi que les MNDREM associées au gène causal impliqué dans les CA. Il est intéressant de noter que les mécanismes physiopathologiques sous-jacents sont variés. Ils peuvent concerner :

- Un dysfonctionnement mitochondrial, avec les variants pathogènes du gène *FRDA* (maladie de Friedreich), les variants pathogènes des gènes mitochondriaux (Kearn-Sayres syndrome, Myoclonic Epilepsy with Ragged Red Fibers syndrome etc..) ou encore les variants pathogènes du gène *AGFL3* dans SCA28.
- Un défaut de la réparation de l'ADN, avec les variants pathogènes dans le gène *ATM* (ataxie télengiectasie), *APTX* et *SETX* (Apraxies oculomotrices AOA1 et AOA2), etc…

- Un défaut de dégradation ou un phénomène d'accumulation, avec les variants pathogènes par expansions de chaînes Glutamines, dans les SCA 1 à 3, 6, 7 et 17 par exemple.
- Une anomalie des canaux ioniques, avec les variants pathogènes du gène *KCNA1* et *CACNA1A* dans les Ataxies épisodiques, ou *PKC* dans la SCA14
- Une anomalie dans la protéine Tau-tubulin kinase 2, codée par le gène *TTBK2* muté dans SCA11
- Une anomalie du Glutamate, gène *SPTBN2* dans SCA5
- Des déficits en vitamine E par variants pathogènes du gène *TTPA*.

Certains ARCA présentent des biomarqueurs offrant parfois des possibilités thérapeutiques :

- L'ataxie avec déficit en Vitamine E, débute entre 2-50 ans avec une moyenne à 17 ans. Elle se présente comme une AF avec un taux plasmatique de Vit E effondrée. La supplémentation préventive ou très précoce permet de limiter les symptômes.
- L'ataxie télangiectasie, débute avant 5 ans et s'accompagne de taux l'alpha-fœtoprotéine élevé avec un déficit immunitaire variable se traduisant par une diminution des IgA.
- L'AOA1, se déclare entre 1-20 ans avec une moyenne d'âge de 7 ans. Les taux d'albumine sont bas alors que le taux de cholestérol est élevé.
- L'AOA2, débute entre 7-25 ans avec une moyenne d'âge à 15 ans. Ici, le taux d'alpha-foetoprotéine est augmenté.

Entité clinique	Gène, protéine	locus	Rôles	Type de variants pathogènes	Autres MNDREM
ATAXIES DOMINANTES					
SCA1	ATXN1, ataxin-1	6p23	régulateur transcription et voie notch	CAG expansion 39-83 (norme: 6-36)	
SCA2	ATXN2, ataxin-2	12q24.1	régulateur endocytose	CAG expansion 34-400 (norme:15-31)	parkinson late-onset
SCA3	ATXN3, ataxin-3	14q21	enzyme déubiquitinilante	CAG expansion 53-86 (norme <47)	
SCA4	SCA4	16q22.1	?	?	SCA 31?
SCA5	SPTBN2, secretory carrier associated membrane protein 4	11p13	squelette membranaire des neurones	SNVs	autosomique récessive ataxia 14
SCA6	CACNA1A, altered 1A subunit of the oltage dependent calcium channel	19p13	canal calcique voltage dépendant	CAG expansion 20-33 (norme <18)	Encéphalopathie épileptique infantile (42), ataxie épisodique de type 2, migraine hémiplégique familiale de type 1
SCA7	ATXN7, ataxin-7	3p14.1-p12	stabilise les microtubules, active transcription	CAG expansion 37 - >300 (norme: 4-35)	
SCA8	KLHL1AS, putative protein ATXN8OS	13q21	gène non codant	CTG expansion 100-250 (norme 15-52)	parkinson late-onset
SCA10	ATXN10, ataxin-10	22q13	maintien de la glyscosylation, neuronigenèse	ATTCT expansion 280 - >4500 (norme: 10-22)	
SCA11	TTBK2, tau tubulin kinase 2	15q14-q21.3	Ser/Thr kinase regulateur de ciliogenèse	SNVs	
SCA12	PPP2R2B, Serin-thréonine protéine phosphatase 2A	5q31	activités proapoptotique; régulateur survie neuronale et équilibre fusion/fission mitochondriale	non coding CAG expansion 45-63 (norme: 7-31)	
SCA13	KCNC3, potassium voltage-gated channel subfamily C member 3	19q13.3-q14.4	canal potassique voltage-dépendant	SNVs	
SCA14	PRKC, Ser/Thr protéine kinase C	19q13.4	protéine proapoptotique	SNVs	
SCA15	ITPR1, inositol-1,4,5-triphosphate receptor type 1	3p26	Canal calcique intracellulaire		SCA29, Gillespie syndrome
SCA16	?	3p26.2pter	?	?	SCA15
SCA17	TBP, TATA-box binding protein	6q27	partie du système de transcription	CAG expansion 63 (norme 25-42)	HDL4, parkinson tardif
SCA18	?	7q22-q32	?	?	Neuropathie sensitive et motrice avec ataxie
SCA19	KCND3, potassium voltage-gated channel subfamily D member 3	1p13	sous-unité pore formant le canal potassique voltage dépendant	SNVs	SCA22, Brugada syndrome 9
SCA20	?	11p13-q11	?	?	Chromosome 11q12 duplication syndrome
SCA21	TMEM240, transmembrane protein 240	1p36.33	protéine transmembranaire	SNVs	
SCA22	?	1p21-q23	?	?	SCA19
SCA23	PDYN, proenkephalin-B	20p13	perception de la douleur et réponse au stress	SNVs	

ATAXIES DOMINANTES SUITE…

Entité clinique	Gène, protéine	locus	Rôles	Type de variants pathogènes	Autres MNDREM
SCA25	?	2p21-q15	?	?	
SCA26	EEF2, eukaryotic elongation factor 2 kinase	19p13.3	protéine Thr kinase régulant élongation	SNVs	
SCA27	FGF14, Fibroblast growth factor 14 related	13q33.1	développement système nerveux	SNVs	
SCA28	AFG3L2, AFG3-like protein 2	18p11.21	protéase ATP-dépendante	SNVs	atrophie optique 12, ataxie spastique 5 autosmique récessive
SCA29	?	?	?	SCA15	
SCA30	?	4q34.3-q35.1	?	?	
SCA31	BEAN, brean expressed protein associating with Nedd4 homolog	16q22.1	proteine cérébrale	(TGGAA)n repeat insertion	
SCA32	?	7q32-q33	?	?	
SCA35	TGM6, protein-glutamine gamma-glutamyltransferase	20p13	catalyse polymérisation peptidique	SNVs	
SCA36	NOP56, nucleolar protein 56	20p13	biogenèse ribosome sous-unité 60S	SNVs	
Spinocerebellar ataxia linked to 16q22	PLEKHG4, puratrophin	16q22	signalisation intracellulaire et cytosquelette Golgi	SNVs	
Dentatorubropallidoluysian atrophy (DRPLA)	ATN1, atrophin-1	12p13.31	co-répresseur de transcription	triplets repeats expansion	hypotonie congénitale, épilepsie, retard développement et anomalies digitales

TABLEAU 5, *Génétique des ataxies spinocérébelleuses DOMINANTES*

Entité clinique	Gène, protéine	locus	Rôles	Varia..	Autres MNDREM
ATAXIES RECESSIVES					
Friedreich's ataxia	FRDA, frataxin	9q13	Mitochondrial iron metabolism	SNVs	
Ataxia with vitamin E deficiency	TTPA, alpha-tocophérol transfer protein	8q13.1-13.3	Vitamin E homeostatisis	SNVs	
Abetalipoproteinaemia	MTTP, Microsomal triglyceride transfer protein	4q22-24	Lipoprotein metabolism	SNVs	Metabolic syndrome X
Refsum's disease	PHYH, Phytanoyl-CoA hydroxylase / PEX7, peroxin 7	10pter-11.2 / 6q21-22.2	Fatty-acid oxidation / Peroxisomal protein importation	SNVs	
late-onset Tay-Sachs disease	HEXA, beta-hexosaminidase A	15q23-24	Glycosphingolipid metabolism	SNVs	
cerebrotendinous xanthomatosis	CYP27, sterol 27-hydroxylase	2q33-ter	Bile-acid synthesis	SNVs	
DNA polymerase gamma disorders (mitochondrial recessive ataxia syndrome)	POLG, DNA polymerase gamma	15q22-26	Mitochondrial DNA repair and replication	SNVs	ophtalmoplégie externe progressive autosomique dominante, syndrome dépélétion ADNmt type 1, 4B, scelrose progressive poliodystrophie, ataxie sensitive avec neuropathie, dysarthrie et opthalmopmarésie syndrome
Spinocerebellar ataxia with axonal neuropathy	TDP1, Tyrosyl-DNA phosphodiesterase 1	14q31-32	DNA repair	SNVs	
ataxia telengiectasia	ATM, Ataxia telengectasia-mutated	11q22-23	DNA damage response	SNVs	cancer du sein familial
ataxia telengiectasialike disorder	MRE11, meiotic recombination 11	11q21	DNA damage response	SNVs	
ataxia with oculomotor apraxia, type 1	APTX, Aprataxin	9p13	DNA repair, possibliy RNA processing	SNVs	
ataxia with oculomotor apraxia, type 2	SETX, Senataxin	9q34	Possibly DNA repair, DNA transcription, or RNA processing	SNVs	Amyotrophic Lateral Sclerosis 4
autosomal recessive ataxia of Charlevoix-Saguenay	SACS, Sacsin	13q11	Possibly protein folding	SNVs	
Infantile-onset spinocerebellar ataxia	C10Orf2, Twinckle, twinky	10q24	DNA replication	SNVs	ophtalmoplégie externe progressive 3 autosomique dominant, mitochondrial DNA depletion syndrome 7 (hepatocerebral type), perrault syndrome 5
Cayman ataxia	ATCAY, Caytaxin	19q13.3	Possibly neurotransmitter metabolism	SNVs	
Marinesco-Sjögren's syndrome	SIL1, BiP-associated protein	5q31	Possibly protein folding	SNVs	

TABLEAU 6, *Génétique des ataxies spinocérébelleuses RECESSIVES*

3.2. Les maladies de Huntington et Huntington-like

3.2.1. Symptômes

La maladie de Huntington (HD) a été caractérisée en 1872 par George Huntington. Il s'agit d'une MNDREM progressive incluant des mouvements choréiques et une dystonie, des troubles de la coordination et un déclin cognitif associé à des troubles du comportement qui évoluent vers la démence. L'âge de début est variable car les formes familiales sont associées à un phénomène d'anticipation. Les 1es signes peuvent être la dépression. [36]

HD est typiquement associée à la perte sélective de cellules et l'atrophie du noyau caudé et du putamen, constituant le striatum. D'autres zones cérébrales peuvent être touchées comme : la substantia nigra, les couches corticales 3, 5 et 6, la région CA1 de l'hippocampe, le gyrus angulaire dans le lobe pariétal, les cellules de Purkinje du cervelet, les noyaux tubéreux latéraux de l'hypothalamus et du complexe centromédial-parafasciculaire du thalamus.

3.2.2. Diagnostic, suivi de l'évolution, diagnostic différentiel

L'IRM et le scanner par Tomodensitométrie (TDM) de routine dans les stades modérés à sévère montrent une perte de volume striatal et une augmentation de la taille des cornes frontales des ventricules latéraux.

L'une des caractéristiques pathognomoniques de la maladie de Huntington est l'apparition d'inclusions nucléaires et cytoplasmiques de huntingtine et de la polyglutamine mutantes. Ces inclusions prédisent le dysfonctionnement cellulaire. L'évolution de la maladie semble corrélée à des stades intermédiaires d'agrégats de polyglutamine.

L'échelle d'évaluation clinique de la maladie de Huntington, « Unified Huntington's Disease Rating Scale » (UHDRS) a été développée pour évaluer quatre domaines de performance et de capacité clinique : la fonction motrice, la fonction cognitive, les anomalies comportementales et la capacité fonctionnelle.

Le "Diagnostic Confidence Level " (DCL) est une échelle d'évaluation de l'atteinte motrice de l'UHDRS.

Les diagnostics différentiels sont les autres causes de chorée d'origine endogène ou iatrogène. La Chorée-acanthocytose est le diagnostic différentiel majeur. Il s'agit d'une myopathie caractérisée par une augmentation de Créatinine Phospho Kinase (CPK) et la présence d'acanthocytes dans le sang. Des crises d'épilepsie sont fréquentes et l'âge de début se situe vers 30 ans. Les variants pathogènes du gène *VPS13A* sont impliqués dans ce phénotype.

3.2.3. Classification

HD est d'origine génétique, et la majorité des cas est associée à une variation délétère du gène *HTT* codant pour la protéine Huntingtine. Chez environ 1% des personnes présentant les caractéristiques de la maladie de Huntington, aucune variation pathogène du gène HD n'a été identifiée.

Quatre syndromes Huntington Like (HDL) sont décrits. Ils sont désignés comme la maladie de Huntington-Like 1 (HDL1) à la maladie de Huntington-Like 4 (HDL4). Les syndromes HDL correspondent aux caractéristiques d'une maladie de Huntington qui n'est pas causée par une variation pathogène du gène *HTT*.

HDL1 est causée par des variants pathogènes dans le gène *PRNP*, tandis que les variants pathogènes du gène *TBP* sont responsables de la HDL4, également connue sous le nom d'ataxie spinocérébelleuse de type 17. La cause génétique de HDL3 est inconnue à ce jour.

Au début des années 2000, HDL2 est lié au locus 16q.24.2. Holmes SE et al. confirment l'implication des variants pathogènes par expansion de triplets CTG dans le gène *JPH3* codant pour la Junctophilin dans l'HDL2 [37]. HDL2 est une forme de HD fréquemment retrouvée dans les populations d'ascendance africaine. [38], [39]

3.2.4. Epidémiologie

Typiquement, l'âge de début de HD se situe entre 45-50ans. Mais sa prévalence varie selon les ethnies notamment à cause des phénocopies HDL: 5.1, 2.1 and 0.25 pour 100,000 individus respectivement chez les caucasiens, métisses and descendants noirs-africains.

Les syndromes HDL sont beaucoup moins fréquents que HD. Des quatre syndromes HDL décrits, HDL4 semble être le plus courant. HDL2 est le deuxième plus commun et se produit presque exclusivement chez les personnes d'origine africaine (en particulier les Sud-Africains noirs) [40]. HDL1 est très rare. HDL3 a été trouvé dans deux familles, toutes deux originaires d'Arabie saoudite. [41]

3.2.5. Génétique et physiopathologie

HD et les HDL se transmettent sur le mode autosomique dominant. Les gènes impliqués sont : *HTT, PRNP, JPH3 et TBP*. Mais les variants pathogènes ponctuels de ces gènes génèrent des pathologies à caractère moteur :

- Les variants délétères p.Phe198Ser, p.Glu200Lys, p.Val210Ile du gène *PRNP* sont également associés au phénotype HDL1. Les autres variants pathogènes ponctuels sont associées à des formes familiales de maladie à prions, y compris la maladie de Creutzfeldt-Jakob (MCJ), le syndrome de Gerstmann-Sträussler-Scheinker (GSS) et l'insomnie familiale fatale (FFI).
- Les variants bi alléliques pathogènes faux-sens de *HTT* seraient à l'origine du syndrome de neurodéveloppemental autosomique récessif du syndrome de Lopes-Maciel-Rodan (LOMARS). Les personnes atteintes de LOMARS présentent dans l'enfance un phénotype semblable au syndrome de Rett avec une déficience intellectuelle.
- Les variants dans *TBP* sont à l'origine de formes génétiques de maladie de Parkinson, dite à début tardif. Nous en parlons dans les chapitres suivants.

En revanche, les variants pathogènes des gènes *HTT, PRNP, JPH3 et TBP* impliqués dans les phénotype HD sont toutes des variants par expansions de motifs répétés. Le TABLEAU 6 ci-après, présente les variants pathogènes par expansion de motifs,

trouvés dans les gènes en cause dans les maladies de Huntington et Huntington-like et les MNDREM associées.

Les variants pathogènes par expansion de motifs répétés sont des « variations dynamiques » ; ce qui signifie qu'il existe une possibilité de glissement lors des divisions méiotiques à l'origine du phénomène d'anticipation observée dans les familles atteintes.

Les individus porteurs de variants d'expansion compris dans le « zone grise » sont le plus souvent asymptomatiques. On parle également de défaut de pénétrance de ces variants. Cependant, ces sujets sont à risque de transmettre des expansions plus longues et donc pathogènes.

Sur le plan fonctionnel, plusieurs hypothèses sont avancées pour comprendre le mécanisme par lequel les variants pathogènes par expansion induisent la présence d'inclusions cellulaires et la perte sélective neuronale. A priori, les variants pathogènes par expansion de triplets CAG causerait un gain de fonction protéique. Autrement dit, la protéine mutée acquerrait de nouvelles compétences biologiques via une nouvelle conformation tridimensionnelle. Notamment, elle serait alors capable à la fois de polymériser, mais aussi de coopérer avec des protéines type facteurs de transcription, protéines signal, ARNs et autres compartiments cellulaires, avec lesquels la protéine sauvage n'interagit pas naturellement. Créant ainsi progressivement un ensemble de perturbations cellulaires à l'origine *in fine* de l'apoptose sélective de cellules nerveuses du striatum, organe fonctionnel de la motricité notamment. [13]

La compréhension fine et détaillée des processus cellulaires en cause dans HD est nécessaire pour élargir les pistes thérapeutiques de cette maladie ; mais également peut-être aussi les maladies de même mécanisme moléculaire que sont les maladies par expansions de motifs répétés.

Entité clinique	Gène, protéine	locus	Fonctions	Variants	Autres MNDREM
HD	HTT, Huntingtin	4p16.3	Rôle dans le transport médié par les microtubules ou dans la fonction vésiculaire/impliqué dans les processus apoptotiques	Expansion CAG, Exon 1 >36 (norme : 18-26 / zone grise : 27-35)	
HDL1	PRNP, Prion protein	20p13	Sa fonction physiologique principale n'est pas claire. Rôle dans développement neuronal et plasticité synaptique. Entretien de la gaine de myéline. Absorption et homéostasie du fer.	Insertion en phase de 48pb ou entre les codons 51 et 91, >4 motifs GlnGlyGlyGlyGlyTrpGlyGln Soit : p.Gln59_Pro60insGlnGlyGlyGlyGlyTrpGlyGlnGlnGlyGlyGlyGlyTrpGlyGlnGlnGlyGlyGlyGlyTrpGlyGlnGlnGlyGlyGlyGlyTrpGlyGln)	Maladies à Prion, maladie de Wilson
HDL2	JPH3, junctophilin 3	16q24.2	Contribue à la formation de complexes membranaires jonctionnels (JMC) qui relient la membrane plasmique au réticulum endoplasmique ou sarcoplasmique dans les cellules excitables. Support de la communication entre surface cellulaire et canaux calciques. Spécifique au cerveau. Rôle actif dans neurones impliqués dans la coordination motrice et la mémoire.	Expansion CAG/CTG >41 dans Intron 1 (norme : 6-28 / zone grise : 29-39)	
HDL3					
HDL4/ SCA17	TBP, TATA-box binding protein	6q27	Assure médiation transcription de la plupart des protéines ribosomiques. Semble médier la transcription de NF1. Ne lie pas la boîte TATA.	Expansion CAG/CAA >47 dans Exon 3 (norme : 25-42 / zone grise : 42-45)	Parkinson à début tardif

TABLEAU 6, *expansion de motifs répétés dans les maladies de Huntington et huntington-like (HDs)*

3.3. La Sclérose Latérale Amyotrophique

3.3.1. Symptômes

La sclérose latérale amyotrophique (SLA) a été décrite pour la première fois en 1873 par le Jean-Martin Charcot. La dénomination de la pathologie décrit ses principales caractéristiques : fonte musculaire (amyotrophique), et perte des motoneurones supérieurs et de leurs voies axonales corticospinales (sclérose latérale). Les second motoneurones, inférieurs, ceux de la corne antérieure de la moelle épinière, sont également touchés [42]. La SLA provoque une paralysie progressive de l'ensemble de la musculature striée des membres, du tronc (y compris les muscles respiratoires) et de l'extrémité céphalique.

Trois formes cliniques de SLA sont décrites :

- Spinale (2/3), l'apparition se manifeste par une faiblesse distale ou proximale et une perte musculaire focale dans les membres inférieurs et supérieurs. Progressivement, la spasticité se développe dans les membres atrophiés et affaiblis, altérant la dextérité manuelle et la démarche.
- Bulbaire, caractérisée par une dysarthrie initiale et une dysphagie aux solides ou aux liquides. Les symptômes dans les membres peuvent apparaître presque simultanément avec les symptômes bulbaires ou chez la grande majorité des patients, un à deux ans plus tard.
- Pseudopolynevrétique, les patients atteints de ce type, développent initialement un déficit distal et une amyotrophie des membres inférieurs imitant la polyneuropathie motrice. La paralysie est progressive et entraîne généralement une insuffisance respiratoire et la mort en deux à cinq ans.

Depuis cette première description, la SLA clinique s'est enrichie de plusieurs caractéristiques: âge variable de l'instabilité, membre supérieur ou inférieur en premier affecté, durée variable de la maladie, association avec des caractéristiques bulbaires, des troubles cognitifs ou des mouvements anormaux comme l'ataxie [43].

3.3.2. Diagnostic, suivi de l'évolution, diagnostic différentiel

En 1994, les critères El Escorial ont été établis pour le diagnostic de la SLA. Puis en 2008, Awaji et al. proposent des recommandations pour l'utilisation d'études électromyographiques dans le diagnostic de la sclérose latérale amyotrophique. L'objectif était de permettre un diagnostic plus précoce, d'éviter les retards de diagnostic et de favoriser une entrée plus précoce dans les essais cliniques [44]. Ensemble, ces deux méthodes uniformisent le diagnostic universel de la SLA. Le diagnostic clinique de la SLA est en fait basé sur la révision des critères EL Escorial - 2015 de Ludolph et al.

Le diagnostic de sclérose latérale amyotrophique (SLA) est défini par au moins l'un des critères suivants : déficits progressifs des motoneurones supérieur et inférieur (UMN / LMN), dans au moins un membre ou une région du corps humain ; ou une évidence clinique de déficits de motoneurones inférieurs dans une région du corps humain et / ou d'ENMG neurologique dans deux régions du corps. [42].

Pour standardiser les étapes du diagnostic et clarifier la complexité des caractéristiques, trois niveaux de diagnostic sont désormais reconnus sur la base des preuves cliniques uniquement: SLA possible, probable et définitive [45], [46].

SLA définitive : atteinte des motoneurones supérieurs (UMN pour Upper Motor Neuron) ainsi que des signes d'atteintes des motoneurones inférieurs (LMN pour Lower Motor Neuron), dans 3 régions du corps.

SLA probable : atteinte des motoneurones supérieurs ainsi que des signes d'atteintes des motoneurones inférieurs dans au moins 2 régions avec certains signes d'atteinte des motoneurones supérieurs nécessairement rostraux (au-dessus) des signes des motoneurones inférieurs.

SLA possible : dysfonction UMN et LMN trouvée conjointement dans une seule région ou signes UMN trouvés seuls dans 2 régions ou plus ; ou les signes LMN se trouvent rostraux aux signes UMN. Un autre diagnostic doit avoir été exclu pour accepter un diagnostic de SLA cliniquement possible.

Le diagnostic repose sur la connaissance des antécédents cliniques, de l'électromyogramme et de l'élimination, par des examens appropriés, des diagnostics différentiels de : paralysie bulbaire progressive, neuropathie motrice multifocale, syndrome du bras / de la jambe du fléau, la maladie de Kennedy, atrophie musculaire progressive ou encore la myélopathie spondylotique cervicale, par exemple. La présence de lésions des motoneurones supérieures et inférieures inexpliquées par une autre cause suggère une SLA.

3.3.3. Classification

La SLA peut survenir dans un contexte familial (Familial Amyotrophic Lateral Sclerosis, FALS) ou sporadique (Sporadic Amyotrophic Lateral Sclerosis, SALS). La FALS se distingue de la SLA héréditaire (Hereditary Amyotrophic Lateral Sclerosis, HALS) et d'origine génétique (Genetically determined Amyotrophic Lateral Sclerosis GALS). Le terme de FALS doit être utilisé si des antécédents familiaux de SLA ou de démence frontotemporale (FTD) sur trois générations sont documentés. On parle de HALS quand au moins un parent au premier ou au deuxième degré souffre de SLA et / ou de FTD. Le terme SLA primaire génétique (GALS) est réservé aux situations dans lesquelles un variant génique pathogène ségrége dans une famille. [42]

3.3.4. Épidémiologie

Les études épidémiologiques de la SLA montrent une incidence mondiale de 2 à 3 par an pour 100 000 habitants de plus de 15 ans et un risque global de développer une SLA de 1/350 pour les hommes et de 1/400 pour les femmes. Le risque augmente avec l'âge, avec un pic autour de 75 ans [47].

La fréquence de la SLA semble être constante dans les populations d'origine européenne dans le monde. Pour les autres populations, pour lesquelles des données suffisantes sont disponibles, le risque semble être plus faible que dans les populations dérivées européennes [48].

En Europe, les hommes sont légèrement plus fréquemment touchés que les femmes (taux d'incidence de 1,4). La durée moyenne d'évolution est généralement de 2 à 4 ans. L'âge moyen d'apparition dépend des formes de SLA, la SLA typique débute à l'âge adulte (61,8 ± 3,8 ans ; extrêmes 54–67 ans) mais ~ 1% des cas de SLA apparaissent avant 25 ans (juvéniles), ~ 10% des cas de SLA surviennent avant 45 ans (jeunes). L'âge moyen au diagnostic est de 64,4 ± 2,9 ans (extrêmes 58–68 ans) [49].

3.3.5. Génétique et physiopathologie

La SLA est le résultat de multiples mécanismes physiopathologiques et de dysfonctionnements cellulaires, notamment: mauvais repliement et agrégation des protéines, altération du traitement de l'ARN (principalement perturbation de l'épissage et de la signalisation de l'ARNm), défauts de transport axonal, métabolisme anormal et accumulation d'espèces réactives de l'oxygène, dysfonctionnements mitochondriaux, mécanismes neuro-inflammatoires microgliaux, excitotoxicité directe du glutamate par les astrocytes, effets modulateurs anormaux d'autres cellules gliales, perturbations de l'autophagie, anomalies du système ubiquitine-protéasome et anomalies des canaux ioniques primaires et secondaires [43].

Les données de la littérature renforcent l'idée d'un modèle dans lequel une charge génétique préexiste et est influencée par les expositions temporelles et environnementales jusqu'à un point de basculement où la neurodégénérescence commence. Il s'agit du concept de modèle gène-temps-environnement de la SLA [47].

Les données d'étude de jumeaux suggèrent que l'héritabilité contribue à 60% au risque de développer la SLA et l'environnement à 40%, mais une analyse de trois études d'association à l'échelle du génome a révélé une héritabilité inférieure de 21%. [47]. La communauté scientifique avaient pris l'habitude d'indiquer que 10% des individus atteints ont des antécédents familiaux de SLA et d'autres conditions neurodégénératives, mais une méta-analyse récente des données de la littérature ne trouve pas plus de 5% de formes familiales (FALS) [48], [50].

Les facteurs de risque environnementaux sont difficiles à étudier car l'espace de recherche est infini et aucun facteur de risque environnemental définitif et reproductible pour la SLA n'a encore été identifié. Les facteurs environnementaux fréquemment étudiés sont : l'exercice physique, le football, le tabac, les métaux lourds, les pesticides et les produits chimiques, le métier, les services de l'armée et le déploiement, les chocs électriques, les clusters géographiques ou encore les cyanotoxines. [47].

Bien que la grande majorité des cas soient sporadiques, un gain substantiel a été observé dans la connaissance des formes génétiques de la maladie, en particulier des formes familiales. À ce jour, environ 32 gènes sont impliqués dans les formes héréditaires de SLA, environ 15 loci sont associés à des formes sporadiques et à des gènes de sensibilité. [43]. Souza et al. dressent un schéma de la répartition des causes génétiques les plus fréquentes de la SLA en fonction de l'âge d'apparition. Cinq gènes représentent à eux seuls plus de 50% des cas de FALS: *SOD1*, *TDP-43*, *FUS*, *VCP* et *C9ORF72*.

Le TABLEAU 7 résume les informations relatives à ces gènes et les autres qui sont actuellement explorés en routine diagnostic.

Entité clinique	Gène, protéine	Localisation	Rôles	Type de variants	Autres MNDREM
ALS1	SOD1, Superoxyde Dismutase 1	21q22.11	Détruit les radicaux produits	SNVs	Paraplégie spastique et hypotonie axiale progressive
ALS10	TDP-43, TAR DNA binding Protein	1p36.22	Protéine de liaison à l'ARN. Biogenèse et traitement de l'ARN impliqués dans la survie neuronale. Maintien de l'homéostasie mitochondriale. Formation et régénération muscles squelettiques.	SNVs	Démence frontotemporale
ALS6	FUS, FUS RNA binding protein	16p11.2	Protéine de liaison à l'ADN / ARN. Régulation transcription, épissage/transport ARN, réparation de l'ADN et la réponse aux dommages)	SNVs	Tremblement essentiel de type 4
ALS14	VCP, Vasolin Containing Protein	9p13.3	Nécessaire à la fragmentation des empilements de Golgi pendant la mitose et à leur remontage après mitose. Régule l'activité de l'ubiquitine-protéine ligase E3.	SNVs	
DFT/ALS	C9ORF72, C9orf72-SMCR8 complex subunit	9p21.2	Régule l'autophagie, la maturation des phagosomes en lysosomes (par similitude), de la dynamique de l'actine dans les motoneurones, de l''extension des axones et de la taille des cônes de croissance des axones dans les motoneurones spinaux.	Hexanucleotide GGGGCC expansion >250-2000 (norme : 2-19)	Démence frontotemporale
ALS2	ALS2, Alsin	2q33.1	La protéine fonctionne comme un facteur d'échange de nucléotides guanine pour la petite GTPase RAB5. La protéine se localise avec RAB5 sur les compartiments endosomaux précoces et fonctionne comme un modulateur de la dynamique endosomale	SNVs	paralysie spastique ascendante infantile.
ALS9	ANG, Angiogenin	14q11.2	Membre de la superfamille RNase A. Puissant médiateur de la formation de nouveaux vaisseaux sanguins et est donc, en plus du nom RNase5. Induit une angiogenèse après liaison à l'actine à la surface des cellules endothéliales.	SNVs	
ALS17	CHMP2B, CHMP family, member 2b	3p11.2	Composant du complexe hétéromérique ESCRT-III (Endosomal Sorting Complex Required for Transport III) qui fonctionne dans le recyclage ou la dégradation des récepteurs de surface cellulaire.	SNVs	Démence frontotemporale
–	DCTN1, Dynactin subunit 1	2p13.1	La plus grande sous-unité de la dynactine. Complexe macromoléculaire constitué de 10 sous-unités dont la taille varie de 22 à 150 kD. Se lie à la fois aux microtubules et à la dynéine cytoplasmique. Impliquée dans un large éventail de fonctions cellulaires, y compris le transport RER-Golgi, le mouvement centripète des lysosomes et des endosomes, la formation du fuseau, le mouvement des chromosomes, le positionnement nucléaire et l'axonogenèse	SNVs	Neuronopathie motrice héréditaire distale de type 7B ; Perry Syndrom
ALS11	FIG4, Polyphosphoinositide phosphatase	6q21	Appartient à la famille des gènes de protéines contenant le domaine SAC, activité phosphoinositide phosphatase	SNVs	Charcot-Marie-Tooth de type 4J ; polymicrogyrie temporooccipital bilatréale ; syndrome de Yunis-Varon
ALS12	OPTN, Optineurin	10p13	Peut jouer un rôle dans le glaucome à tension normale et le glaucome primaire à angle ouvert de l'adulte. Médie l'apoptose, l'inflammation ou la vasoconstriction. Impliquer dans la morphogenèse cellulaire et le trafic membranaire, le trafic des vésicules et l'activation de la transcription grâce à ses interactions avec les protéines RAB8, huntingtine et facteur de transcription IIIA	SNVs	facteur de susceptibilité au glaucome a tension normale ; Glaucome primaire à angle ouvert

Gènes SLA suite…

Entité clinique	Gène, protéine	Localisation	Rôles	Type de variants	Autres MNDREM
ALS4	SETX Probable helicase senataxin	9q34.13	Nommée pour son homologie avec la protéine Sen1p des champignons qui a une activité ARN hélicase. Contient un domaine ADN / ARN hélicase à son extrémité C-terminale, ce qui suggère qu'elle pourrait être impliquée à la fois dans le traitement de l'ADN et de l'ARN	SNVs	Ataxie spinocerebelleuse autosomique récessive avec neuropathie axonale de type 2
SPG20	SPG20 Spartin	13q13.3	Contient un domaine MIT (Microtubule Interacting and Trafficking Molecule), et est impliqué dans la régulation du trafic endosomal et de la fonction des mitochondries. Localisée dans les mitochondries et partiellement avec les microtubules.	SNVs	Syndrome de Troyer
FTD/ALS3	SQSTM1 Sequestosome 1	5q35.3	Protéine multifonctionnelle qui se lie à l'ubiquitine et régule l'activation de la voie de signalisation du facteur nucléaire kappa-B (NF-kB). Fonctionne comme une protéine d'échafaudage / adaptateur de concert avec le facteur 6 associé au récepteur TNF pour médier l'activation de NF-kB en réponse à des signaux en amont.	SNVs	Myopathie distale avec vacuoles cerclées ; Démence frontotemporale, Neurodégénérescence avec ataxie, dystonie et paralysie du regard, début de l'enfance ; Maladie osseuse de Paget de type 3
FTD/ALS4	TBK1 TANK binding kinase	12q14.2	La protéine codée par ce gène est similaire aux kinases IKB et peut médier l'activation du NFKB en réponse à certains facteurs de croissance	SNVs	Sensibilité à l'encéphalopathie aiguë induite par une infection (spécifique aux herpes) ; ALS4 Démence frontotemporale
ALS15	UBQLN2 Ubiquilin-2	Xp11.21	Protéine de type ubiquiline qui contient un domaine de type ubiquitine N-terminal et un domaine associé à l'ubiquitine C-terminal. Ils s'associent physiquement à la fois aux protéasomes et aux ubiquitine ligases pour affecter la dégradation des protéines in vivo.	SNVs	ALS15 avec ou sans démence frontotemoporale
ALS8	VAPB Vesicle-associated membrane proteinassociated protein B/C	20q13.32	Protéine membranaire de type IV trouvée dans les membranes plasmiques et vésiculaires intracellulaires, et impliqué dans le trafic de vésicules	SNVs	Atrophie musculaire spinale proximale autosomique dominant de l'adulte,

TABLEAU 7, *Principaux gènes étudiés dans le SLA*
(SNVs, Single Nucleotid Variant (variant intéressant un seul nucléotide)

3.4. Les Syndromes parkinsoniens

3.4.1. Symptômes

Le syndrome parkinsonien (SPARK) fait référence à des symptômes de la maladie de Parkinson dite Idiopathique (MPI) causés par une étiologie identifiable ne remplissant pas tous les critères de la MPI. La MPI représentait une unité clinique apparemment indivisible jusqu'au progrès anatomopathologique et l'avènement de la génétique. La plupart des SPARK sont des MNDREM dont le diagnostic est rendu complexe par la proximité des symptômes et les subtilités de diagnostics différentiels.

La maladie de Parkinson Idiopathique (MPI) est la MND modèle des MND avec atteinte motrice. Elle est la 1e cause du handicap moteur du sujet âgé. Elle se définit par une triade de symptômes moteurs associant : Akinésie, Rigidité, Tremblement. La perte de 50 à 60 % des neurones du locus niger (substance noire), entraîne la survenue des signes moteurs de la triade parkinsonienne. D'expression unilatérale ou asymétrique. Le diagnostic est confirmé cliniquement par l'observation d'une nette réduction des signes moteurs à l'instauration du traitement dopaminergique.

Dans la maladie de Parkinson cette sensibilité aux traitements est prolongée au-delà de 5 ans. Quand le processus neurodégénératif dépasse la voie nigro-striatale, d'autres signes moteurs (par exemple axiaux), et non moteurs (par exemple les troubles cognitifs) peuvent survenir et résistent au traitement dopaminergique. On relève également la présence de corps de Lewy, ces inclusions intraneuronales contenant des agrégats anormaux d'alpha-synucléine.

L'étiologie de la MPI est actuellement inconnue, mais plusieurs facteurs génétiques et environnementaux sont identifiés, comme les variants pathogènes du gène de la parkine ou de l'α-synucléine. [51]

Un SPARK doit être évoqué quand : la triade apparue après 40 ans ou avant (sans être une maladie de Wilson) ne répond pas au traitement, progresse rapidement avec des chutes fréquentes, s'accompagne d'un syndrome cérébelleux et/ou pyramidal, de troubles oculomoteurs, de signes corticaux, âge de début avant 60 ans, présente une composante cognitive et/ou d'autres signes neurologiques comme une dysautonomie ou une atteinte pseudobulbaire…

3.4.2. Diagnostic, suivi de l'évolution, diagnostic différentiel

Le diagnostic différentiel des maladies de Parkinson peut être fait selon deux approches différentes et complémentaires : l'approche sémiologique « ciblée » tend à délimiter les « bornes » des différents syndromes parkinsoniens à l'aide de la sémiologie et des examens complémentaires. Le TABLEAU 8 rappelle les caractéristiques de ces diagnostics.

	Maladie de Parkinson	*Démence à corps de Lewy*	*Atrophies mutisystématisées*	*Paralysie Supranucléaire Progressive*
• akinésie et rigidité	asymétrique	+	sévère	axial
• tremblement	repos	rare	irrégulier /postural	–
• réponse à la lévodopa	+++	±	±	±
Instabilité posturale et chutes	±	+	++	+++
Ophtalmophégie supranucléaire	–	–	±	+++
Dysarthrie	±	–	précoce	S. pseudo-bulbaire
Dystonie	–	–	antécolis	rétrocolis
Hypotension orthostatique	±	±	++	–
Troubles génito-sphinctériens	–	–	+++	±
Syndrome cérébelleux	–	–	±	–
Hallucinations	–	+++	–	–
Syndrome frontal	–	–	±	+++
Démence	±	+++	–	++

TABLEAU 8, *sémiologie des SPARK* [51]

3.4.3. Classification

Les syndromes parkinsoniens (SPARK) sont des maladies de Parkinson peu ou pas dopa-sensibles regroupées sous les entités cliniques suivantes :

- SPARK du sujet jeune
- Atrophie multisystématisée
- Paralysie supranucléaire progressive
- Démence à corps de Lewy
- Dégénérescence cortico-basale
- Parkinsonisme des Caraïbes

L'apport de la biologie moléculaire permet de classer ces entités cliniques en sous-groupes selon les protéines accumulées : taupathies ou synucléinopathies.

- ✓ **Synucleinopathies**

Démence a corps de Lewy

La démence à corps de Lewy (DCL) est la seconde démence neurodégénérative la plus fréquente après la maladie d'Alzheimer. C'est une maladie complexe qui emprunte certains symptômes à la maladie d'Alzheimer et à la maladie de Parkinson, donc difficile à reconnaître ou à diagnostiquer. Non rare en Europe, elle ne sera pas dans les centres d'intérêt de ces travaux.

Syndrome parkinsonien du sujet jeune

Les Syndromes Parkinsoniens de l'Adulte Jeune (SPAJ) débutent avant l'âge de 50 ans. Ils se caractérisent par des postures dystoniques, souvent inaugurales, prédominant aux membres inférieurs, une bonne réponse au traitement antiparkinsonien, l'absence de trouble cognitif ou d'atteinte axiale, l'apparition plus fréquente et plus précoce de fluctuations et de dyskinésies, une fréquence accrue des troubles de l'humeur et des perturbations de la vie sociale, une progression plus lente de la maladie.

Les diagnostics différentiels sont aussi divers que les dystonies Dopa-sensibles, la maladie de Wilson, la maladie de Gaucher type 3, la neurodégénérescence associée à la panthoténate kinase, les formes juvéniles de maladie de Huntington ou d'atrophie dento-rubro-pallido-luysienne, les lésions des noyaux gris centraux.

Atrophie multisystématisée

L'atrophie multisystématisée (AMS) se caractérise par l'association de troubles dysautonomiques à un syndrome parkinsonien peu dopa sensible et/ou un syndrome cérébelleux. Les causes de décès les plus fréquentes sont les infections broncho-pulmonaires et les morts subites [52]

Selon la prédominance du syndrome parkinsonien ou du syndrome cérébelleux, il y a deux formes cliniques : AMS-P et AMS-C.

L'AMS est diagnostiquée à partir de critères cliniques définissant un niveau « probable » ou« possible » établis en 1998 et révisé en 2008 [53]. Le diagnostic de certitude repose sur une confirmation anatomopathologique post-mortem du cerveau retrouvant notamment des inclusions intra cytoplasmiques oligodendrogliales (ou glial cytoplasmic inclusion : GCI) majoritairement constituées

d'alpha-synucléine. Elles sont présentes dans l'ensemble du système nerveux central avec une forte densité dans les régions appartenant au système olivo-ponto-cérébelleux et la voie nigro-striée [54], [55]

Le diagnostic différentiel repose sur la possibilité de distinguer les alphasynucléinopathies partageant certaines caractéristiques cliniques avec la démence à corps de Lewy et la maladie de Parkinson Idiopathique.

✓ Taupathies

Paralysie supranucléaire progressive

Les signes neuropathologiques de la PSP sont une perte neuronale, une gliose avec des plaques astrocytaires en étoile, et des dégénérescences neurofibrillaires immunoréactives à la protéine tau, avec une distribution cérébrale spécifique. Les degrés d'accumulation et de distribution de la protéine tau phosphorylée sont corrélés avec les cinq variantes cliniques [56], [57]. La Paralysie Supranucléaire Progressive (PSP) a cinq variantes cliniques décrites selon des corrélations clinicopathologiques :

PSP classique : la plus fréquente, se manifeste par une marche trébuchante, des chutes précoces par instabilité posturale, un déclin cognitif et une lenteur des saccades oculaires verticales. Le tableau s'enrichit peu à peu de troubles de la parole, d'une ophtalmoplégie supra-nucléaire et de troubles de la déglutition.

PSP de type parkinsonien (PSP-P), se caractérise par un syndrome parkinsonien au premier plan et peu ou pas de troubles cognitifs. Les signes typiques de PSP classique apparaissent après plusieurs années d'évolution.

PSP-akinésie pure avec « freezing » de la marche (PSP-PAGF), se caractérise par un « freezing » progressif de la marche, de la parole et de l'écriture, précoce dans le cours de la maladie, avec peu ou pas de rigidité axiale et d'amimie faciale, plus tardifs, puis une parésie de la verticalité du regard vers le bas qui peut apparaître après une décennie.

PSP-syndrome corticobasal (PSP-CBS), se caractérise par une dyspraxie asymétrique progressive avec rigidité des membres, bradykinésie et instabilité posturale progressive.

PSP-aphasie progressive non fluente (PSP-PNFA), se caractérise par des troubles du langage (apraxie de la parole, agrammatisme et erreurs phonémiques), les symptômes moteurs étant plus tardifs.

Les diagnostics différentiels sont la maladie de Parkinson et les autres syndromes parkinsoniens atypiques, dont l'atrophie multisystématisée et la dégénérescence cortico-basale. Une maladie de Niemann-Pick de type C tardive et la maladie de Whipple peuvent également être confondues.

Dégénérescence cortico-basale

La Dégénérescence Cortico-Basale (DCB) représente 4 à 6 % des cas de parkinsonisme. La DBC touche principalement les régions cérébrales sous-corticales, puis corticales et les noyaux gris centraux. Elle présente quatre sous-types :

- Le type **classique**, présente typiquement un dysfonctionnement moteur asymétrique ainsi que des troubles cognitifs
- La variante **frontale**, caractérisée par l'apparition précoce d'une perte de dynamisme et d'une perturbation de l'attention, de troubles du comportement et de la personnalité ainsi que d'une désorientation spatiale
- La variante **linguistique**, caractérisée par l'apparition précoce de troubles de la parole, de la compréhension du langage et de l'élocution
- La variante **PSP**, caractérisée par une rigidité du tronc, des problèmes d'équilibre marqués présentant un risque de chute, une incontinence urinaire et des troubles de la mobilité oculaire

L'imagerie médicale cérébrale, met en évidence une atrophie unilatérale du lobe frontal et du lobe pariétal et peut s'avérer utile pour exclure d'autres maladies neurodégénératives.

Parkinsonisme des Caraïbes

Un SPARK atypique a été identifié en Guadeloupe dans les années 1999-2000. Cette entité clinique nouvelle est reconnue par une bradykinésie symétrique, une rigidité musculaire à prédominance axiale, une instabilité posturale avec des chutes précoces et un déclin cognitif avec des signes de dysfonction des lobes frontaux. La plupart des patients ne répondent pas à l'administration de L-dopa [7].

Elle pourrait s'avérer être une tauopathie identique ou très proche de la paralysie supranucléaire progressive, mais la topographie des lésions d'IRM cérébrale est distincte de la paralysie supranucléaire progressive [58].

Plus récemment, un cluster de parkinsonisme atypique de Guadeloupe a été mis en évidence en Martinique [59]. Ce qui suggère qu'il existe dans ces îles francophones de la Caraïbe, un SPARK atypique dont l'origine est inconnue à ce jour. Des facteurs de susceptibilités génétiques ou/et des facteurs environnementaux sont suspectés, notamment, des plantes tropicales contenant des inhibiteurs du complexe I mitochondrial [60].

3.4.4. Épidémiologie

La prévalence de la MPI est de 150 pour 100 000 habitants dans la population générale, s'élevant à 1,5 % au-delà de 65 ans (chiffre croissant en raison du vieillissement de la population). Généralement, elle débute après 60 ans, mais peut survenir à tout âge (10 % avant 40 ans). La MP est la deuxième cause de handicap moteur chez le sujet âgé.

L'âge de début moyen varie selon les SPARK :

- L'incidence du SPAJ est évaluée à 1.5/100 000 personnes/an entre 30 et 50 ans. Il y aurait entre 1000 et 2000 malades en France. Les SPAJ peuvent être familiaux ou sporadiques.
- Classiquement la PSP touche environ 1 personne sur 16 600.
- La prévalence de l'AMS est faible, dix fois moindre que la MPI, estimée entre 1,6 à 5 cas pour 100 000 habitants avec une incidence estimée entre 0,6 et 1,6 cas pour 100 000 habitants par an [61]. Son pronostic est plus grave que celui de la MPI avec une survie médiane d'environ 6 à 9 ans [62], [63]
- La prévalence de la DCB est entre 5-7/100 000.

3.4.5. Génétique et physiopathologie

L'accumulation de protéines anormales et toxiques semble être un mécanisme important de la mort des neurones dopaminergiques, indépendamment des variants

pathogènes spécifiques ou de facteurs environnementaux. De fait, dans la MPI, on observe une accumulation d'α-synucléine, de parkine, d'ubiquitine et d'UCH-L1 (ubiquitin C-terminal hydrolase L1) dans les corps de Lewy. Ceci suggère qu'un défaut de protéolyse en lien avec un dysfocntionnement du système ubiquitine-protéasome pourrait participer au processus dégénératif [51], [64].

La parkine, codée par *PARK2*, joue un rôle important dans le processus d'ubiquitinylation. On rappelle que ce mécanisme permet de dégrader en substrats non toxiques des protéines intracellulaires potentiellement toxiques du fait de leurs anomalies de conformation ou de leur agrégation anomale via le protéasome. De ce fait des variants pathogènes de *PARK2* entraînent une perte neuronale très sélective des neurones dopaminergiques.

L'α –synucléine, codée par *SNCA*, est une des protéines substrats de la parkine : en excès ou dans une conformation anormale, l'α-synucléine va être dégradée par la voie du système ubiquitine-protéasome, dans laquelle la parkine a un rôle de ligase (E3). Dans des conditions de stress oxydatif ou lors de variants pathogènes *SNCA*, l'α- synucléine peut s'agréger sous formes de composés intermédiaires toxiques.

L'UCH-L1, ubiquitin C-terminal hydrolase L1, ou *PARK5*, joue un rôle dans la formation de polymère d'ubiquitine. Quand *PARK5* est muté, l'ubiquitinylation des protéines est altérée, l'activité du protéasome est diminuée, conduisant à l'accumulation toxique de protéines.

En parallèle, on retient que ces processus de dégradation des protéines sont hautement consommateurs d'énergie (ATP). Ainsi, tout dysfonctionnement de la chaîne énergétique mitochondriale endogène ou exogène toxiques peuvent interférer avec la survie cellulaire.

On connaît les gènes *PARK2, PARK6* et *PARK7* en cause dans les formes autosomiques récessives associées à un début précoce du SPAJ. Le gène *LRRK2*, qui code pour la dardarine, peut être en cause dans les formes sporadiques ou familiales dominantes de SPAJ. Le gène *MAPT* est le gène majeur impliqué dans la PSP. Aucun gène n'est associé au diagnostic de l'AMS ou dans la DCB.

Le TABLEAU 10 récapitule les gènes impliqués actuellement dans le diagnostic de la MPI et des SPARK.

Entité clinique	Gène, protéine	Localisation	Type de variants	Rôles	Autres MNDREM
PARK2	PRKN, parkin RBR E3 ubiquitin protein ligase	6q26	SNVs, petites délétions	Composant du complexe ubiquitine ligase multiprotéine E3 pour la dégradation du protéasome	Cancer du poumon et néoplasies de l'ovaire
PARK6	PINK1, PTEN induced kinase 1	1p36.12	SNVs	Sérine / thréonine protéine kinase localisée dans les mitochondries. Protection contre dysfonctionnement mitochondrial induit par le stress.	Aucune
PARK7	DJ1, Parkinsonism associated deglycase	1p36.23	SNVs	Famille des protéines peptidase C56. Régulateur positif de la transcription dépendante des récepteurs aux androgènes. Protection neurones du stress oxydatif et mort cellulaire	Aucune
PARK8	LRRK2, leucine rich repeat kinase 2	12q12	SNVs	Protéine majoritairement cytoplasmique mais ausside la membrane externe mitochondriale	Aucune
FTD	MAPT, microtubule associated protein tau	17q21.31		Protéine tau associée aux microtubules (MAPT) dont l'épissage alternatif génèredes transcrits exprimés sélectivement dans SNC, selon le stade de maturation neuronale et le type de neurone.	Démence frontotemporale, SPARK + démence, maladie de Pick, ophtalmoplégie supranucléaire progressive
PARK5	UCHL1, ubiquitin C-terminal hydrolase L1	4p13	SNVs	Famille des peptidases C12. Thiol protéase, hydrolyse la glycine C-terminale de l'ubiquitine. Spécifiquement exprimé dans neurones et cellules du système neuroendocrinien diffus.	Paraplégie spastique 79 (autosomique récessive)
PARK1	SNCA, synuclein alpha	4q22.1	SNVs	Famille des synucléines. Abondamment exprimées dans cerveau. Aide à l'intégration de la signalisation présynaptique et au trafic membranaire.	Démence à corps de Lewy

TABLEAU 10, *Gènes impliqués dans le diagnostic de MPI et SPARK*

4. Chevauchement clinique et moléculaire

Les chapitres précédents confirment la redondance tant sur le plan phénotypique que moléculaire des MNDREM : des gènes de PARK sont impliqués dans des démences, des gènes de SLA dans des CA ou des neuropathies… On peut donc véritablement parler de chevauchement clinique et aussi moléculaire entre toutes les MNDREM d'intérêt.

Le chapitre qui suit souligne quelques descriptions atypiques et chevauchantes de la littérature ayant trait à toutes les MNDREM de ces travaux.

4.1. Atteinte de la corne antérieure et troubles cognitifs

Dans les années 90, on commence à décrire des formes de SLA associées à des perturbations du système limbique. Quelques cas de SLA sporadiques sont décrits avec des démences et certains aspects neuropathologiques inattendus : une perte neuronale dans le noyau basolatéral de l'amygdale, le noyau accumbens et le subiculum de l'hippocampe, une gliose et une spongiosité des couches superficielles dans tout le cortex cérébral. Des inclusions immunoréactives à l'ubiquitine étaient présentes dans certains neurones des couches cellulaires granulaires de l'hippocampe. La gliose fibreuse était étendue dans la substance blanche sous-corticale et profonde des lobes frontotemporaux caractérisant une entité clinique dénommée Dégénérescence Lobaire Fronto-temporale (DLFT) [65] distinguable par une atteinte démentielle type Alzheimer, affectant la mémoire, le langage, la pensée abstraite, le comportement et les capacités motrices. On considère alors une nouvelle forme de SLA avec démence non Alzheimer, tau- ou alpha-synuclein-négative mais ubiquitine positive.

C'est dans la première décennie du XXe siècle que l'on relie une expansion hexanucléotidique GGGGCC pathogène du gène *C9ORF72* à cette entité clinique originale alliant atteinte du motoneurone et troubles cognitifs [66]. Il est désormais reconnu que la SLA et la dégénérescence lobaire frontotemporale (FTLD) forment un spectre clinique de la maladie avec des caractéristiques cliniques, pathologiques et génétiques se chevauchant. Actuellement, la fonction normale de *C9ORF72* n'est pas claire. De même, le mécanisme cellulaire par lequel l'expansion répétée

pathogène conduit à la neurodégénérescence est également un processus non élucidé.

D'autres associations entre démence et pathologies neuromusculaires ont été établies. Comme cette myopathie à inclusions qui résulte de la conjoncture entre maladie de Paget des os et Démence Fronto-temporale (DFT), associée à des variants pathogènes des gènes *HNRNPA1*, *HNRNPA2B1* ou *VCP*. Ng AS et al signent que la DFT est un pont entre démence et maladies neuromusculaires. [67]

4.2. Troubles cognitifs, syndrome parkinsonien et *ATXN2*

L'ataxie spinocérébelleuse de type 2 (SCA2) est associée à une perte neuronale dans le cervelet, la substantia nigra, le striatum et le globus pallidus, due à l'agrégation intranucléaire de protéines à « strectch » de polyglutamin. En effet, une expansion répétée de CAG dans le gène Ataxin-2 (*ATXN2*) code pour un tronçon de polyglutamine (polyQ). Trente-quatre répétitions CAG ininterrompues ou plus sont associées à l'ataxie cérébelleuse, aux saccades lentes et au parkinsonisme, commençant avant 60 ans.

Interrompues par des motifs CAA, les expansions CAG pathogènes produisent un parkinsonisme isolé sensible à la lévodopa. Young et al décrivent une famille coréenne présentant un phénotype clinique typique de Maladie de Parkinson associée à 40 répétitions CAG et 4 interruptions CAA dans le gène *ATXN2*. Cette expansion est à priori la plus longue expansion associée à un phénotype Parkinson sans ataxie. L'imagerie structurale a montré un cervelet normal. Les images fonctionnelles sont en faveur d'une dégénérescence dopaminergique nigrostriatale et une activité de liaison au récepteur D2 normale, en accord avec le phénotype clinique [68].

Ceci prouve que les variants pathogènes de *ATXN2* en particulier les expansions interrompues devraient être considérées comme une cause du phénotype typique de la maladie de Parkinson même en l'absence d'ataxie cérébelleuse.

Dans les familles où ataxie, démence et parkinsonisme cohabitent, il a été prouvé que si l'expansion dans *ATXN2* génère l'ataxie, la démence est probablement liée à la fois aux expansions répétées de *C9ORF72* et *ATXN2*, laissant penser que la

répétition pathologique ininterrompue d'*ATXN2* peut ne pas avoir le même effet que les allèles interrompus intermédiaires [69].

Les allèles intermédiaires supérieurs à 26 et jusqu'à 39 répétitions CAG représentent un fort facteur de risque de sclérose latérale amyotrophique (SLA) associé aux inclusions neuronales de TDP-43 (TAR DNA binding Protein 43) [70]. Les répétitions intermédiaires *ATXN2* polyQ diminuent le délai de survie dans la SLA [71].

4.3. Accumulations protéiques et atteintes mitochondriales

Si certains variants de l'ADN mitochondrial sont incriminés depuis longtemps dans les maladies d'Alzheimer et Parkinson [72], l'implication de la mitochondrie dans les autres pathologies ND a été démontré au fil du temps, par le biais des processus de stress oxydatif, ou par son implication indirecte avec le système ubiquitine-protéasome.

Dans le registre des MNDREM alléliques à composante mitochondriale, on cite : le syndrome « Fragile X Tremor Ataxia » (FXTAS), qui est la conjonction de tremblement intentionnel et ataxie débutant à l'âge adulte0 Il est associé à une prémutation, c'est-à-dire une expansion en zone grise du gène *FMR1*. Lorsqu'il est muté, ce gène est responsable du syndrome de l'X fragile, première cause de retard mental chez les garçons. FMR1 code pour une protéine impliquée dans le trafic d'ARNm du noyau vers le cytoplasme et joue un rôle central dans le développement neuronal et la plasticité synaptique grâce à la régulation de l'épissage alternatif de l'ARNm.

L'accumulation de protéines à polyglutamine mutantes dans les inclusions intraneuronales est une caractéristique des MNDREM à expansion de stretch de Glutamine.

Les analyses en protéomique menées sur des cellules souches embryonnaires humaines affectées par la maladie de Huntington révèlent une pathologie évolutive impliquant un dysfonctionnement mitochondrial et des perturbations métaboliques [73]. En utilisant des modèles de souris et des tissus cérébraux humains de la maladie de Huntington (HD) et de l'ataxie spinocérébelleuse de type 3 (SCA3), il a été montré que l'*UBQLN2* est sélectivement recrutée pour des inclusions dans la HD

mais pas dans SCA3. La protéine adaptatrice de liaison à l'ubiquitine-protéasome, *UBQLN2*, participe à l'homéostasie des protéines et se localise dans les inclusions de diverses maladies neurodégénératives. Ce qui suggère qu'il existe une hétérogénéité dans les inclusions des MNDREM à polyQ et que des composants du contrôle de la qualité des protéines neuronales peuvent être perturbés de manière différente selon les maladies à polyQ [74].

4.4. Autres chevauchements de MNDREM

Dans les Paraplégies Spastiques Héréditaires (PPSH), 86 gènes sont impliqués. Parmi lesquels des gènes codant pour des molécules de signalisation cellulaire, des protéines de cytosquelette et de trafic cellulaire ou encore des protéines impliquées dans la machinerie mitochondriale.

Le même constat peut être fait pour la Neurodégénérescence avec Surcharge Cérébrale en fer (Neurodegenerative Brain Iron Accumulation NBIA) qui rassemble un groupe de MNDREM caractérisé par un dysfonctionnement progressif du système extrapyramidal (dystonie, rigidité, choréoathétose), et une accumulation de fer dans le cerveau avec présence de sphéroïdes axonaux, généralement limités au système nerveux central. Onze gènes sont actuellement connus, dont *ATP13A2* impliqués dans la PPSH.

✓ Cas spécifique : anticipation dans les MNDREM

La plupart des pathologies à expansion de motifs répétés se transmettent avec une forte probabilité de glissement moléculaire, entraînant une expansion augmentée de la taille du motif et une aggravation des signes cliniques notamment un âge de début précoce. C'est ce qui définit le phénomène d'anticipation. On le retrouve dans les ADCA à variants pathogènes d'expansion ou dans la maladie de Huntington.

Dans la SLA, des corrélations génotype-phénotype ont permis de mettre en évidence des variants de *SOD1* responsables de début précoce [75] ou de diminution du délai de survie. Cependant, dans cette pathologie, un phénomène d'anticipation est décrit dans des familles japonaises avec le variant pathogène p.Gly94Ser, sans qu'aucun mécanisme moléculaire ne puisse le justifier [76].

5. Quid du continuum clinique et moléculaire dans les MNDREM…

Les Maladies Neurodégénératives à Expression Motrice (MNDEM) sont un groupe très hétérogène de pathologies neurodégénératives de physiopathologie acquise ou innée, pouvant se déclarer à tout âge. Elles impactent de façon ciblée, les structures neurologiques centrales ou périphériques impliquées dans l'activité motrice. Ces structures sont fonctionnellement et topographiquement proches des centres de cognition. C'est ainsi que dans certaines conditions, la neurodégénérescence des structures motrices impacte les fonctions cognitives. Elles se traduisent par des troubles de la marche et /ou de l'équilibre, une diminution ou une absence de mouvement, une atteinte cognitive variée pouvant aller jusqu'à la démence.

La revue de la littérature laisse penser que les MNDREM sont un véritable éventail de pathologies neurologiques dont les mécanismes moléculaires sont en fait étroitement liés. Les premières descriptions basées sur la sémiologie n'ont pas résisté à la complexification descriptive en lien avec les chevauchements cliniques et cytopathologiques entre MNDREM. Même les classifications moléculaires ont trouvé leurs limites puisqu'il existe des MNDREM alléliques, c'est à dire, des pathologies causées par un même gène, ou des MNDREM dont les variants génétiques concourent à des modifications phénotypiques d'autres pathologies neurodégénératives.

La représentation en éventail de la FIGURE 11 ci-après, est une proposition schématique du continuum phénotypique des MNDREM en fonction des types d'atteintes motrices vis à vis des processus cellulaires impactés. La sélectivité des symptômes s'entend de la faible redondance phénotypique ; postulant que les perturbations des processus de communication cellulaire et les autres processus, se retrouvent associés à plus de MNDREM que les processus altérant la fonction mitochondriale.

Le chevauchement clinique et moléculaire dans les MNDREM est à la fois une intrigue et une clé de compréhension des mécanismes sélectifs de dégénérescence. Il laisse présager des voies de signalisation moléculaire croisées dont les éléments communs agiraient comme des régulateurs capables de modifier des symptomatologies à façon. Il interroge l'oligogénisme.

Cet état des lieux des connaissances actuelles pose la question de la précision du diagnostic clinique et moléculaire dans les MNDREM. Pour pallier cette question et compte tenu de la fragilité des classifications cliniques, le recours aux nouvelles technologies polychrones est naturel.

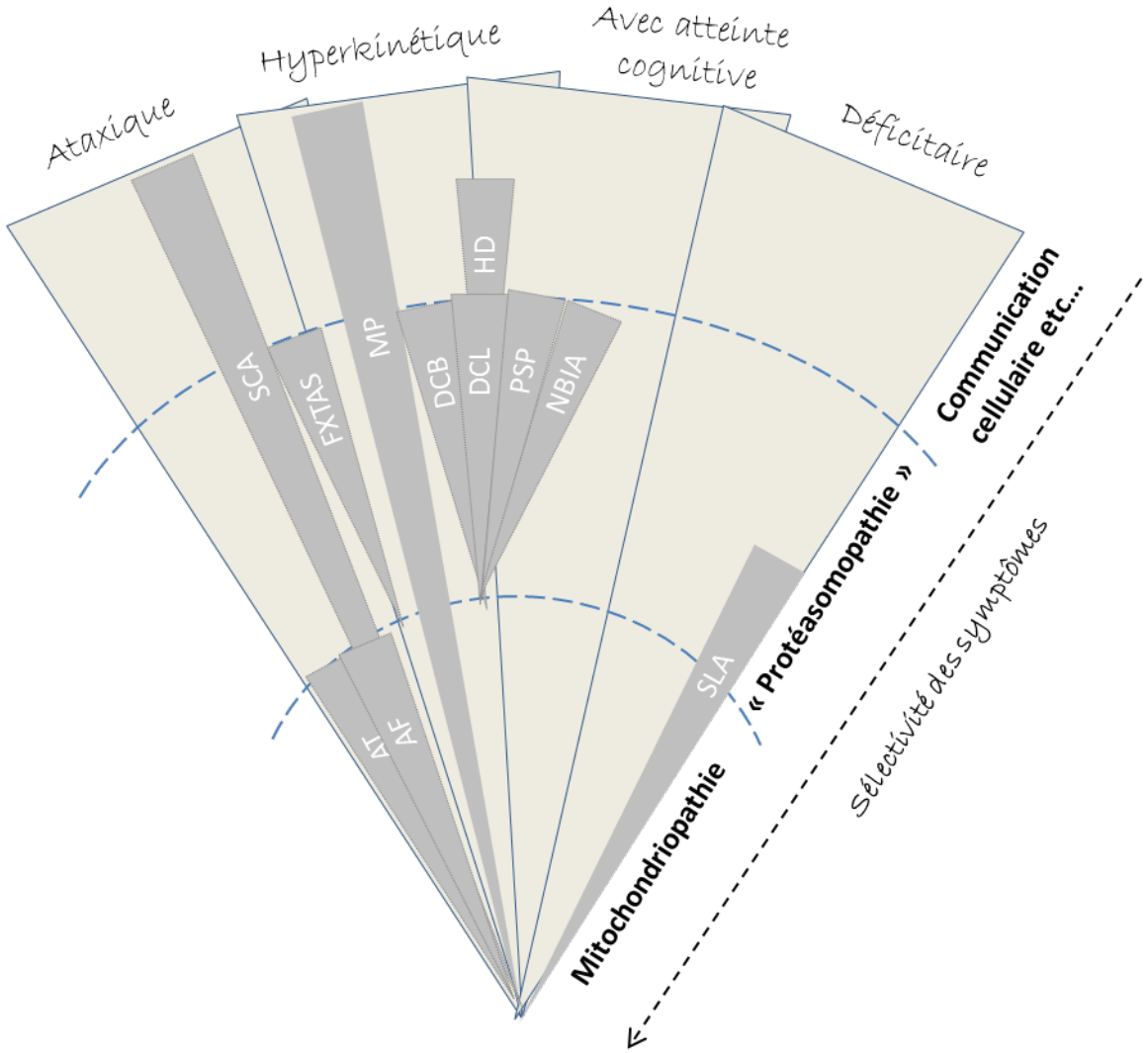


FIGURE 11, *représentation schématique en éventail du chevauchement clinique des MNREM d'intérêt en fonction de l'atteinte motrice associée*

Actuellement, dans les pratiques médicales, le Séquençage Haut Débit (SHD) ou Next Generation Sequencing, (NGS) permet de surseoir aux limites du séquençage Sanger qui ne permet d'explorer qu'un gène à la fois. Le NGS, lui, autorise le séquençage de panel de gènes, de toutes les séquences exoniques ciblées (exome ou Whole Exome Sequencing, WES) voire du génome complet (ou Whole Genome Sequencing, WGS). S'imposant ainsi comme une technologie de choix dans la problématique des symptômes chevauchants ou atypiques, mais également dans celle des gènes modificateurs ou dans le plurigénisme.

Ce premier chapitre montre que la plupart des variants pathogènes impliqués dans les MNDREM sont des SNVs mais on relève tout de même quelques redondances étiologiques moléculaires comme : la nature des variants pathogènes, avec une forte représentation des variants pathogènes par expansions de motifs répétés ; ou l'ontologie des gènes, souvent associés aux processus mitochondriaux, au système ubiquitine-protéasome ou à tout autre processus cellulaire impactant le cytosquelette ou les systèmes de communication cellulaire.

Les CNVs (Copies Variations Number), sont peu décrits et explorés alors qu'ils pourraient faire partie des mécanismes étiologiques dégénératifs, comme les variants pathogènes dynamiques, en particulier le raccourcissement des télomères par exemple, comme nous l'avons vu. A ce jour, le NGS permet principalement la détection de SNVs. Mais il est en perpétuel perfectionnement et autorise de plus en plus la détection de CNVs et de variants par expansion de motifs répétés, notamment grâce à l'adaptation des méthodes de capture et des outils bio informatiques.

L'absence de données génétiques et démographiques sur la population de notre territoire et plus largement dans la région des Antilles françaises, nous a conduits à mettre en œuvre une stratégie d'investigation des MNDREM ayant pour objectif les suivants :

- Dresser la cartographie des Maladies Neurodégénératives à Expression Motrice (MNDREM)
- Trouver des variants géniques pathogènes responsables de MNDREM sans étiologies

Pour cela, nous avons collecté et analysé les données clinico-biologiques des patients atteints de MNDREM vus dans le service de référence de diagnostic et de

suivi de ces pathologies en Martinique, le CERCA. Et nous avons mis en œuvre pour la première fois en Martinique, les méthodes d'analyses bioinformatiques nécessaires au décryptage des données NGS de 2nde génération. Pour tenter de mettre en évidence de nouveaux gènes impliqués dans les MNDREM de notre population, nous avons sélectionné un échantillon de patients selon des crirères précis. Nous avons employé la méthode CoDEseq (Copy number variation Detection and Exome sequencing) qui permet de détecter SNVs et CNVs dans l'exome d'un individu.

Enfin, pour interroger la problématique de chevauchement clinique, nous avons fait l'essai d'une méthode de modélisation des Interaction Protéines-Protéines (IPPs) en utilisant les résultats de la caractérisation moléculaire des patients atteints de MNDREM. Il s'agit d'une méthode de prédiction bioinformatique largement utilisée dans les questions de recherche de voies de signalisation et de biomarqueurs pour *in fine*, définir des molécules à impact thérapeutique. Les interactions entre gènes impliqués dans les MNDREM, en particulier *JPH3*, ont été recherchées et modélisées par ce moyen.

CHAPITRE II

Conditions de mise en œuvre

CHAPITRE II – Conditions de mise en œuvre

Dans ce chapitre, nous présentons les moyens et méthodes mis en œuvre pour l'étude démographique et génomique des MNDREM d'intérêt.

Le travail s'est déroulé principalement en 2 phases : l'une clinique, l'autre expérimentale.

Nous avons sous-traité les analyses expérimentales à la plateforme GENOSCREEN-LIGAN à Lille. Nous avons par contre développé au sein du pôle Martinique de l'université des Antilles, les méthodes bioinformatiques d'analyses de données issues du Séquençage Haut Débit (NGS) sans lesquelles le travail expérimental serait complètement inexploitable.

Nous avons donc déployé pour la 1ere fois en Martinique, les moyens informatiques nécessaires à l'analyse des données bioinformatiques de NGS.

1. Collecte de données cliniques, biologiques et génétiques

Comme le montre la FIGURE 12 ci-après, les travaux se sont donc déroulés en 2 phases :

- La collecte de données pour la caractérisation clinique et moléculaire des MNDREM d'intérêt
- La caractérisation génomique dans le cadre de l'étude par la méthode CoDE-seq d'un échantillon de patients sélectionnés

PHASE 1	PHASE 2
CLINICAL CHARACTERISATION	MOLECULAR CHARACTERIZATION
Data collection	CoDE-seq
1st Inclusions criteria • Rare NDs motor impairement +/- dementia • All onset ages	2nd Inclusions criteria • Early onset • Familial history • No identified molecular cause
CERCA *Retrospective study, Descriptive and analytic study, Followed cases of*	**INSERM UMR 8199, LIGAN-Genoscreen©**, Lille *Next Generation Sequencing, Medexome* **BIOSPHERES**, Martinique *Bioinformatics analysis*

FIGURE 12, *Présentation de la méthodologie des travaux, en 2 phases*

1.1. Critères d'inclusion et d'exclusion

1.1.1. Étude démographique des MNDREM

Le travail de caractérisation clinique a débuté par la collecte rétrospective de données cliniques et biologiques de patients suivis au centre de référence. Parfois une actualisation des données clinico-biologiques et généalogiques de la population ciblée a été possible par le bais de consultations de suivi ou de conseil génétique.

Les critères d'inclusion de cette première phase concernaient :

- Les pathologies rares à expression motrice +/- démence
- De tous âges de début

Les critères d'exclusion portaient sur les :

- Formes congénitales
- Myopathies
- Syndromes neurodéveloppementaux
- Les étiologies acquises (vasculaire, immunologique, etc…)

1.1.2. Étude génomique d'un échantillon de MNDREM

Pour la sélection des cas à étudier sur le plan génomique, nous avons exclu :

- Les atteintes motrices causées par des anomalies de la substance blanche (leucopathies, encéphalopathies),
- Les syndromes auto-immuns (comme les ataxies anti-GAD par exemple)

Nous avons appliqué les critères suivants :

- Cas sans étiologies génétiques
- Patients majeurs
- Dont l'ADN était disponible au centre de référence

Et présentant à la fois :

- Des arguments d'héritabilité (antécédents familiaux évocateurs)
- Et, des arguments de prédisposition (âge de début précoce, symptomatologie atypique)

Nous avons utilisé un schéma d'inclusion représenté par la FIGURE 13.

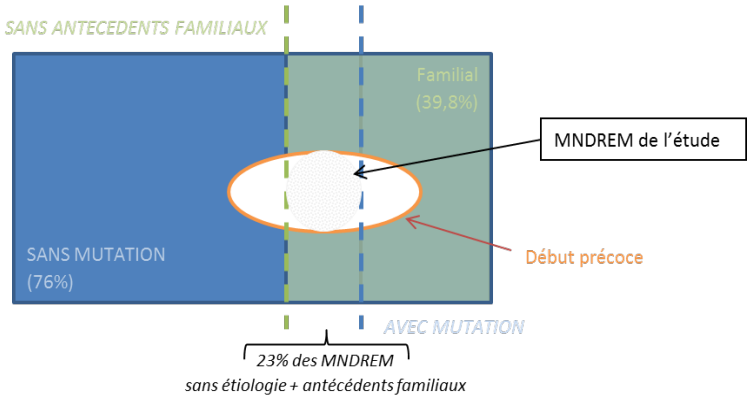


FIGURE 13 – *Schéma représentatif de l'échantillonnage de l'étude*

1.2. Cadre législatif et éthique

Tous les patients inclus sont issus de la filière de soin du centre de référence. La caractérisation des données cliniques et génomiques fait partie intégrante de la filière de soins. Mais conformément aux exigences éthiques et legislatives, des dipositions de déclaration et d'information des instances et des participants ont été prises. Les documents relatifs à ces dispositions sont en annexe.

1.2.1. Autorisation et accord pour l'utilisation des données

Les travaux ont reçu un avis favorable N°2020/076 de l'IRB du CHU de Martinique (annexe 1). Une déclaration CNIL correspondant au référentiel MR-004, et placée sous la responsabilité d'un délégué à la protection des données du CHUM (Mr Gérald GALIM), a été effectuée.

Une fiche information (annexe 2), conforme aux dispositions de l'article 13 ou 14 du règlement général sur la protection des données (RGPD), a été rédigée, adressée aux patients concernés, et affichée dans les salles d'attente et de consultation du centre de référence.

1.2.2. Caractérisation génétique des individus

Dans le cadre de leur diagnostic, les patients signent un consentement éclairé à l'issu de la consultation de conseil génétique.

Ce consentement (annexe 3), conforme à par l'arrêté du 27 mai 2013 définissant les règles de bonnes pratiques applicables à l'examen des caractéristiques génétiques d'une personne à des fins médicales, autorise la caractérisation génétique et les recherches en lien avec les pathologies dont ils souffrent.

L'étude de la caractérisation génétique des individus sélectionnés pour notre étude s'inscrit dans le cadre de l'utilisation des échantillons biologiques, conservés au décours des investigations diagnostiques, hors loi Jardé. Dans ce cadre une

déclaration CNIL correspondant au référentiel MR-001 a été faite au n°2219621v0 (annexe 4).

1.3. Système d'encodage des données collectées

Au moyen d'un fichier Excel, les informations identitaires et clinico-biologiques des individus présentant les MNDREM d'intérêt, ont été collectées et encodées à partir du dossier des patients, comme le montrent les figures ci-après.

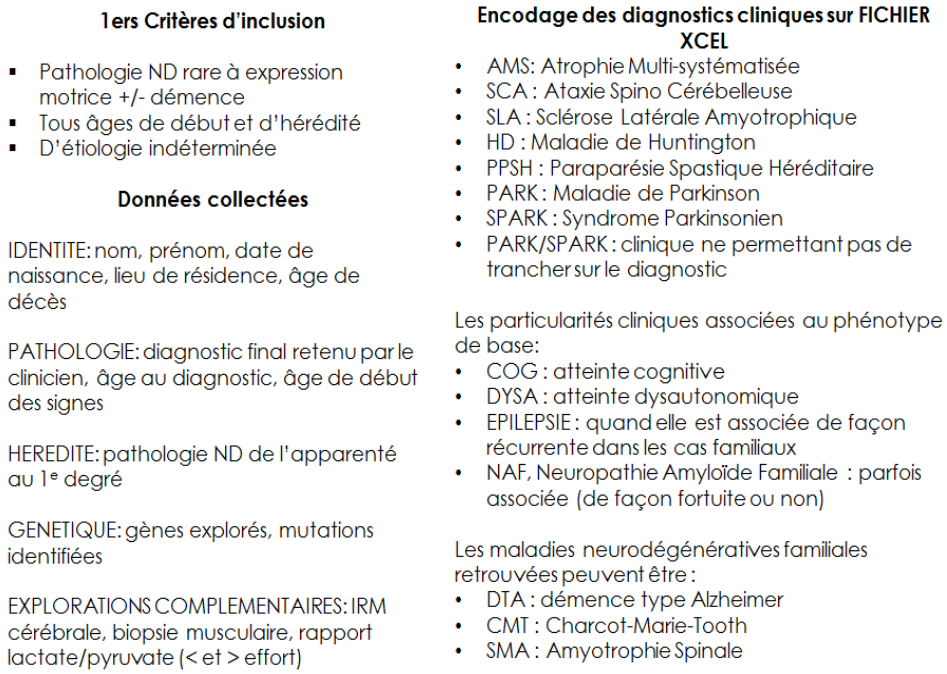
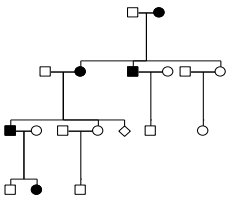


FIGURE 14, *données clinico-biologiques et encodage de la base de données de MNDREM d'intérêt*

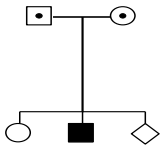
Il a été parfois nécessaire de compléter les informations familiales en établissant un arbre généalogique en propre lors de consultations de conseil génétique. On rappelle ci-après, les modes de transmission mendéliens et mitochondriaux, ainsi que les notions de pénétrance, expressivité variable et hétéroplasmie leur correspondant.

1.3.1. Modes de transmission mendéliens



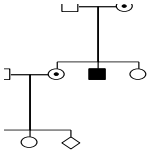
a) Autosomique dominante

La pathologie se transmet sur plusieurs générations à partir d'un individu atteint et quel que soit le sexe du descendant.



b) Autosomique récessif

La pathologie apparaît dans la descendance de 2 individus porteurs d'un allèle pathogène dans le gène responsable. La maladie s'exprime chez un individu qui hérite conjointement des alléles délétères parentaux.



c) Lié à l'X

La pathologie affecte les individus de sexe masculin qui héritent d'un variant pathogène d'un gène porté par le gonosome X maternel. Les femmes transmettrices sont a- ou pauci-symptomatiques.

FIGURE 15, pedigree type en fonction des modes de transmission
a)Dominant ;b) Récessif ; c) Lié à l'X
carré : hommes ; rond : femme ;
noircit : atteint ; partiellement noircit : porteur

Expressivité variable

Se dit de l'étendue interindividuelle des signes de la pathologie. Ils peuvent être plus ou moins présents d'un individu à l'autre.

Pénétrance

S'entend de la probabilité qu'un variant pathogène s'exprime. On parle de pénétrance incomplète lorsque statistiquement, tous les individus porteurs de variants pathogènes ne sont pas atteints. La pénétrance peut varier avec l'âge et le sexe. Elle peut être dépendante de facteurs génétiques ou environnementaux.

1.3.2. Mode de transmission mitochondrial

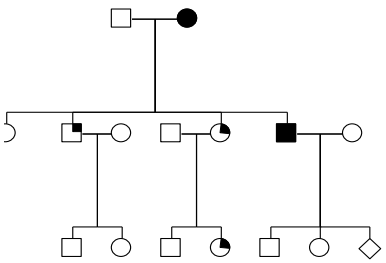


FIGURE 13, pedigree type transmission mitochondriale
carré : hommes ; rond : femme ; noircit : atteint ; partiellement noircit : porteur

Quand le variant causal est situé sur l'ADN mitochondrial, il ségrège via l'organelle qui est transmis uniquement par le milieu nutritif des ovules maternels. Le pedigree ci-contre est un exemple de transmission familiale.

Ainsi, une pathologie causée par une variation délétère de l'ADNmt touchera potentiellement la descendance des femmes atteintes, quel que soit le sexe. En revanche, un individu atteint de sexe masculin ne peut pas transmettre la variation délétère. Il existe une grande variabilité d'expression phénotypique dans une même famille, notamment à cause du phénomène d'hétéroplasmie.

Hétéroplasmie

Signe la présence de génome différent dans une cellule : les variants pathogènes de l'ADNmt peuvent ne concerner qu'un certain nombre de mitochondries. L'hétéroplasmie est responsable des variabilités d'expression de la pathologie. L'atteinte étant d'autant plus importante que le nombre de génomes mitochondriaux mutés est important. Le phénomène d'hétéroplasmie impacte aussi les tissus. Ainsi, certains peuvent être plus touchés que d'autres.

2. Généralités sur le séquençage 2e génération

Le Séquençage 2eme génération ou Next Generation Sequencing (NGS) est une méthode de Séquençage dite Haut Débit (SHD), c'est-à-dire permettant le ciblage de plusieurs régions du génome en même temps. Elle permet le séquençage dit « massif en parallèle », c'est-à-dire sur plusieurs ADN en même temps.

Mis à l'épreuve à l'époque du challenge du séquençage du génome humain au début des années 2000, il est aujourd'hui largement répandu dans les laboratoires de recherche et de diagnostic moléculaire.

2.1. La technique

Les résultats du séquençage 2eme génération sont obtenus après la réalisation des étapes suivantes : Préparation de librairie, Amplification clonale, Séquençage, Analyse bioinformatique, Interprétation de variants.

2.1.1. Préparation de librairie et enrichissement

Cette étape permet d'une part de combiner plusieurs patients dans une même analyse de séquençage (Parallélisation des échantillons), d'autre part de cibler les régions d'intérêt à amplifier (préparation de librairie).

La parallélisation des échantillons permet de garantir le séquençage de plusieurs patients en même temps. Cette étape est rendue possible grâce aux index ou codes-barres, qui sont des séquences nucléotidiques synthétiques ajoutées au fragment d'ADN. Un index unique est attribué à chaque patient.

La librairie est l'ensemble des séquences génomiques à analyser, prêt à être séquencés. L'enrichissement se dit de l'amplification des séquences cibles.

D'un point de vue technique, constitution de librairies et enrichissement sont étroitement liés. Communément, trois approches permettent la constitution de

librairie. La FIGURE 17, ci-après, présente les différentes méthodes de préparation de librairies. [77]

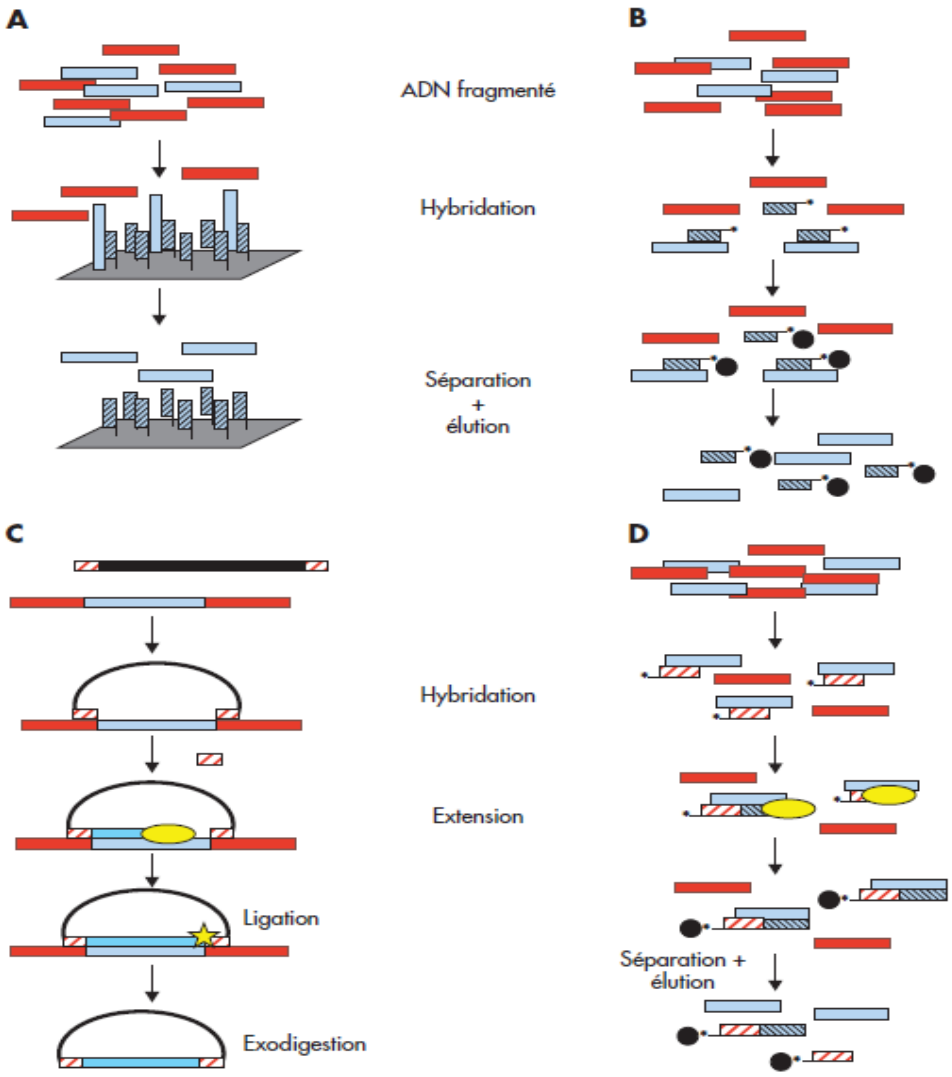


FIGURE 17. *Différentes méthodes d'enrichissement préliminaires à la préparation des librairies, selon Teer JK et al.*[77] *; Les barres bleu clair représentent les séquences génomiques d'intérêt et les barres rouges les séquences non désirées. Enrichissement par capture en phase solide (A) ou phase liquide (B) ; enrichissement par circularisation (C) ; enrichissement par PCR multiplex (D).*

- ✓ **Enrichissement par capture**

En phase solide

Des sondes complémentaires aux séquences génomique cibles, sont hybridées sur un support solide, une puce appelée « flowcell ». Après fragmentation de l'ADN à analyser, il y aura hybridation selon la règle de complémentarité de brins. Puis le lavage de la puce permettra d'éliminer les fragments non hybridés.

L'amplification clonale est opérée par la technique de **Bridge PCR :**

Les fragments sont amplifiés par PCR avec formation de ponts entre séquences adaptatrices. Il y a alors obtention de groupes d'amplicons issus d'un même fragment initial à la surface de la flowcell, ce sont des clusters. Cette méthode est largement utilisée par la technologie Illumina. Elle permet d'utiliser par la suite un séquençage de type reversible terminator.

En phase liquide

Les sondes d'hybridation complémentaires à l'ADN, sont marquées à la biotine. Après, fragmentation et hybridation des séquences cibles, des billes de streptavidine s'accrochent à la biotine par liaison affine. Les fragments non accrochés aux billes sont élués.

L'amplification clonale est opérée par la technique de **PCR Emulsion**:

Suite à la formation d'une émulsion (mélange huile/eau), un fragment d'ADN et une bille sont capturés au sein d'une goutte. Des adaptateurs présents sur la bille permettent l'amplification de l'ADN sélectionné. Après plusieurs étapes d'amplification par PCR, on obtient de nombreux amplicons à la surface de la bille. Cette méthode est prônée par Roche et autorise le séquençage type semi-conducteur, pyroséquençage ou séquençage par ligation. La FIGURE 18 ci-après, schématise les 2 méthodes d'amplification. [78]

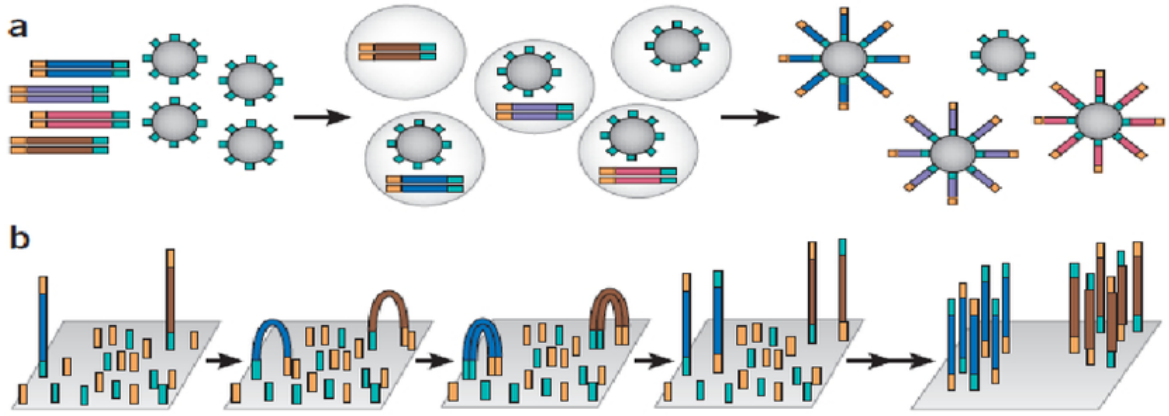


FIGURE 18 - *Schéma des méthodes d'amplification clonale des enrichissements par capture* [78]

a) PCR Emulsion ; b) Bridge PCR

✓ Enrichissement par amplicons

Des amorces biotynilées sont mises en présence de l'ADN fragmenté. Une polymérase permet l'amplification par PCR Multiplex des régions d'intérêt. Des billes de streptavidine permettent la rétention des amplicons après élution. Ensuite les amorces sont digérées par une enzyme. Puis il y a ligation des adaptateurs et barre-codes. Cette approche est simple mais nécessite de bonnes quantités et qualités d'ADN de départ.

✓ Enrichissement par circularisation

L'ADN est fragmenté par un cocktail d'enzymes de restriction. Les sites de coupure peuvent être prédits par un algorithme informatique. On synthétise donc une sonde de séquence universelle double brin, flanquée aux extrémités de séquences complémentaires de la région d'intérêt simple brin. L'hybridation des sondes par leurs extrémités, entraine une circularisation du fragment d'ADN. L'ADN ainsi capturé est alors amplifié par PCR multiplex.

Chaque méthode présente des avantages et inconvénients que Mertes F et al. ont comparées dans la Figure 19 ci-après. [79]

Technique d'enrichissement	Fournisseur	Caractéristiques	Avantages	Inconvénients	Nombre de loci (taille cible)	Préparation de la librairie
Capture d'hybrides						
Support solide	Agilent SureSelect, Roche NimbleGen Sequence Capture Arrays	Régions cibles moyennes à grandes, personnalisées ou préconfigurées, multiplexage possible, kits prêts à l'emploi	Facile à produire, ensemble de cibles importantes	Grande quantité d'ADN nécessaire (3-10 µg), équipement dédié	10^4-10^6 (1-50 Mb)	Avant enrichissement
Phase liquide	Agilent SureSelect, FleXGen FleXelect, MYcroarray MYselect, Roche NimbleGen SeqCap EZ		Facile à utiliser, faible quantité d'ADN nécessaire (< 1-3 µg)			
Circularisation	HaloGenomics	Kits personnalisés et panel cliniquement pertinent	Appareils non dédiés, spécificité importante, faible quantité d'ADN nécessaire (< 1 µg)	Kit exome non disponible à ce jour	10^2-10^4 (0,1-5 Mb)	Pendant enrichissement
PCR						
Multiplex	Life technologies Ampliseq, Multiplicom MASTR, Fluidigm Acess Array	Régions cibles plus courtes, couverture par tiling, PCR multiplex de 150-200 amplicons (150-450 bp)	Facile à exécuter, coût raisonnable, couverture identique	Efficacité largement influencée par conditions de PCR, > 10 µg d'ADN nécessaire pour grand panel	10^2-10^4 (0,1-5 Mb)	Après enrichissement
Micro-gouttelettes	RainDance	Régions cibles plus courtes, couverture par tiling, PCR en émulsion jusqu'à 20 000 amplicons (150-1 500 bp)	Couverture identique, faible quantité d'ADN nécessaire	Relativement coûteux, équipement dédié	10^3-10^4 (jusqu'à 10 Mb)	

FIGURE 19 - *Comparaison des différentes techniques d'enrichissement – tiré de la littérature* [79]

2.1.2. Méthodes de Séquençage

Les technologies 2eme génération sont basées sur le principe du « wash and scan ». C'est à dire une succession de cycles permettant dans un premier temps, l'ajout de polymérase et de réactif pour incorporer une base, dans un deuxième temps, le lavage. Au contraire, les technologies de 3e génération qui permettent le séquençage de molécules uniques ne nécessitent pas de « wash and scan ». Trois sociétés se disputent le leadership de ces technologies et sont spécialisées chacune dans une méthode de séquençage.

✓ *Pyroséquençage (Roche 454/GS)*

Développé par la société Roche Diagnostics, il est basé sur le principe de la PCR en émulsion et le pyroséquençage.

Le pyroséquençage s'appuie sur la détection, par quantification de lumière émise, du pyrophosphate libéré lors de la réaction de polymérisation en chaîne. Le milieu réactionnel est composé de la matrice d'ADN simple brin hybridée à une amorce, l'ADN polymérase, l'ATP sulfurylase, la luciférase et l'apyrase, les substrats 3'-phosphoadénosine 5'-phosphosulfate (PAPS) et luciférine. L'ADN polymérase incorpore les dNTP complémentaires sur la matrice. Cette incorporation libère un pyrophosphate. L'ATP sulfurylase convertit le pyrophosphate en ATP en présence du PAPS. Cet ATP agit comme substrat pour la conversion de la luciférine en oxyluciférine catalysée par la luciférase, ce qui génère de la lumière visible en quantités proportionnelles à la quantité d'ATP. Les nucléotides non incorporés (en excès) sont dégradés par l'apyrase, et la réaction peut reprendre avec un nouveau nucléotide.

La lumière produite au cours de cette réaction est détectée par une caméra et analysée par un programme informatique qui traduit le signal sous forme d'un pic sur un pyrogramme. La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps.

✓ *Semiconducteur (Life technologies Ion Torent)*

Promu par la société Life science, le séquençage semiconducteur est basé sur la libération de protons lors de l'ajout d'un nucléotide. Cette technique débute par l'hybridation d'une amorce universelle et par utilisation de nucléotides non marqués. Pour chaque cycle de séquençage, il y a injection séquentielle d'une des 4 bases séparément et mesure du pH. Si la base est intégrée dans le fragment d'ADN en cours de synthèse, il y a libération de proton et variation du pH. L'intégration de plusieurs bases identiques lors d'une même injection est possible si la séquence est un homopolymère. Les modifications de pH sont ensuite converties en signal pouvant être lu et analysé.

✓ *Par ligation, SoLiD (Life technologies)*

SoLiD veut dire “Sequencing by Oligonucleotide Ligation and Detection”. La première étape d’amplification de cette méthode est identique à celle du pyroséquençage mais la stratégie de séquençage est très différente.

La réaction de séquençage repose sur un système complexe de cycles de ligation séquentielle d’oligonucléotides et de clivage. Après détermination d’une paire de bases identifiable par un fluorophore, l’oligonucléotide ligué est enlevé et ce processus est répété plusieurs fois.

✓ *Par enzyme terminator (Illumina)*

Le principe du séquençage « Illumina » est équivalent au séquençage Sanger : utilisation de nucléotides modifiés marqués par fluorescence, mais avec un terminateur de chaine réversible. Il débute par hybridation d’une amorce universelle. Pour chaque cycle de séquençage, il y a présentation des 4 bases simultanément et incorporation d’une base et une seule dans le fragment d’ADN en cours de synthèse avec émission d’une fluorescence et acquisition du signal lumineux.

La FIGURE 20, ci-après, présente une comparaison de performance des différentes méthodes de séquençage.

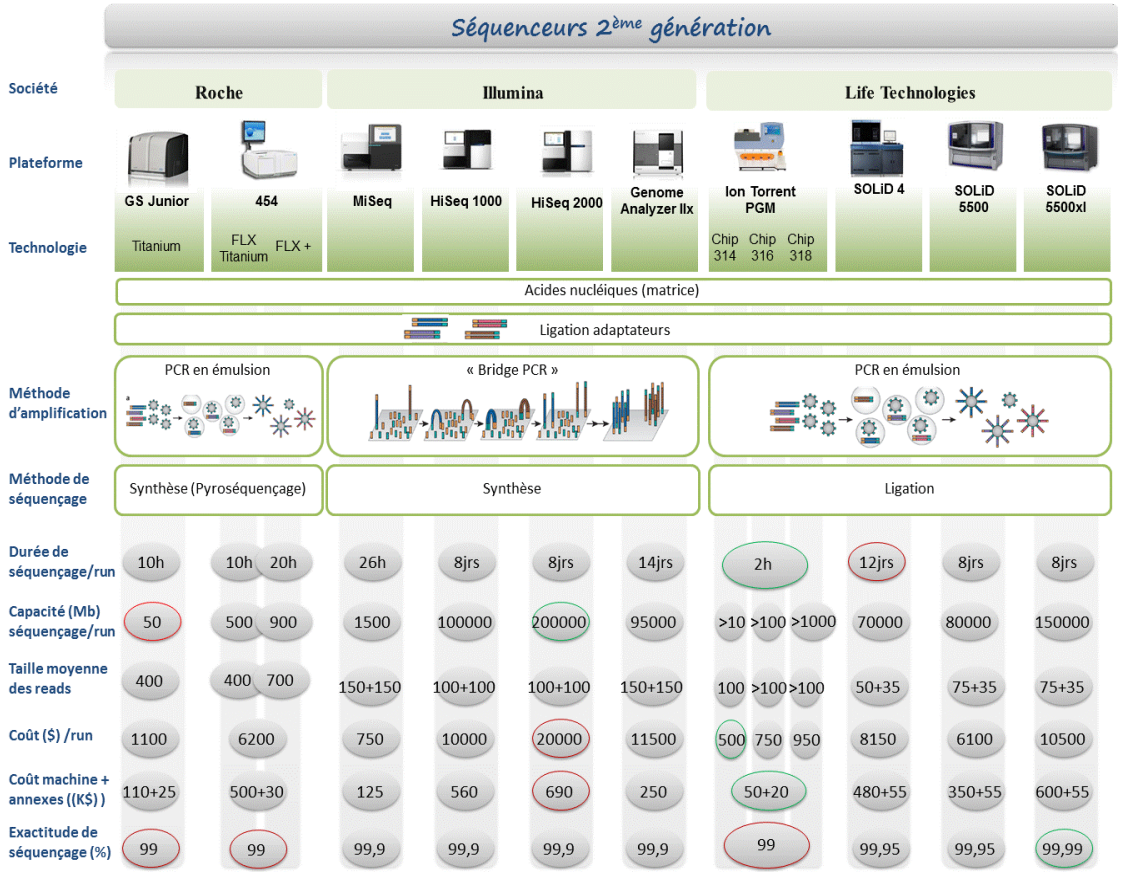


FIGURE 20 – *Données comparatives des caractéristiques générales des séquenceurs nouvelle génération, produites par la société Life Technologies pour la promotion de l'Ion Torrent (http://www.biorigami.com/?tag=principe-ion-torrent-pgm*

2.1.3. Champ lexical du SHD

Des méthodes nouvelles et un champ lexical adapté sont nés de ces technologies :

Avec cette nouvelle technologie, de nouvelles notions sont apparues. La FIGURE 21, ci-après, schématise les plus importantes à partir des chémas de séquençage Paired-end et single-end inspiré de Zhernakova et al. [80].

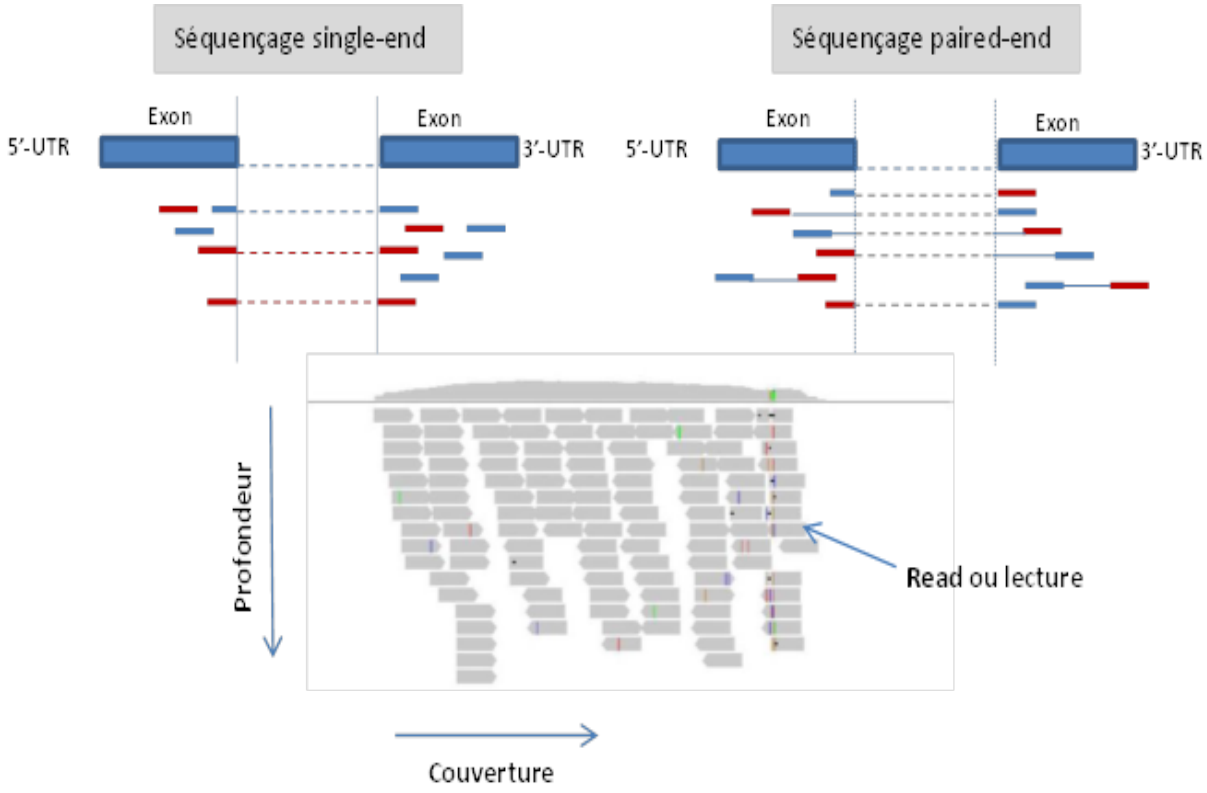


FIGURE 21 – *Représentation visuelle des notions de SHD.*

Lecture ou read : Un read c'est la séquence en elle-même

La longueur de lecture/read : Correspond au nombre de bases successives obtenues pour le séquençage d'un fragment donné.

Paired-end et Single-end : Le séquençage paired-end permet de séquencer les deux extrémités d'un fragment, l'avantage est de produire 2 fois le nombre de lecture. A l'opposé, on peut effectuer un séquençage single-end.

La profondeur : Correspond au nombre de lectures à une position donnée. Elle est quantifiée par le terme X. Par exemple, 20X correspondent à une profondeur de 20 lectures à une position donnée.

La couverture : C'est le nombre de bases séquencées par rapport au nombre de bases de la séquence ciblée initialement.

La qualité : Est exprimée en en QPhred c'est-à-dire, la probabilité p d'erreur de mauvaise identification de la base. QPhred=-10log10p.

2.2. La bioinformatique

En principe, quelle que soit la méthode de séquençage, un signal luminescent sera produit à chaque cycle. Ce signal traduit la base incorporée à un instant donné. Selon la densité de régions ciblées, ou le nombre d'échantillons parallélisés, l'information à traiter sera plus ou moins volumineuse.

Les premières acquisitions informatiques issues du séquençage sont donc des valeurs ou des images qui sont traduites par les calculs informatiques en séquences. Un certain nombre d'étape de calcul va être indispensable avant l'analyse des résultats produits par échantillons.

Le dispositif informatique compte donc pour une bonne partie de la méthode de NGS, et le contrôle qualité est donc important avant toute interprétation. Les données de NGS nécessitent l'utilisation d'un nouveau domaine de compétence, la bioinformatique.

2.2.1. Hardware : Réseau, serveur, calculateur

Les principaux composants informatiques de l'analyse NGS sont les calculateurs hautes performances (HPC), les serveurs de stockage et le réseau. Le choix de capacité des différents éléments est fonction de la taille des séquences analysées.

Les fichiers générés sont très volumineux. En ordre de grandeur, un fichier texte fait environ 500 ko, une photo 10 Mo et un film 750 Mo. Or, un run d'une technologie type terminator enzyme qui représenterait environ 30 milliards de pb génèrerait des données brutes images d'environ 50 Go.

Les fichiers transitent via le réseau grâce à des protocoles de transfert et de partage de fichiers. Les protocoles les plus utilisés sont FTP (File Transfer Protocol), Rsync (Remote Synchronization), SMB (Server Message Block), et NFS (Network File System).

Le stockage à long terme peut se réaliser en local sur des serveurs hébergés en interne ou de façon délocalisée grâce à des serveurs distants type Cloud mais ces serveurs doivent être certifiés « hébergeur de santé » il s'agit de données médicales.

La puissance d'un calculateur est mesurée grâce au FPU (Floating Point operation per second Unit). Ces opérations sont réalisables grâce aux processeurs (CPU : Central Processing Unit) multi-cœurs qui permettent de les réaliser en parallèle, ce qui permet de réduire considérablement le temps d'analyse. Il est également nécessaire de disposer d'un volume important de mémoire vive ou RAM.

2.2.2. Software

On utilise plusieurs outils informatiques pour transformer l'information expérimentale en données interprétables. L'analyse bioinformatique se déroule en 3 grandes étapes ou niveau d'analyse :

- **L'analyse primaire** regroupe les étapes bioinformatiques nécessaires à l'attribution des données à chaque échantillon ;
- **L'analyse secondaire** déroule les analyses bioinformatiques de révélation de séquence ; elle est modulable en fonction des variants à détecter.
- **L'analyse tertiaire**, s'entend de l'examen de la pertinence biologique des variants identifiés.

Chaque niveau d'analyse regroupe un ensemble de méthodes et calculs informatiques automatisables. C'est ce qu'on appelle le pipeline.

✓ Analyse primaire

Le Base Calling

Convertit le signal brut du séquenceur (lumière ou variation de pH) en séquence. Il permet aussi d'estimer la qualité des bases obtenues. Un fichier FASTQ (FASTA Quality) est alors généré. Ce fichier est composé d'un bloc de trois lignes pour chaque séquence : l'identifiant de la séquence, la séquence nucléotidique et le score de qualité.

Le Demultiplexage

Permet d'attribuer chaque séquence lue à un échantillon donné grâce aux index des adaptateurs.

✓ Analyse secondaire

C'est l'ensemble des étapes de calculs allant de l'alignement des séquences contre une séquence de référence, à la génération de fichiers au format Variants Caller Format (VCF), c'est à dire les fichiers permettant l'identification des variations par rapport à la séquence de référence.

Le Trimming

Permet d'éliminer les séquences d'amorce et les bases de mauvaise qualité. La qualité est obtenue par la scorification de Q-Phred. Il s'agit de la probabilité p d'erreur de la base (QPhred = -10log10p). Par exemple Q20 correspond à une probabilité d'erreur de 1% et Q30 de 0.1%.

L'alignement

Consiste à aligner contre une séquence de référence. Celle-ci peut être soit des séquences ciblées, soit la totalité du génome. Actuellement 2 versions du génome humain sont couramment utilisées GRCh37 (ou hg19) et GRCh38 (ou hg20). Plusieurs outils permettent de réaliser l'alignement, comme Bowtie, BWA ou SAMtools. Ils génèrent des fichiers de format SAM (Sequence Alignment Map), qu'il est d'usage de compresser en format BAM (Binary Alignment Map).[80]

SAM (Sequence Alignment / Map) est un format générique permettant de stocker de grands alignements de séquences nucléotidiques.

Le SAM se compose de onze paramètres regroupés sous le terme CIGAR (Concise Idiosyncratic Gapped Alignment Report). Il s'agit d'une codification des alignements par rapport à la séquence de référence. Il est possible d'estimer la qualité des données via le SAM : taux d'alignements, distribution des scores d'alignement (de 70 à 95% attendus) et profondeur de séquençage des cibles.

Le fichier BAM peut être lu directement avec des navigateurs de génome tel qu'IGV (Integrative Genomic Viewer), pour visualiser à la fois la séquence, son sens, ainsi que la couverture et la profondeur de séquençage. [81]

Elimination des duplicats de PCR

La technique de préparation de la librairie peut générer des fragments, hors amplicons, dont les extrémités seraient aléatoires mais de coordonnées et de score CIGAR strictement identiques aux reads d'une paire. On parle de duplicat de PCR.

Il faut les éliminer car leur présence impacterait la détection et la quantité de variants réels. Des outils tels que Picard, Samtools ou GATK peuvent réaliser cette étape.

Réalignement autour des Indels

Des sites génomiques sont concernés par des variations type Indel : indel connues dans les bases de données, indel vues dans les alignements originaux et sites suggérant un indel caché. Il faut donc réaliser un nouvel alignement local autour de ces sites pour éliminer le maximum de faux positifs. Cette étape consiste donc à détecter les sites qui nécessitent un réalignement local. Il y a alors recherche du meilleur alignement pour les sites ciblés en prenant en compte les différentes lectures s'alignant à ces positions. Le réalignement peut être réalisé par GATK.

Variant Calling

Étape d'identification et listing des variants détectés, pouvant être réalisées par des outils comme GATK ou Samtools. Le format de fichier obtenu est le VCF (Variant Call Format). Il possède au minimum huit champs fixes : le chromosome, la position du variant, l'identifiant (numéro Rs, identifiant ClinVar ...), la base de référence, la base alternative, la qualité moyenne des bases alternatives et les critères de passage du filtre informatique qui est prédéfini.

Des champs optionnels permettent d'ajouter des données de filtration selon des critères choisis. Les résultats sont visualisés avec un tableur de type Excel, ce qui permet une exploitation simple et intuitive des données.

Pour la détection des CNVs, l'étape de variant calling est différente car les variants callers de CNVs sont basés sur des stratégies d'investigations des profondeurs de séquençage : à qualité égale de séquence et d'alignement, la profondeur de séquençage permet d'évoquer des variations de nombre de copies (FIGURE 22).

Des outils comme XHMM ou exomedepth permettent la détection de CNVs en passant par des étapes mathématiques de normalisation permettant de retrouver des écarts statistiquement représentatifs de variation de profondeur de séquençage, et donc de délétin ou duplication de région.

La FIGURE 22 ci-après, représente les étapes du pipeline de détection des SNVs.

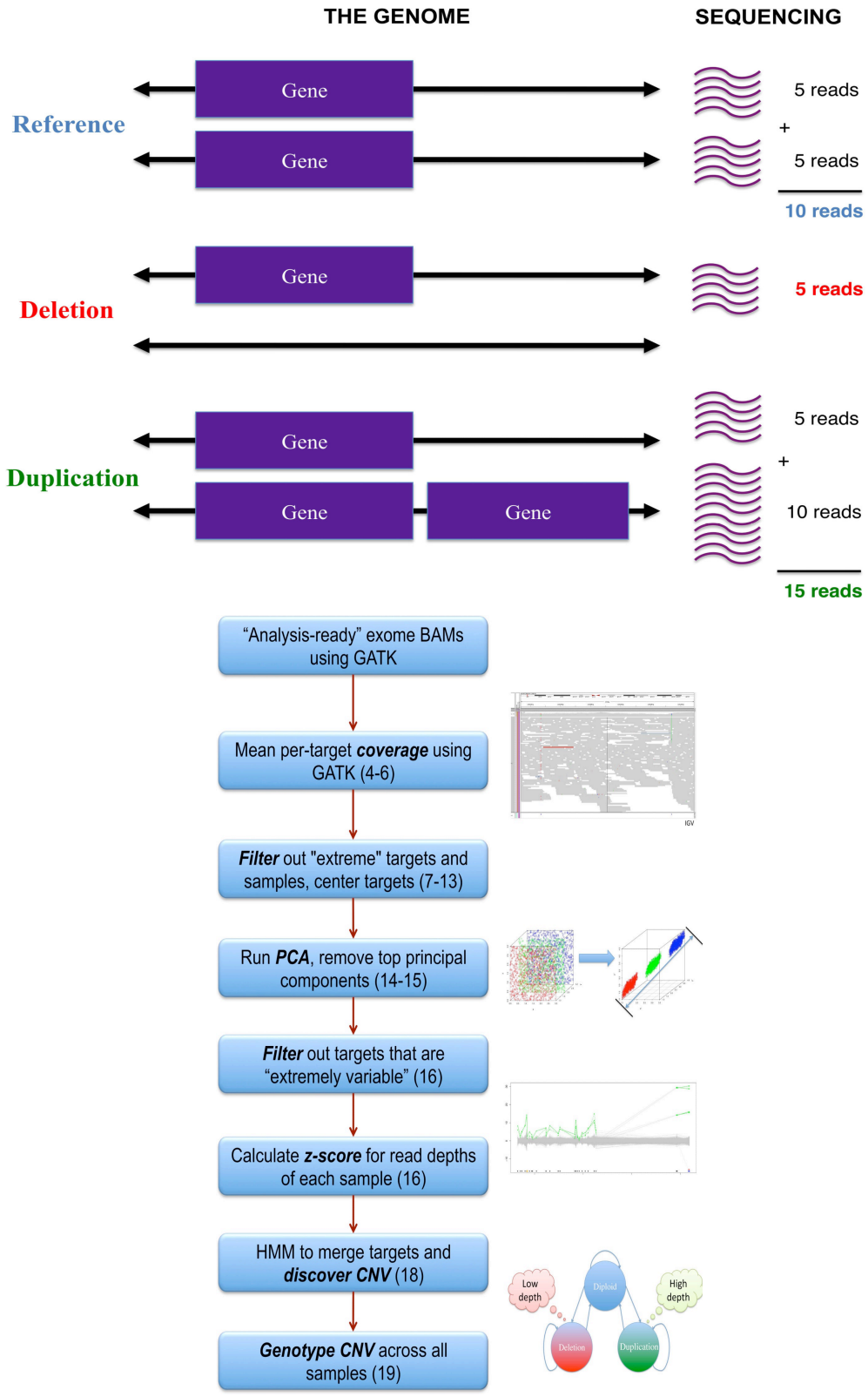


FIGURE 22 – *Principe de détection et succession de tâches de calculs effectués par XHMM pour la détection des CNVs à partir des données de NGS [8]*

✓ Analyse tertiaire

Correspond à l'interprétation des résultats. Elle est basée sur l'annotation des variants et l'interprétation biologique des résultats. L'annotation est une donnée descriptive du variant en termes de nature (missense, frameshift etc..) ou d'impact (pathogène, bénin etc…).

L'annotation du fichier VCF est réalisée de façon systématique en fin de processus d'analyse des données. Les variants du fichier VCF sont annotés grâce à l'interrogation de différentes bases de données ou encore des résultats des outils de prédiction.

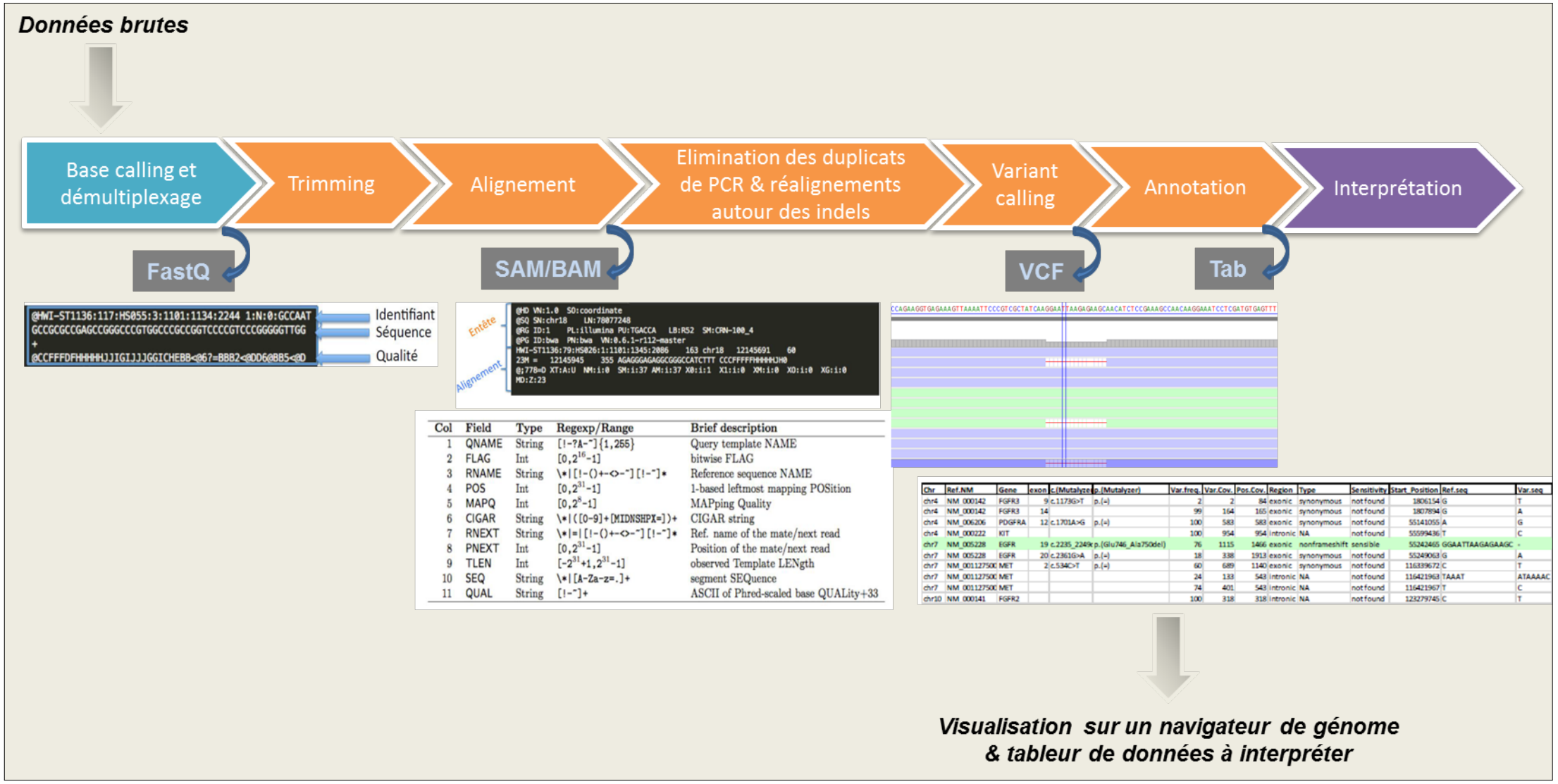
Plusieurs outils sont capables de réaliser l'annotation, tel que Seattle Seq Annotation ou SnpEff. Tandis que Annovar ou VEP (Variant Effect Predictor) permettent de réaliser des prédictions in silico sur plusieurs outils simultanément.

Pour l'annotation des CNVs, on utilise communément cnvscan ou annotSV. Les bases de données phénotypiques classiques sont OMIM et Decipher.

L'ensemble de ces données d'annotation permet de filtrer les variants pour répondre à une question biologique. La sélection des variants d'intérêt se fait tout d'abord sur des critères analytiques. Afin d'éviter les faux positifs on peut :

- Visualiser le fichier BAM sur des outils comme IGV pour éviter des erreurs, telles que des variants retrouvés uniquement sur un seul sens de lecture, des éventuelles amorces non trimées ou des faux positifs liés au pipeline d'analyse.
- Confronter les résultats à des bases de données locales de variants pour mettre en évidence des variants récurrents correspondants à des artefacts. Une faible fréquence allélique peut correspondre à un artefact ou une mosaïque.
- Utiliser des filtres basés sur la profondeur pour exclure les variants avec de faibles profondeurs qui pourraient correspondre à des artefacts

Le taux d'erreur du NGS est de l'ordre de 0,1 à 1%.



Col	Field	Type	Regexp/Range	Brief description
1	QNAME	String	[!-?A-~]{1,255}	Query template NAME
2	FLAG	Int	[0,2^{16}-1]	bitwise FLAG
3	RNAME	String	*|[!-()+-<>-~][!-~]*	Reference sequence NAME
4	POS	Int	[0,2^{31}-1]	1-based leftmost mapping POSition
5	MAPQ	Int	[0,2^{8}-1]	MAPping Quality
6	CIGAR	String	*|([0-9]+[MIDNSHPX=])+	CIGAR string
7	RNEXT	String	*|=|[!-()+-<>-~][!-~]*	Ref. name of the mate/next read
8	PNEXT	Int	[0,2^{31}-1]	Position of the mate/next read
9	TLEN	Int	[-2^{31}+1,2^{31}-1]	observed Template LENgth
10	SEQ	String	*|[A-Za-z=.]+	segment SEQuence
11	QUAL	String	[!-~]+	ASCII of Phred-scaled base QUALity+33

Chr	Ref.NM	Gene	exon	c.(Mutalyzer)	p.(Mutalyzer)	Var.freq.	Var.Cov.	Pos.Cov.	Region	Type	Sensitivity	Start_Position	Ref.seq	Var.seq
chr4	NM_000142	FGFR3	9	c.1173G>T	p.(=)	2	2	84	exonic	synonymous	not found	1806154	G	T
chr4	NM_000142	FGFR3	14			99	164	165	exonic	synonymous	not found	1807894	G	A
chr4	NM_006206	PDGFRA	12	c.1701A>G	p.(=)	100	583	583	exonic	synonymous	not found	55141055	A	G
chr4	NM_000222	KIT				100	954	954	intronic	NA	not found	55599436	T	C
chr7	NM_005228	EGFR	19	c.2235_2249d	p.(Glu746_Ala750del)	76	1115	1466	exonic	nonframeshift	sensible	55242465	GGAATTAAGAGAAGC	-
chr7	NM_005228	EGFR	20	c.2361G>A	p.(=)	18	338	1913	exonic	synonymous	not found	55249063	G	A
chr7	NM_001127500	MET	2	c.534C>T	p.(=)	60	689	1140	exonic	synonymous	not found	116339672	C	T
chr7	NM_001127500	MET				24	133	543	intronic	NA	not found	116421963	TAAAT	ATAAAAC
chr7	NM_001127500	MET				74	401	543	intronic	NA	not found	116421967	T	C
chr10	NM_000141	FGFR2				100	318	318	intronic	NA	not found	123279745	C	T

FIGURE 23 – *Schématisation du procédé bioinformatique de détection des SNVs*

Les couleurs des boites représentent les niveaux d'analyse.

Les boites à fonds noir indiquent le format de fichier; en dessous, des illustrations photos des étapes correspondantes.

3. Rappels sur l'interprétation de variants

L'interprétation est l'étape qui permettra de distinguer un variant pathogène d'une variation de séquence non pathogène ou polymorphisme. Elle est basée sur l'appréciation critique du biologiste ou du chercheur des éléments de filtration. C'est l'étape « humaine » de l'analyse NGS. C'est la raison pour laquelle, elle conduit parfois à des divergences de classification de variants.

La première étape de l'interprétation consiste à déterminer si le variant est référencé dans les bases de données ou décrit dans la littérature. S'il n'est pas connu, on s'intéresse à son impact grâce aux différents outils in sillico. Des algorithmes de classification proposés par l'American College of Medical Genetics (ACMG) permettent au final de déduire la pathogénicité du variant.

3.1. Bases de données

3.1.1. Bases de données de fréquence

Suite au décryptage du génome humain, de nombreux projets basés sur le séquençage massif en parallèle ont vu le jour. On note par exemple le projet « 1000 genomes », débuté en 2008, achevé en 2010, il avait pour but de répertorier les variations génomiques de 1000 individus anonymes, quel que soit leur santé ou ethnie.

Ce type de projet a permis de séquencer des cohortes d'individus sains ou malades. Les résultats de ces travaux de séquençage massif sont répertoriés dans des bases de données publiques. Ces différentes bases offrent une estimation de la fréquence des variants dans une population donnée. On considère qu'un variant est un polymorphisme lorsque sa fréquence dans la population générale est supérieure à 1%. Mais ces bases ne contiennent pas que des variants bénins et les cohortes peuvent contenir des individus présentant diverses pathologies de l'adulte, seuls les phénotypes pédiatriques sévères peuvent être discriminés.

Il faut donc être prudent pour ne pas éliminer des variants pathogènes qui seraient considérés à tort comme bénins (faux négatifs). On peut citer les bases suivantes :

ESP, Exome Sequencing Project,

Comprend actuellement les exomes de 6502 individus.

ExAC, Exome Aggregation Consortium,

Regroupe les exomes de 60 706 individus. Soit 14 cohortes de patients et 6 groupes de populations.

gnomAD, Genome Aggregation Database,

Correspond à une évolution d'ExAC et compile actuellement 123136 exomes et 15946 génomes. Cette base regroupe les données d'ExAC et d'autres projets de séquençage tels que 1000 Genomes et HapMap.

dbSNP, Single Nucleotide Polymorphism Database

Entretenu par le centre de ressource américain NCBI (National Center for Biotechnology Information), c'est à l'origine une base de polymorphismes mais aujourd'hui elle référence également les variations selon leur degré de pathogénicité.

Il existe des bases de données spécifiques de pathologies, issues de réseaux nationaux de recherche. L'avantage de telles bases est leur fiabilité grâce à la validation de variants par des groupes d'expert. Cependant elles comportent des cohortes plus restreintes que les bases internationales et sont donc moins exhaustives.

ClinVar compile des soumissions de variants réalisées directement par les laboratoires ou par des groupes d'experts.

3.1.2. Bases de données de la littérature

La revue de la littérature peut aider à caractériser des variants. Mais cette recherche est parfois longue et fastidieuse. Il existe des bases spécifiques comme :

HGMD (Human Gene Mutation Database)

Consultée à l'adresse : http://www.hgmd.cf.ac.uk/ac/index.php ; extrait et compile les données de la littérature ;

OMIM (Online Mendelian Inheritance in Man)

Consultée à l'adresse : https://www.omim.org/ ; Recense plus de 3800 maladies secondaires à des variants pathogènes dans plus de 3400 gènes. Les données d'OMIM sont issues de la littérature et mises à jour régulièrement. Mais cette base est loin d'être exhaustive.

Decipher, (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources) [82]

Associe surtout CNVs et pathologies humaines.

3.2. Logiciels de prédiction

Lorsque le variant est absent des différentes bases de données ou que son implication dans la pathologie n'est pas clairement définie, on dispose d'outils bioinformatiques permettant de prédire l'impact d'un variant.

Globalement, il y a 2 catégories d'algorithmes prédictifs :

- Ceux qui prédisent l'effet d'un variant missense sur la structure et la fonction protéique ;
- Ceux qui prédisent l'effet d'un variant sur l'épissage.

3.2.1. Algorithmes de prédiction de variant missense

Une variation missense (faux-sens en français) est une altération ponctuelle d'un nucléotide qui aboutit à un changement d'acide aminé lors de la traduction. Le changement d'acide aminé peut avoir ou non des conséquences sur la fonction de la protéine. Plusieurs arguments sont à l'origine des algorithmes prédictifs :

- Un acide aminé (AA) conservé au cours de l'évolution a de grandes chances d'appartenir à un élément fonctionnel jouant un rôle important dans la fonction protéique ;
- Un variant est d'autant plus délétère qu'il transforme l'AA en un autre de propriétés physico-chimiques différentes, impactant la polarité et l'encombrement stérique ;
- Il existe des scores de probabilités de substitution d'AA ;
- Il existe des blocs de nucléotides conservés présentant une certaine intolérance aux changements ;
- Un variant est délétère s'il impacte un AA impliqué dans le maintien de la structure protéique ;

Les outils cités ci-après ne sont pas exhaustifs mais sont ceux qui ont été utilisés dans l'étude.

SIFT (Sorting Intolerant From Tolerant) [83]

Pour prédire si la substitution d'un acide aminé affecte la fonction protéique, SIFT considère la position à laquelle la variation s'est produite et le type d'acide aminé impliqué. Un alignement multiple permet de visualiser la conservation des résidus entre les espèces. Les changements dans des positions bien conservées ont tendance à être prédits délétères. De même, sur la base de comparaison avec des familles de protéine, SIFT calcule la probabilité qu'un AA soit toléré à une position donnée. Enfin, les changements affectant les catégories physico-chimiques des AA ont tendance à être prédits comme moins tolérées.

Le score de Grantham

Le score de Grantham tente de prédire la différence entre deux acides aminés. Les scores de distance vont de 5 à 215. Un faible score reflète une faible distance évolutive. Un score élevé reflète une plus grande distance évolutive. Les scores élevés de Grantham sont considérés comme plus délétères.

Le score de Grantham est basé sur une matrice de transition entre les 20 acides aminés connus. Elle tient compte des propriétés physico-chimiques et du volume moléculaire de chaque acide aminé. Plus ce score est petit plus les propriétés physico-chimiques de la protéine sont respectées.

PhastCons, PhyloP, PROVEAN (PROtein Variation Effect Analyzer) [84]**, ou GERP (Genomic Evolutionary Rate Profiling)** [85]

Sont d'autres outils de prédiction qui se basent sur la conservation de blocs nucléotidiques, par l'utilisation d'une matrice d'intolérance aux changements de résidus.

FATHMM, Functional Analysis through Hidden Markov Models [86]

Capable de prédire les effets fonctionnels des variants faux-sens en combinant la conservation de séquence dans des modèles de Markov (HMM). Ces modèles attribuent des "poids de pathogénicité" en fonction de la conservation des séquences homologues et des domaines protéiques, ce qui traduit la tolérance globale aux variants pathogènes d'une protéine ou d'un domaine protéique.

LRT, Likelihood Ratio Test

Le test du rapport de vraisemblance (LRT) peut identifier avec précision un sous-ensemble de variants délétères perturbant les acides aminés hautement conservés dans les séquences codant pour les protéines, qui sont susceptibles d'être inconditionnellement délétères.

Polyphen [87]

Polyphen (Polymorphism Phenotyping) utilise à la fois des données d'alignements multiples, des annotations connues de séquences et des informations structurales provenant de banque de données de structure (PDB). Il donne la possibilité de prédire si la substitution engendre une déformation probable de la structure de la protéine par utilisation d'un algorithme combinant HBPLUS, qui prédit la formation de ponts hydrogènes entre chaînes latérales et entre chaines latérales et chaine principale ; et l'outil Dist, qui calcule les contacts spatiaux d'un résidu.

MutationTaster [88]**, CADD** [89]

Sont des outils de compilation d'algorithmes. Ils donnent également un score de prédiction. MutationTaster combine les résultats de PhasCons, PhyloP, NNsplice et du score de Grantham. CADD (Combined Annotation Dependant Depletion) combine de multiples données sur la conservation et la tolérance de changement de résidus pour calculer le C-score.

3.2.2. Algorithmes de prédiction des effets sur l'épissage

L'épissage est le mécanisme moléculaire par lequel le spliceosome ôte les introns. Le spliceosome reconnait le site d'épissage à travers un appariement de bases. Les variants ayant des conséquences sur l'épissage se situent principalement sur les sites canoniques d'épissage.

Les logiciels de prédiction calculent un score qui renseigne sur la force du site d'épissage et il y a ensuite comparaison des scores sauvages et de la séquence mutée pour prédire l'effet. On peut citer :

MaxENTScan

Ce logiciel se base sur une approche de modélisation de motifs de courtes séquences impliquées dans l'épissage de l'ARN. Il prend en compte simultanément les dépendances non adjacentes et adjacentes entre les positions. Cette méthode est basée sur le « principe d'entropie maximum » et généralise la plupart des modèles probabilistes de motifs de séquences tels que les modèles de matrice de poids et les modèles de Markov inhomogènes. L'outil est online à l'adresse : http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html

Human Splicing Finder

Se base sur des matrices de prédiction pour les sites de liaison des protéines d'épissage. Il combine 12 algorithmes différents pour identifier et prédire l'effet des variants de motifs d'épissage, y compris les sites d'épissage donneur et accepteur, le point de branchement et les séquences auxiliaires connus pour renforcer ou réprimer l'épissage : Exonic Splicing Enhancers (ESE) et les Exonic Splicing Silencers (ESS). L'outil est online à l'adresse : http://www.umd.be/HSF.

Les performances de ces outils se révèlent assez efficaces lorsque le variant est proche des sites consensus mais les performances s'amenuisent lorsque le variant s'éloigne de ces derniers.

3.3. Étude de la ségrégation familiale

L'étude de la ségrégation d'un variant est une aide à l'interprétation. Lorsque la maladie répond à un mode transmission mendélien, l'étude de la ségrégation familiale est informative :

- Quand le variant n'est retrouvé chez aucun des deux parents sains, il est dit *de novo*, ce qui constitue un argument supplémentaire en faveur de l'implication dans la pathologie du cas index.
- Quand le variant est hérité, il faudra vérifier qu'il ségrége avec la maladie dans la famille en testant d'autres apparentés atteints.

3.4. Classification des variants

A l'échelle d'un génome entier, on retrouve plus de trois millions de variations. Cependant, la grande majorité de ces variations n'aura pas d'effet délétère. Il faudra donc identifier parmi l'ensemble de ces variations celles qui sont associées au phénotype du patient.

L'ACMG propose une méthode de classification des variants (FIGURE 24 ci-après) en pondérant les principaux critères utilisés pour aboutir à la classification en cinq niveaux ou classes :

- Classe 1, bénin
- Classe 2, probablement bénin
- Classe 3, Variant de Signification Incertaine (VSI ou Variant of Unknown Signification, VUS)
- Classe 4, probablement pathogène
- Classe 5, pathogène

Comme le montre la FIGURE 24, ils sont de classe 5 s'ils prédisent un STOP dans un gène intolérant à la perte de fonction et impliqué dans un mécanisme de pathologie humaine. Ils sont de :

- Classe 4 ou 5 selon l'association de plusieurs arguments forts parmi lesquels : de novo ; études fonctionnelles en faveur d'un effet délétère ; un variant dans l'acide aminé concerné est prédit pathogène ; la prévalence dans les populations malades sont statistiquement plus importantes que dans les populations saines.
- Classe 3, les plus fréquents : de signification incertaine. Ce sont les variants pour lesquels l'analyse ne permet pas de confirmer une classe 4 ou 5 ; Ou les variants pour lesquels les critères de classification en bénin ou pathogène sont contradictoires.

Il existe aussi des « sous-classes » : très pathogène, pathogène, modérément pathogène, probablement pathogène ; bénin ou probablement bénin. Elle participe du faisceau d'arguments à employer pour déterminer la pathogénicité d'un variant. L'ACGM propose un algorithme de calcul et d'évaluation de la pathogénicité à partir de ces sous-classes (FIGURE 25).

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder *BA1/BS1* **OR** observation in controls inconsistent with disease penetrance *BS2*			Absent in population databases *PM2*	Prevalence in affecteds statistically increased over controls *PS4*	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact *BP4* Missense when only truncating cause disease *BP1* Silent variant with non predicted splice impact *BP7* In-frame indels in repeat w/out known function *BP3*	Multiple lines of computational evidence support a deleterious effect on the gene /gene product *PP3*	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before *PM5* Protein length changing variant *PM4*	Same amino acid change as an established pathogenic variant *PS1*	Predicted null variant in a gene where LOF is a known mechanism of disease *PVS1*
Functional Data	Well-established functional studies show no deleterious effect *BS3*		Missense in gene with low rate of benign missense variants and path. missenses common *PP2*	Mutational hot spot or well-studied functional domain without benign variation *PM1*	Well-established functional studies show a deleterious effect *PS3*	
Segregation Data	Non-segregation with disease *BS4*		N ≤ 1/8 if 1 family N ≤ 1/4 if > 1 family	N ≤ 1/16 if 1 family N ≤ 1/8 if > 1 family	N ≤ 1/32 if 1 family N ≤ 1/16 if > 1 family	
De novo Data				*De novo* (without paternity & maternity confirmed) *PM6*	*De novo* (paternity & maternity confirmed) *PS2*	
Allelic Data		Observed in *trans* with a dominant variant *BP2* Observed in *cis* with a pathogenic variant *BP2*		For recessive disorders, detected in *trans* with a pathogenic variant *PM3*		
Other Database		Reputable source w/out shared data = benign *BP6*	Reputable source = pathogenic *PP5*			
Other Data		Found in case with an alternate cause *BP5*	Patient's phenotype or FH highly specific for gene *PP4*			

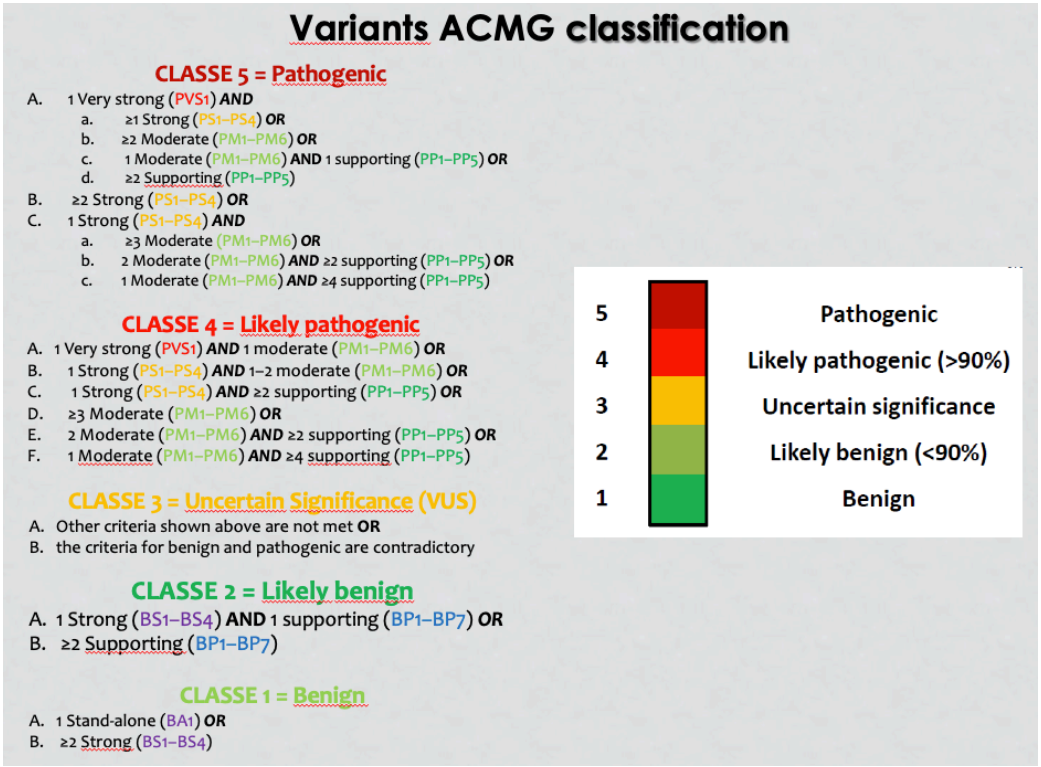


FIGURE 24 – Tableaux de classification des variants selon les recommandations de l'ACMG [90]

Pathogenicity: *PVS, Pathogenic Very Strong or PS, Pathogenic Strong, PM, Pathogenic Moderate, PP, Pathogenic Poor*
Benign character: *BA, Benign stand Alone or BS, Benign Strong, BP, Benign Poor*

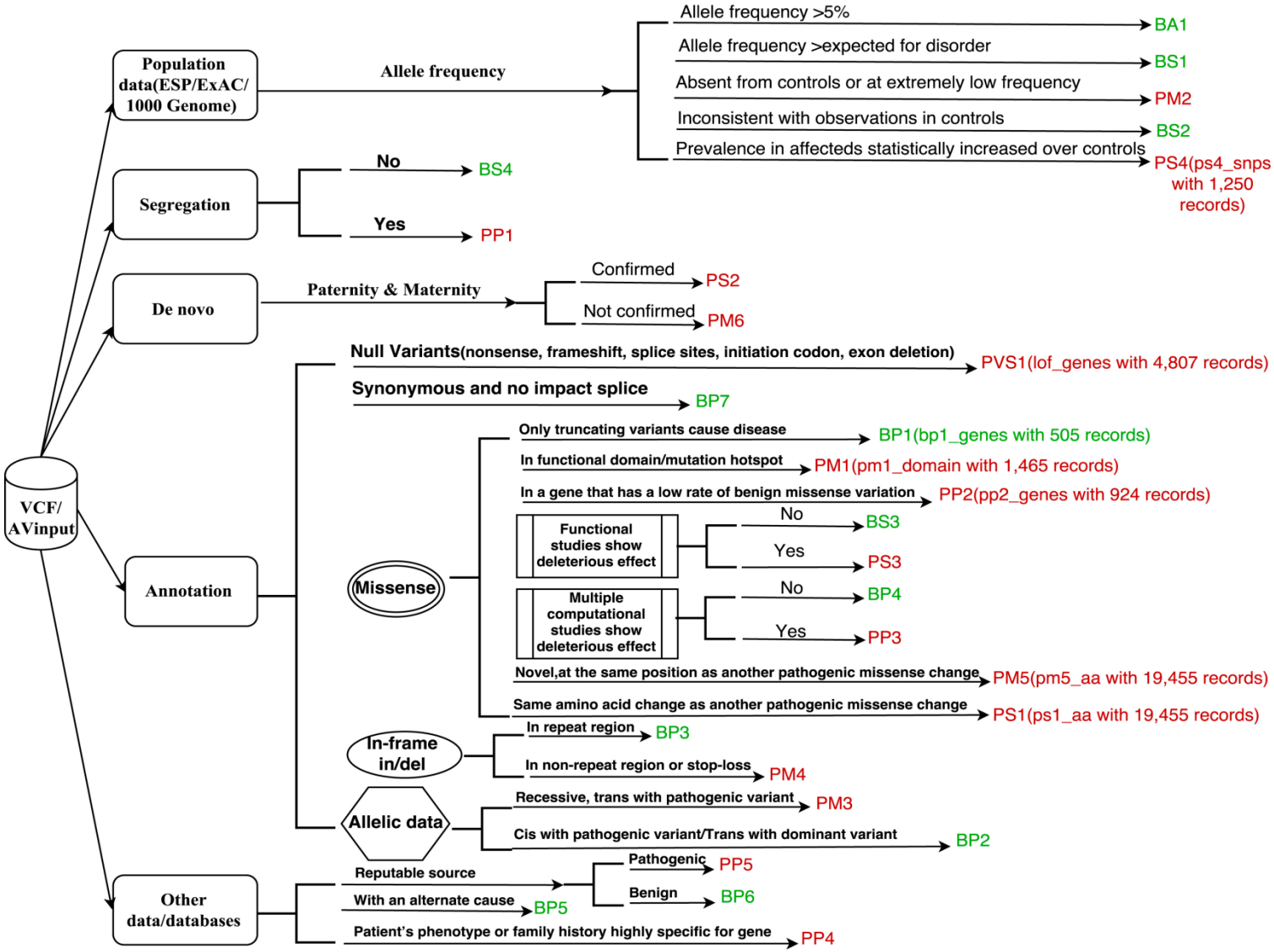


FIGURE 25 – Algorithme de classification ACMG [90]

4. Détection de variants SNVs et CNVs par CoDE-seq

Les expériences ont été sous-traitées à la plateforme GENOSCREEN-LIGAN à Lille. La méthode CoDE-Seq, Copy number variation Detection and Exome Sequencing, mise en oeuvre est un exome complet (WES) augmenté de sondes permettant la détection de CNVs associés à des pathologies humaines.

4.1. Technique de séquençage

4.1.1. Extraction d'ADN

L'extraction est réalisée au laboratoire du centre de référence par une technique manuelle basée sur la purification de l'ADN par le NaCL. Des quantités conséquentes d'ADN sont ainsi extraites à partir de 7 à 8 mL de sang veineux. En moyenne, on obtient des concentrations de 400-500 ng/µL.

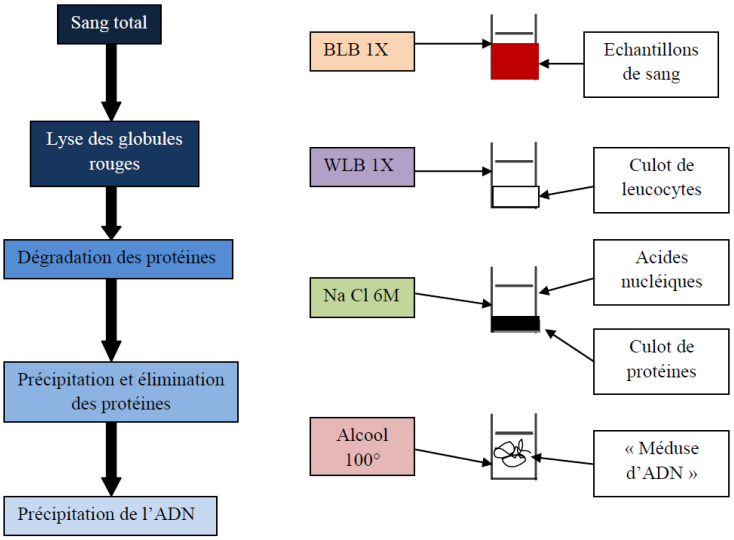


FIGURE 26, *schéma de la méthode d'extraction aux sels employée au laboratoire du centre de référence*
BLB : Blood Lysis Buffer ; WLB : White Lysis Buffer

L'ADN est conservé à -80°C jusqu'à acheminement dans les laboratoires référents de diagnostic ou sur la plate-forme de NGS à laquelle nous avons eu recours pour ces travaux.

4.1.2. Design de la capture

L'ensemble de la capture de l'exome a été effectué avec l'enrichissement NimbleGenSeqCap EZ combiné à la conception maison de NimbleGenSeqCap EZ Choice XL : des sondes d'environ 120pb sont régulièrement espacées dans tout le génome, c'est le squelette de répartition des sondes. Elles sont placées spécifiquement tous les 10 Kb (10.10^3pb) dans les régions génomiques connues pour être affectées par les CNVs.

Ces sondes englobent 881 Mb, soit environ 30% du génome. Pour les 70% restants, des sondes ont été placées tous les 25 Kb. Les sondes ciblant l'exome et le squelette entiers couvraient 65 Mb du génome.

La base de référence utilisée pour le choix de ces régions est DECIPHER.

4.1.3. Capture

L'enrichissement, de type captyure en phase solide, a été effectué avec le KIT NimbleGen SeqCap EZ MedExome, conformément au protocole du fournisseur, Roche diagnosis, pour le séquençage type Illumina.

Un microgramme d'ADN a été fragmenté par sonication (Covaris E220 Focused-ultrasonicator). Les extrémités des fragments d'ADN ont été réparées et liées aux adaptateurs Illumina en utilisant les kits de préparation de librairies « KAPA http », sur l'automate Hamilton Microlab STARlet.

Les échantillons ont ensuite été amplifiés par PCR (Polymérase Chain Reaction) en utilisant des amorces complémentaires des adaptateurs.

Après quantification et sélection de taille des échantillons (Perkin Elmer LabChip GX), ils ont été poolés par cinq. On obtient des pools d'au moins 1 µg qui sont hybridés au mélange de sondes marquées à la biotine du KIT SeqCap EZ. Après 72h

à 47°C, les captures ont été purifiées en utilisant le kit d'hybridation et de lavage SeqCap sur la plate-forme automatisée de manipulation de liquides Agilent Bravo.

Les captures ont ensuite été amplifiées à l'aide du KAPA HiFi HotStart ReadyMix et quantifiées par les dosages de quantification fluorométrique Caliper LifeScience LabChip GX et Qubit (Thermo Fisher Scientific).

4.1.4. Séquençage

Les échantillons ont été séquencés sur le système Illumina HiSeq 4000 avec un débit d'un pool comprenant cinq échantillons par voie et en utilisant un protocole paired-end 2 x 150 pb. Le séquençage est de type enzyme terminator.

4.2. Dispositif informatique

4.2.1. Hardware

Le démultiplexage a été assuré par la plate-forme de séquençage, en utilisant bcl2fastq Conversion Software (Illumina ; version 2.19.1). Les données brutes au format FASTQ, ont été sauvegardées sur un disque dur externe. A BIOSPHERES, par un protocole SSH, les analyses bioinformatiques ont été menées sur un processeur 32 virtuels de 90G de RAM dont 45 pour un disque de sauvegarde, avec un système d'exploitation Centos 7.5 64b.

4.2.2. Pipeline

Il existe des outils online ou payant permettant d'automatiser les pipelines, comme « Galaxy ». Pour notre étude, nous avons élaboré un script d'automatisation du pipeline. Avec le soutien des universités de Bourgogne et de Lyon 1, nous avons analysé les données brutes avec un pipeline qui respecte les étapes ci-après. Les commandes utilisées sont en annexe 5 de ce chapitre.

✓ Détection des SNVs

Dans un 1e temps on procède à l'initialisation des variables (Commande 1)

Contrôle Qualité des FASTQ

FastQC vise à fournir un rapport de qualité (QC) à partir des fichiers FASTQ qui peut repérer les problèmes qui proviennent du séquenceur ou du matériel de la bibliothèque de départ. (Commande 2)

Nettoyage des reads

Trimmomatic effectue une variété de tâches de découpage utiles pour les données illumina paired-end et single-end. La sélection des étapes de rognage et leurs paramètres associés sont fournis sur la ligne de commande. Il fonctionne avec FASTQ (en utilisant des scores de qualité phred + 33 ou phred + 64, selon le pipeline Illumina utilisé). Pour les données appariées, deux fichiers d'entrée sont spécifiés, et 4 fichiers de sortie : 2 « appariés » correspondants au traitement des deux lectures et 2 « non appariés » correspondants au traitement de l'une ou l'autre lecture uniquement. (Commande 3)

Les paramètres usuels de « trimming » sont :

- ILLUMINACLIP : Elimine l'adaptateur et les autres séquences spécifiques à Illumina (ILLUMINACLIP : TruSeq3-PE. fa: 2: 30: 10)
- SLIDINGWINDOW : Scanne la lecture avec une large fenêtre coulissante à 4 bases, coupant lorsque la qualité moyenne par base tombe en dessous de 15 (SLIDINGWINDOW : 4 :15)
- LEADING : Elimine les bases en début de lecture quand elles sont de qualité moyenne inférieure à 3 (LEADING : 3)
- TRAILING : Elimine les bases en fin de lecture quand elles sont de qualité moyenne inférieure à 3 (TRAILING : 3)
- CROP : Coupe la lecture à une certaine longueur de séquence
- HEADCROP : Elimine un nombre spécifié de bases en début de lecture
- MINLEN : Supprime la lecture si elle est inférieure à une longueur de 36 bases (MINLEN : 36)
- TOPHRED33 : Convertit les scores de qualité en Phred-33
- TOPHRED64 : Convertit les scores de qualité en Phred-64

Alignement

BWA, Burrows-Wheeler Aligner (version 0.7.15-r1140) a été utilisé pour mappées les FASTQ sur le génome humain de référence version hg20/GRCh38. BWA est un logiciel permettant de cartographier des séquences faiblement divergentes par rapport à un grand génome de référence, tel que le génome humain. Il se compose de trois algorithmes : BWA-backtrack, BWA-SW et BWA-MEM.

Le premier algorithme est conçu pour une séquence Illumina lit jusqu'à 100 pb, tandis que les deux autres pour des séquences plus longues allant de 70 pb à 1 Mbp. BWA-MEM et BWA-SW partagent des fonctionnalités similaires telles que la prise en charge de la lecture longue et l'alignement fractionné, mais BWA-MEM, qui est le dernier et celui que nous avons utilisé, est généralement recommandé pour les requêtes de haute qualité car il est plus rapide et plus précis. (Commande 4)

Elimination des duplicats de PCR et réalignement autour des Indels (Commande 5)

SAMTOOLS fournis divers utilitaires pour manipuler les alignements au format SAM, y compris le tri, la fusion, l'indexation et la génération d'alignements dans un format par position.

Variant calling

Nous avons utilisé 2 variant-callers :

GATK version 3.8-0-ge9d806836, Genome Analysis tool kit, développé par le Broad Institute fournit un ensemble d'outils pour analyser les données de NGS. (Commande 6)

FREEBAYES, détecte des variants génétiques sur la base d'un algorithme bayésien centré sur l'analyse des haplotypes : il appelle les variants selon l'alignement de séquences littérales, et non sur leur alignement précis. Cette méthode évite l'un des problèmes fondamentaux de la détection de variants basée sur l'alignement, à savoir que des séquences identiques peuvent avoir plusieurs alignements possibles. Freebayes utilise des alignements à lecture courte pour n'importe quel nombre d'individus d'une population et un génome de référence afin de déterminer la combinaison de génotypes la plus probable pour la population à

chaque position dans la référence. Il rapporte les positions qu'il trouve prétendument polymorphes au format VCF. (Commande 7)

Nous avons ensuite rassemblé les résultats dans un seul fichier. (Commande 8)

Annotation

snpEff version 4.3t 2017-11-24 a été utilisé pour l'annotation des variants avec différentes bases de données de fréquences et de prédiction in sillico. (Commande 9).

Pour annoter les gènes d'intérêt, nous avons nous-mêmes effectué une recherche par mots-clés sur la base de données Gene de NCBI, pour lister les gènes indexés aux pathologies NDs et en particulier à la SLA que nous avons étudiée de façon spécifique par rapport à une situation atypique observée au cours de l'étude. Le tableau correspondant à la liste des gènes en question est en annexe (gene_result_ALS_01072019.txt et gene_result_ND_01072019.txt). (Commande 10)

✓ Détection des CNVs

Méthode de détection

La détection des CNV a été effectuée à l'aide de 3 méthodes :

XHMM (eXome Hidden Markov Model) version 1.0, est un programme de détection des CNVs à partir des données d'Exome [91]. Ici aussi, on débute par l'initialisation des variables. (Commande 11). Pour utiliser XHMM, on passe par des « metrics » de GATK, en particulier les valeurs de profondeur de séquençage (Commande 12). Les données obtenues par XHMM ont été croisées avec les panels de gènes NDs/SLA, que nous avions défini afin de ne cibler que les régions contenant des gènes d'intérêt.

Exomedepth v1.1.15, utilise un modèle robuste de comptage de reads pour créer un ensemble de référence optimisé et ainsi maximiser la puissance de détection des CNV. Un script R a été monté pour cette étape. (Commande 13)

CoNIFER utilise les données de séquençage de l'exome pour trouver des variants de nombre de copies et génotyper le nombre de copies des gènes dupliqués. Comme les réactions de capture d'exomes sont sujettes à des biais de capture forts et systématiques entre les lots d'échantillons, nous avons mis en œuvre une décomposition en valeur singulière (SVD) pour éliminer ces biais dans les données d'exomes. Associé à un aligneur de lecture courte tel que mrsFAST qui peut aligner les lectures sur plusieurs emplacements, CoNIFER peut détecter de manière robuste les CNV rares et estimer le nombre de copies de gènes dupliqués jusqu'à ~ 8 copies avec les kits de capture d'exomes actuels [92] (Commande 14)

Enfin, les variants détectés sont rassemblés dans un seul fichier (Commande 15).

Filtration et annotation

cnvscan v0.2.0 a été utilisé pour filtrer les résultats obtenus par les outils de détection des CNVs. Ce logiciel a été développé pour améliorer la prédiction informatique des CNVs dans le cadre d'une utilisation clinique en réduisant le nombre de faux positifs et en fournissant des annotations. cnvScan se compose de deux étapes: la détection et l'annotation des CNVs. [93] Nous avons utilisé l'outil pour chacun des résultats des méthodes de détection. (Commande 16)

Dans le but de n'obtenir qu'un fichier unique, nous avons rassemblé les données. (Commande 17)

AnnotSV v2.2 est un programme autonome conçu pour annoter et classer les variations structurelles. Il compile des informations fonctionnellement et cliniquement pertinentes, et vise à fournir des annotations utiles pour interpréter la pathogénicité potentielle des variations structurelles et filtrer les faux positifs potentiels [94]. Les bases de données utilisées pour l'annotation sont, OMIM et Decipher. (Commande 18)

✓ **Détection des Tandem Repeats**

La detection des motifs répétés est possible par l'utilisation d'outils bioinformatique adaptés. En 2020, Halman et al. publient leur comparaison d'outils en la matière. Ils concluent que tous les outils ont leurs forces et faiblessses mais qu'un bon usage et une bonne sélection des paramètres sont indispensables à l'appel effectif de variants.[95]

Dans notre étude, nous avons opté pour l'outil **GangSTR** pour la détection de ce type de variants par rapport à sa grande capacité d'appel de variants [96].

✓ **Détection de variants communs**

geneFilter version 1.6.0. a été utilisé pour filtrer spécifiquement les variants qui seraient communs à plusieurs échantillons. Nous nous en sommes servi dans l'étude des patients atteints de SLA avec variant pathogène *SOD1* appartenant à une famille où la maladie ségrège avec une anticipation, dans le but de recherche un oligogénisme [97].

4.2.3. Interprétation

Pour les arguments de fréquence, nous avons utilisé la base de données GnomAD. Nous avons annoté les données de fréquence de populations africaine et caucasienne.

Pour les arguments de pathogénicité, nous avons utilisé les algorithmes in sillico suivants : PROVEAN, PolyPhen-2, SIFT, MutationTaster, MutationAssessor, LRT, CADD, FATHMM, GERP, PhyloP (100 & 20), and PhastCons (100 & 20).

Le TABLEAU 11 ci-après, récapitule les outils de prédiction utilisés dans notre étude des MNDREM, et les scores de prédiction associés au statut probablement pathogène à pathogène.

L'interprétation des CNVs non répertoriés dans Decipher est essentiellement basée sur les données de la littérature.

OUTILS	Méthodes de prédiction	Score pathogène	Site web	Références
PROVEAN	Calcul d'un score d'alignement des changements de résidus et conservation	-1 à > 2.5	http://provean.jcvi.org/index.php	https://doi.org/10.1371/journal.pone.0046688
SIFT	Conservation inter-espèce	<0.95	http://sift.jvci.org	https://doi.org/10.1038/nprot.2009.86
Polyphen2	Conservation structurale de la protéine et conservation inter-espèce	>0.85	http://genetics.bwh.harvard.edu/pph2	https://doi.org/10.1038/nmeth0410-248
Mutation taster	Combinatoire : Grantham, PhastCons, PhyloP, Exac, NNsplice	1 à >0.5	http://www.mutationtaster.org	https://doi.org/10.1038/nmeth0810-575
Mutation Assessor	L'impact fonctionnel est évalué sur la base de la conservation évolutive de l'acide aminé affecté dans les homologues de protéines.	- 5.545 à >0.65	http://mutationassessor.org/r3/	10.1186/gb-2007-8-11-r232.
LRT	test du rapport de vraisemblance	D = deleterious, N = Neutral, U = unknown	http://www.genetics.wustl.edu/jflab/lrt_query.html	doi: 10.1101/gr.092619.109
CADD	Calcul du c-score : conservation physico-chimique et conservation inter-espèce	0 à 1	http://cadd.gs.washington.edu	10.1093/nar/gky1016
FATHMM	conservation des séquences homologues et des domaines protéiques	>0.45	http://fathmm.biocompute.org.uk/	10.1093/bioinformatics/btx536
GERP	Conservation nucléotidique multi-espèces	0 à 1	http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html	https://doi.org/10.1038/nmeth0410-250
PhyloP(20/100)	conservation de blocs nucléotidiques	>0	http://hgdownload.cse.ucsc.edu/goldenPath/hg18/phyloP44way	https://doi.org/10.1007/11732990_17
PhastCons(20 ou100)	Conservation phylogénétique de blocs nucléotidiques (20 ou 100 espèces)	>0.5	http://compgen.bscb.cornell.edu/phast/	10.1101/gr.3715005

TABLEAU 11*- Outils de prédictions utilisés dans l'étude des MNDREM*

5. Vers la recherche de SNVs et CNVs dans les MNDREM…

Le séquençage première génération est l'expression dédiée à l'automatisation de la méthode de séquençage de Sanger (1975-1979). Basé sur l'amplification par l'élongation base par base d'une région unique du génome. Les performances technologiques amenaient rapidement les automates de séquençage première génération, à la possibilité de séquencer plusieurs ADN dans une seule expérience. Les données bioinformatiques générées étaient supportées par les disques durs d'ordinateurs personnels.

Le séquençage nouvelle-génération, NGS/SHD (2000 à nos jours), lui, permet l'amplification de plusieurs régions du génome en même temps. Le volume de données de séquençage généré par le NGS est notablement plus important. Et les logiciels informatiques nécessaires à l'analyse des données, nettement plus complexes. Avec le NGS, le bioinformaticien devient incontournable.

La course au décryptage du génome humain a permis un développement exponentiel des moyens de séquençage. En particulier les coûts ont nettement diminué. Alors que des milliards avaient été engagés au début du projet du séqunçage du génome humain, aujourd'hui, le coût d'un génome avoisine seulement 1000 dollars. Nous-mêmes, en sous-traitant nos analyses, nous avons obtenu l'exome de nos patients pour environ 700 euros.

Le NGS est particulièrement intéressant dans l'optique de répondre à la problématique de l'hétérogénéité génétique des MNDREM. Il est aussi une méthode de choix dans la recherche de transmission oligogénique. Il est jusqu'alors utilisé pour la recherche de variants de séquences. Très peu répandu dans la détection des CNVs ou des tendem repeats.

Pour notre étude, nous avons donc opté pour la méthode CoDE-seq qui allie séquençage d'exome et détection de CNVs. Nous l'avons mise en œuvre pour la première fois dans les MNDREM. Il s'agit d'une méthode optimisée permettant la détection de CNVs et SNVs par des sondes placées habilement sur le génome. L'enrichissement s'est fait par capture en phase solide et le séquençage était de nature terminator. Le séquenceur utilisé est un HiSeq 4000 de chez Illumina. Il s'agit d'une méthode de séquençage robuste, rapide et sensible (99,9%), largement utilisée dans les laboratoires de diagnostic et de recherche, fournie par le leader

actuel en matière de NGS. Nous avons utilisé un pipeline élaboré avec l'aide des universités de Dijon et de Lyon pour la détection des variants.

En principe, dans les méthodes de NGS 2e génération, la détection de variants par expansions de motifs répétés reste limitée. A priori, les expansions de grandes tailles, quand elles sont synthétisées, s'alignent de façon aspécifique sur le génome de référence et sont probablement éliminées par les pipelines informatiques de détection de SNVs. Cependant, la détection de ce type de variants est possible par l'adaptation et les manipulations d'outils bioinformatiques. Dans ce registre, nous avons testé l'outil GangSTR en regard d'un patient déjà investigué par la filière diagnostic.

Les résultats sont présentés dans le chapitre suivant.

CHAPITRE III

Étude démographique et génomique des MNDREM

CHAPITRE III - Étude démographique et génomique des MNDREM

Ce chapitre énoncera les principaux résultats des travaux.

Tout d'abord, nous présenterons la caractérisation démographique des MNDREM vues au centre de référence, CERCA. Elle a été réalisée par la collection, principalement rétrospective, de données médicales et biologiques.

Ensuite, nous présenterons les résultats de l'analyse de CoDESeq, NGS 2e génération permettant la détection de Variations de Nombre de Copies (Copies Number Variations, CNVs), et de Simples Variants Nucléotidiques (Single Nucleotid Variant, SNVs). Cette analyse a été pratiquée de façon sélective sur les MNDREM possiblement génétiques, sans variant délétères pathogènes identifiés par la filière diagnostic moléculaire ; et les MNDREM présentant des atypies cliniques.

1. Population générale de MNDREM

De manière majoritairement rétrospective, les données cliniques et les données d'explorations diagnostiques complémentaires ont été collectées dans les dossiers de patients atteints de MNDREM. Les informations cliniques ont été encodées pour collecter et exploiter facilement les données d'identité, d'âge d'apparition des signes, âge au diagnostic, antécédents familiaux de Maladies Neurodégénératives, et examens biologiques ou d'explorations complémentaires.

Nous décrivons ci-après la population initiale à partir de laquelle nous appliquerons les critères d'inclusion/exclusion à l'étude génomique des MNDREM.

1.1. Données descriptives

1.1.1. Données générales

Notre population comporte 309 patients, avec presque autant d'hommes (168) que de femmes (141).

Le sexe ratio est donc de 1,2.

Les principaux phénotypes de la population sont répartis comme suit :

- Maladies de Hungtington (HD) : 24/309 (8%)
- Sclérose Latérale Amyotrophique (SLA) : 61/309 (20%)
- Syndromes parkinsoniens (PARK/SPARK) : 145/309 (47%)
- Paraparésie Spastique Héréditaire (PPSH) : 5/309 (1%)
- Ataxies spinocérébelleuses (CA) : 74/309 (24%)

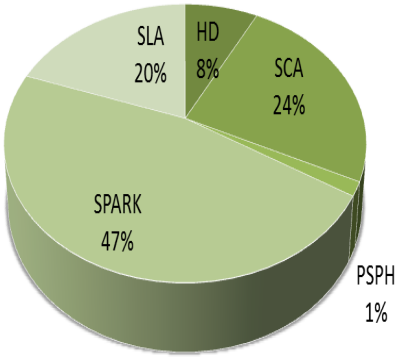


FIGURE 27 – Répartition des MNDREM vues au CERCA

L'âge moyen au diagnostic est d'environ 66 ans, quel que soit le sexe.

L'âge d'apparition moyen des premiers signes est de 61 ans 2 mois, avec un écart-type moyen de 21 ans 9 mois.

La durée moyenne d'évolution est de 15 ans.

Les formes familiales de MDREM sont définies par l'existence d'un apparenté au 1e degré (parents, enfants) ou d'un frère ou d'une sœur, atteints de MNDREM allélique ou encore de toute autre Maladie Neuro-Dégénérative (MND) ou neurologique chronique. On dénombre 40% (124/309) de patients avec une histoire familiale de MND. La FIGURE 28 ci-après montre la répartition des MND retrouvées dans les antécédents familiaux.

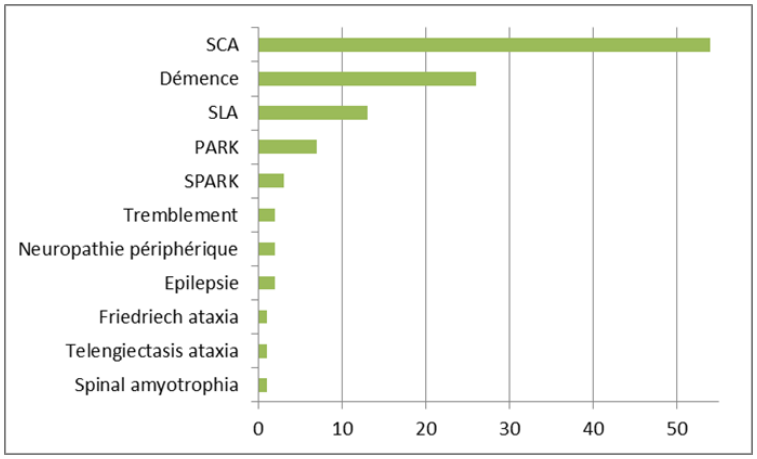


FIGURE 28 – Antécédents familiaux

1.1.2. Modes de transmission

On définit le mode de transmission selon l'ascendant atteint :

- Dominant quand un père ou une mère est atteint ;
- Récessif quand un frère et/ou une sœur sont seuls atteints sur 2 générations ;
- Sporadique quand aucun apparenté n'est atteint ;
- Lié à l'X quand dans la branche d'hérédité maternelle, un oncle ou un cousin germain issu d'un individu de sexe féminin, est atteint ;
- Mitochondrial quand une anomalie de cet organite a pu être mise en évidence.

Dans la population des MNDREM vus au CERCA, les modes de transmission, sont répartis comme suit :

- 59% sporadiques (183/309),
- 31% dominant (95/309),
- 5% récessif (15/309),
- 2% potentiellement mitochondrial (5/309)
- 3% indéterminés (11/309)

Nous n'avons pas détecté de transmission liée à l'X par notre méthode d'investigation qui ne permettait pas de rapporter de façon systématique, les antécédents familiaux de 2nd degré (oncle maternel atteint).

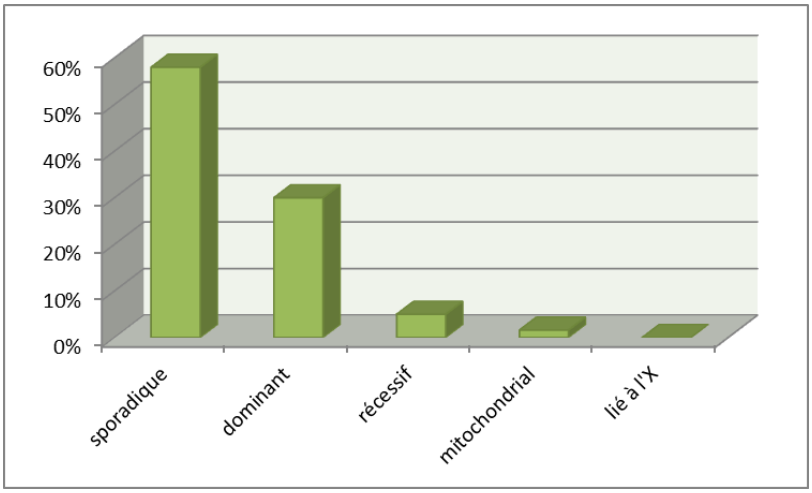


FIGURE 29 – Répartition des modes de transmission dans la population

1.1.3. Données de pathologies

Dans la maladie de Huntington, le sexe ratio est élevé. Il est de 2 (16/8). 100% des cas sont de transmission dominante. Un phénomène d'anticipation est remarquable. L'âge moyen au diagnostic est de 61 ans. L'âge moyen aux 1es signes est de 55 ans 7 mois. La durée d'évolution moyenne est de 15 ans.

Dans les CA, le sexe ratio est de 0,85 (34/40). Les modes de transmissions sont variables : 61% (45/74) dominant, 11% (8/74) récessif, 7% (5/74) mitochondrial, 11% (8/74) sporadique, 11% (8/74) non précisé. L'âge moyen au diagnostic est de 59 ans 7 mois. L'âge moyen aux 1es signes, 54 ans. La durée d'évolution moyenne, 15 ans.

Dans les SPARK, le sexe ratio est de 1,5 (86/56). Les modes de transmission sont variables : 92% (131/142) sporadique, 6% (8/142) dominant, 2% (3/142) récessif. L'âge moyen au diagnostic est de 60 ans. L'âge moyen aux 1es signes est de 54 ans. La durée d'évolution moyenne est de 15 ans.

Dans la SLA, le sexe ratio est de 0,9 (29/32). Les modes de transmission sont essentiellement sporadiques 67% (41/61), si non, dominant 23% (14/61). Les autres sont des censures ou des informations non précisées 10% (6/61). L'âge moyen au diagnostic est de 60 ans. L'âge moyen aux 1es signes est de 54 ans 6 mois. La durée d'évolution moyenne est de 15 ans.

La FIGURE 30 ci -après montre la répartition des modes de transmission par pathologies.



FIGURE 30 – Répartition des modes de transmission par pathologies

1.2. MNDREM génétiques : présentations cliniques

Dans notre population, 24% des MNDREM ont un variant pathogène identifié (75/309). Mais la proportion de ces variants varie selon les phénotypes :

- Dans HD, 100% (24/24) des cas sont identifiés
- Dans les CA, 53% (39/74)
- Dans la SLA, 13% (8/61)
- Dans les PPSH, 20% (1/5)
- Et aucune dans les SPARK (0/145).

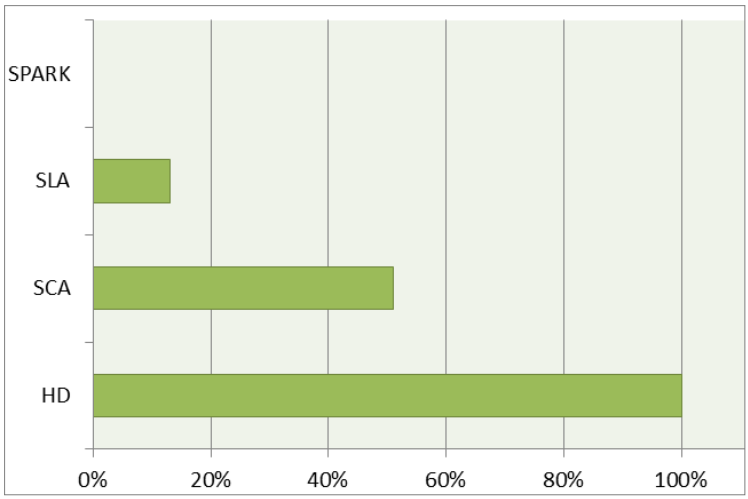


FIGURE 31 – Proportion de variant causal identifiée par phénotype

Nous détaillons ci-après, les présentations cliniques des MNDREM génétiques de la population.

1.2.1. La maladie de Huntington

✓ HD et HDL2 exclusivement

On dénombre 24 cas de patients atteints de maladie de Huntington. Il y a 14 cas dans HD (*HTT*) et 10 dans HDL2 (*JPH3*). Donc deux gènes sont exclusivement impliqués : *HTT* et *JPH3*.

HTT code pour la protéine Huntingtin qui joue un rôle dans le transport ou la fonction vésiculaire médiée par les microtubules. Tandis que JPH3 code pour une protéine de la famille des junctophilines qui contribuent à la formation de

complexes membranaires jonctionnels. Ces complexes relient la membrane plasmique au réticulum endoplasmique ou sarcoplasmique dans les cellules excitables. JPH3 est spécifique au cerveau et semble avoir un rôle actif dans certains neurones impliqués dans la coordination motrice et la mémoire.

Les variants pathogènes responsables du phénotype sont des expansions de motifs répétés. La moyenne de répétitions parmi le cas *HTT* est de 44 triplets (41 à 48 CAG, patho>36) et 45 pour *JPH3* (42 à 58 CAG/CTG, patho>41). Tandis que l'allèle sain comporte un nombre de répétitions compris entre 13 et 19 (moyenne 15,7).

Le sexe ratio est nettement plus important dans HDL2 que dans HD, respectivement 9 (9 hommes pour 1 femme) contre 1 (7 hommes pour 7 femmes).

Une histoire familiale est identifiable dans 13 cas sur 14. Dans l'HDL2, une seule grande famille martiniquaise comporte tous les individus atteints. La transmission de HD est toujours dominante avec un phénomène d'anticipation.

Les mouvements anormaux observés sont principalement la dystonie, les mouvements involontaires, la chorée. Une dysarthrie est souvent présente. Sur le plan neuropsychologique, les troubles mnésiques (5/14), et des troubles comportementaux sont récurrents. Ceux-ci sont associés à des changements d'humeur (3/14), des anosognosies (1/14), des hallucinations (2/14), des attaques de panique (1/14), des dépressions (4/14).

A cause des mouvements involontaires et grâce à la génétique, l'IRM cérébrale n'est pas systématique puisqu'elle n'est pas limitante pour le diagnostic. Elle a cependant été réalisée dans 6 cas sur 14 : 2 cas de *JPH3* et 4 cas de *HTT*. Toutes les IRM étaient comparables, montrant systématiquement une atrophie corticale diffuse, parfois accompagnée d'atrophie des noyaux caudés, d'une dilatation ventriculaire ou d'hypersignaux ponctiformes.

- ✓ **Cas particuliers : forme précoce**

On dénombre une forme précoce, débutée avant 20 ans dans HD et HDL1. Avec un nombre de triplets égal à, respectivement, 61 et 58 CAG.

Le patient HD est une femme chez qui les signes sont apparus à l'âge de 21 ans par des difficultés dans les études supérieures, des troubles dépressifs associés à « un

syndrome hallucinatoire stéréotypé » et quelques rares mouvements involontaires. Les signes moteurs décelés à l'âge de 26 ans sont un tableau akinéto-rigide sans véritables mouvements anormaux. L'UHDRS était alors à 20 et il est passé à 35 en un an, malgré les traitements. A ce jour, une véritable chorée n'est jusqu'alors pas manifeste chez cette patiente âgée de 27 ans.

Chez le patient HDL2, les signes sont apparus vers l'âge de 18 ans. Le patient, de sexe masculin, présentait alors des signes moteurs à type de dystonie et mouvements brusques involontaires latéralisés droits. Il avait une instabilité à la marche provoquant des chutes répétées. L'IRM cérébrale réalisée à 19 ans montre une atrophie cérébelleuse associée à des signes d'ischémie, probablement en lien avec une pratique toxicomane. Ceci dit, l'atteinte cognitive est marquée : sur le plan développemental, le patient aurait eu des difficultés d'apprentissage et des crises convulsives dès l'âge de 3 ans. Les troubles neuropsychiatriques et comportementaux se sont renforcés à partir de 20 ans avec des hallucinations auditives et visuelles, des troubles du comportement avec une agressivité notoire envers autrui. Le patient souffre par ailleurs de maladie de Crohn. A 38 ans, l'UHDRS est de 66.

1.2.2. Les ataxies spinocérébelleuses

Une proportion de 53% (39/74) des cas d'ataxies cérébelleuses ont une étiologie génétique identifiée. Ces cas se transmettent sur le mode majoritairement dominant 77% (30/39), mais il y a également des formes récessives 15% (6/39) ou encore mitochondriales 8% (3/39).

✓ Ataxies dominantes

Les ataxies cérébelleuses dominantes sont représentées par SCA2 (26/30), SCA3 (2/30), SCA1 (1/30), et SCA6 (1/30). Les variants délétères sont toujours des expansions pathogènes de motifs répétés dans les gènes, respectivement, *ATXN2*, *ATXN3*, *ATXN1*, *CACNA1A*.

ATXN2 agit comme un régulateur de l'Epithelial Growth Factor Receptor (EGFR) ;

Tandis que *ATXN3* est une enzyme de désubiquitinylation impliquée dans l'homéostasie des protéines, la transcription, la régulation du cytosquelette, la myogenèse et la dégradation des substrats chaperons mal repliés ;

ATXN1 est un facteur de liaison à la chromatine capable de réprimer la signalisation Notch et peut-être impliqué dans le métabolisme des ARN et le développement cérébral ;

Enfin *CACNA1A* code pour un canal calcique voltage dépendant qui permet l'entrée des ions calcium dans les cellules excitables. Il est impliqué dans divers processus comme la contraction musculaire, la libération d'hormones ou de neurotransmetteurs, l'expression génique, la motilité cellulaire, la division et la mort cellulaire.

✓ Ataxies récessives

Dans notre population, les ataxies récessives génétiquement déterminées sont en lien avec, soit un variant délétère du gène *ATM*, responsable de l'ataxie télangiectasie (2/6) ; soit un variant délétère du gène *FRDA*, responsable de l'ataxie de Friedreich (1/6) ; ou encore, un variant délétère du gène *POLG*, impliqué dans les ataxies d'origine mitochondriale (3/6).

Variant pathogène du gène *POLG*

POLG est un gène nucléaire situé en 15q26.1 codant pour une sous-unité catalytique de l'ADN polymerase gamma. Il s'agit de l'unique ADN polymérase active dans les mitochondries et donc capable de répliquer l'ADNmt.

Les variants pathogènes du gène *POLG* sont responsables d'un continuum de phénotypes et qui ont été définis cliniquement bien avant que leur base moléculaire ne soit connue : le syndrome d'Alpers-Huttenlocher, le spectre des ataxies neuropathiques, le spectre des myocérébrohépatopathies infantiles, l'épilepsie myoclonique avec myopathie et ataxie sensorielle, l'ophtalmoplégie externe progressive (récessive ou dominante) et le syndrome de Leigh.

La combinaison des signes cliniques suivants, peut faire évoquer les variants pathogènes de ce gène : hypotonie, retard de développement, épilepsie, trouble du

mouvement (myoclonie, dysarthrie, choréoathétose, parkinsonisme…), myopathie (Ptose, ophtalmoplégie, faiblesse proximale plutôt que distale des membres avec fatigue et intolérance à l'exercice…), ataxie, neuropathie périphérique, régression psychomotrice épisodique, maladie psychiatrique (Dépression, trouble de l'humeur…), endocrinopathie (diabète sucré, insuffisance ovarienne prématurée…).

Dans notre population, trois cas sont rencontrés dans une fratrie issue d'un couple d'origine martiniquaise a priori non consanguin.

L'aînée a développé une polyneuropathie dysesthésiante et ataxiante à l'âge de 38 ans. Elle a bénéficié de soutien psychologique pour des troubles anxiodépressifs. La biopsie musculaire n'avait pas été faite chez cette patiente.

Le cadet, a été diagnostiqué à 41 ans. Il présentait un syndrome parkinsonien juvénile associé à une neuropathie ataxiante et paresthésiante aux 4 membres dont les signes seraient apparus vers l'âge de 38 ans. Les troubles ont évolué vers une faiblesse musculaire et l'invalidité motrice vers l'âge de 46 ans, nécessitant un fauteuil roulant. La biopsie musculaire montrait une prédominance des fibres de type 2 et une discrète surcharge lipidique en rapport avec l'atteinte mitochondriale. Mais la chaîne respiratoire était normale. L'étude de l'ADNmt mettait en évidence des délétions multiples visibles par PCR « grands fragments » et Southern-Blot traduisant une instabilité de l'ADNmt, probablement en rapport avec l'atteinte mitochondriale.

La benjamine a débuté la maladie vers l'âge de 33 ans par une atteinte motrice à type d'intolérance à l'effort et faiblesse musculaire proximale marquée. Le périmètre de marche autonome était limité à 7 minutes. La neuropathie hypoesthésiante distale s'est installée rapidement, accompagnée d'une ataxie proprioceptive. La biopsie musculaire n'a pas été pratiquée chez cette patiente.

Aucun des membres de la fratrie ne présentait de trouble oculomoteur ni d'épilepsie.

L'étude du gène *POLG* révèle un variant c.695G>A p.Arg232His à l'état hétérozygote, présent chez tous les cas de la fratrie, associé à un polymorphisme p.Arg546Cys. En l'état des connaissances, le variant p.Arg232His est responsable de tableau clinique type Alpers et CMT quand il est associé à un autre variant pathogène. La présence de ces 2 variants semble donc insuffisante pour conclure à une atteinte

du gène *POLG*. Cependant, la combinaison entre instabilité de l'ADNmt et variant pathogène pourrait justifier la pathologie.

Aussi, la recherche de variant structurelle est donc toujours en cours dans le laboratoire de référence. En particulier, les grands réarrangements de l'ADNmt, notamment chez les apparentés non testés à ce niveau (aînée et benjamine).

Variant pathogène du gène *FRDA*

Les variants pathogènes du gène *FRDA* sont associés à la maladie de Friedreich. Cette affection neurodégénérative se caractérise habituellement par une démarche ataxique, une dysarthrie, une dysphagie, un trouble oculomoteur, une perte des réflexes tendineux profonds, des signes du faisceau pyramidal, une scoliose, et dans certains cas, une cardiomyopathie, un diabète sucré, une cécité et un trouble de l'audition.

Dans notre population, il y a un cas d'origine Européenne (France), présentant les signes de la maladie et une variation délétère du gène *FRDA*, non retrouvée dans le dossier du patient, qui avait été diagnostiqué avant son arrivée en Martinique. Les premiers signes sont apparus de façon classique vers l'âge de 22 ans par des troubles de l'équilibre. Il présente quelques troubles dysarthriques. Des troubles du rythme cardiaque ont conduit à la pose de pacemaker à l'âge de 57 ans. Par ailleurs, il présente des nodules thyroïdiens et une hypothyroïdie. L'histoire familiale rapporte une sœur cadette atteinte également.

Variant pathogène du gène *ATM*

Le gène *ATM* code pour une protéine kinase de la famille PI3/PI4. Il s'agit d'un des contrôleurs principaux des voies de signalisation des points de contrôle du cycle cellulaire. Ces points de contrôle sont indispensables à la réponse cellulaire aux dommages de l'ADN et à la stabilité du génome. Les variants pathogènes de ce gène sont responsables de pathologies cancéreuses (leucémie ou gliome) ou encore de l'apraxie oculomotrice et de l'Ataxie Télangiectasie.

Deux cas d'ataxie télangiectasie, non apparentés a priori, avec variants délétères du gène *ATM* sont répertoriés dans la population MNDREM.

Le premier, est un patient perdu de vue, originaire de Guyane probablement d'origine surinamaise. Il a été dépisté au décours de l'enquête familiale débutée après le diagnostic de son frère aîné qui souffrait notamment de déficit immunitaire et diarrhée persistante, accompagnés de troubles moteurs. Une variation pathogène du gène *ATM* est responsable de ce tableau, mais le résultat moléculaire de l'analyse du gène n'est pas présent dans le dossier.

Le second cas est également originaire de Guyane. Le diagnostic a été fait à l'âge de 5 ans, face à un syndrome cérébelleux statique et des télangiectasies conjonctivales. La maladie avait débuté par des tremblements qui se sont aggravés lentement jusqu'aux troubles de l'équilibre. La variation pathogène a été mise en évidence dans un laboratoire de référence et n'est pas disponible dans le dossier du patient.

- ✓ **Ataxies d'origine mitochondriale**

Les ataxies mitochondriales sont en lien avec des variants pathogènes de l'ADNmt, en propre. Nous rapportons des variations délétères des gènes *ARNt Leu* et *MT-CO1*. Ces gènes sont situés sur l'ADNmt. Ils se transmettent donc sur le mode mitochondrial, c'est-à-dire, hérités par les mères. Le conseil génétique est rendu complexe par le phénomène d'hétéroplasmie (répartition aléatoire du nombre de variants pathogènes de l'ADNmt dans les tissus du corps entier).

Variant pathogène de l'ARNt Leu

Associé au syndrome de MELAS : Myopathie mitochondriale, Encéphalopathie, Acidose Lactique et des tableaux neurologiques aigus ressemblant à des accidents ischémiques cérébraux. La prévalence n'est pas connue mais probablement extrêmement faible. Les symptômes récurrents sont : myocardiopathie, surdité, diabète, petite taille, faiblesse musculaire, retard mental, troubles de l'apprentissage, de la mémoire ou de l'attention.

Le syndrome de MELAS est provoqué par des variants pathogènes de l'ADN mitochondrial. Au moins 10 variants différents ont été identifiés mais 80% des patients sont porteurs du variant délétère 3243A>G dans le gène de l'ARN de transfert de la leucine (ARNt Leu).

Du fait de l'hétéroplasmie, la proportion de variants pathogènes peut varier considérablement d'un tissu à l'autre mais, dans le syndrome de MELAS elle est supérieure à 90% de la population d'ADN mitochondrial dans tous les tissus et peut donc être cherchée dans le sang.

Nous rapportons le cas d'une fratrie issue d'un couple d'origine martiniquaise, a priori non consanguin. Sur les 6 enfants de ce couple, au moins 2 ont été vus dans le cadre du syndrome de MELAS avec variant pathogène 3243A>G dans le gène de l'ARN Leu.

L'aînée des deux cas est une patiente qui a présenté à l'âge de 59 ans, un syndrome cérébelleux statique associé à une dysarthrie, une hypopallesthésie distale des membres inférieurs et une abolition des réflexes Achilléens. Elle avait développé un diabète insulinodépendant à l'âge de 32 ans et souffrait d'hypoacousie bilatérale appareillée depuis l'âge de 10 ans. L'IRM cérébrale montrait des calcifications des noyaux gris centraux associées à une atrophie vermienne.

Son cadet a été vu pour la première fois à 47 ans pour un tableau similaire : une ataxie évoluait depuis plusieurs années, une surdité depuis l'âge de 25 ans, une IRM montrant une atrophie médullaire et vermienne et une biopsie musculaire montrant des Ragged Red Fibers (RRF) caractéristiques de l'atteinte mitochondriale. Il présentait en plus, un diabète non insulinodépendant, un antécédent récent d'accident vasculaire cérébral postérieur et des épisodes d'épilepsie généralisée.

Variant pathogène du gène *MT-CO1*

MT-CO1 code pour la Cytochrome C Oxydase 1 mitochondriale. La cytochrome C oxydase est le composant de la chaîne respiratoire qui catalyse la réduction de l'oxygène en eau. Les sous-unités 1 à 3 forment le noyau fonctionnel du complexe enzymatique. CO I est la sous-unité catalytique de l'enzyme.

Les variants pathogènes du gène *MT-CO1* sont impliqués dans l'anémie sidéroblastique autosomique récessive réfractaire à la pyridoxine 2, les myoglobinuries récurrentes, la surdité sensorielle non syndromique d'origine mitochondriale, et un large spectre de symptômes associés au déficit en complexe IV de la chaîne respiratoire. Parmi lesquels principalement des atteintes

neuromusculaires, sensorielles et cardiaques allant de la myopathie isolée, à la maladie multi systémique sévère, d'apparition juvénile ou à l'âge adulte.

L'individu est vu pour la pemière fois à l'âge de 50 ans. Il présente une marche ataxique, une hypoacousie et des troubles mnésiques. L'examen clinique révèle une abolition des réflexes aux membres inférieurs, des pieds creux, une amyotrophie distale nette et une grande akinésie avec amimie sans tremblements. En réalité, les troubles de la marche sont d'aggravation progressive depuis l'âge de 18 ans.

L'interrogatoire retrouvait des antécédents familiaux de tremblements et de marche ataxique chez 2 sœurs et la mère. Un scanner cérébral effectué à l'âge de 38 ans suite à une chute, montrait des calcifications des noyaux gris centraux. Par la suite, l'ENMG a montré des signes de neuropathie périphérique, et l'IRM cérébrale dans la limite de la norme, montre toutefois, une atrophie diffuse et un hypersignal périventriculaire. A 53 ans, le syndrome cérébelleux est franc et le score SARA est de 16.

La biopsie musculaire présente des anomalies évoquant un processus pathologique mitochondrial, notamment plusieurs fibres présentent des amas mitochondriaux anormaux leur donnant un aspect rouge légèrement déchiqueté pour certaines, compatibles avec l'aspect typique de Ragged Red Fibers, caractéristique de l'atteinte mitochondriale ; ainsi que des fibres claires à la réaction de Cytochrome Oxydase, orientant vers un déficit de cette enzyme.

L'étude de la chaîne respiratoire met en évidence un effondrement de l'activité du complexe IV. L'analyse de l'ADNmt révèle une variation pathogène m.7197G>A, p.Gly432Ser du gène *MT-CO1*, compatible avec les anomalies histologiques observées et le phénotype du patient.

- ✓ **Cas particulier : variant pathogène *DMPK***

On retrouve une ataxie chez un patient atteint d'une Dystrophie Myotonique de type 1 de forme infanto-juvénile. Cette pathologie neuromusculaire est causée par une variation délétère avec expansion de motifs répétés dans la région non traduite 3'-UTR du gène *DMPK* (DM1 Protein Kinase).

DMPK code pour une sérine-thréonine kinase qui est étroitement apparentée aux kinases qui interagissent avec les membres de la famille Rho des petites GTPases.

La région 3'UTR de *DMPK* contient 5-38 copies d'une répétition de trinucléotides CTG. L'expansion instable de ce motif pouvant aller de 50 à 5000 copies, provoque une DM1 d'autant plus grave que le nombre de trinucléotides est élevé. L'expansion répétée est associée à la condensation de la structure de la chromatine locale perturbant l'expression des gènes dans cette région. Des formes congénitales de pronostic neuromusculaire et neurodéveloppemental très sévère sont observées.

Dans notre cas, c'est à l'âge de 18 ans, face à une hypotonie faciale, un ptosis bilatéral, une limitation de la verticalité du regard, un déficit moteur des ceintures et une myotonie du thénar droit, que le diagnostic de DM1 est suspecté.

L'ENMG confirme l'existence de décharges myotoniques abondantes. Mais une instabilité à la marche avec véritable ataxie complète ce tableau de façon inhabituelle. L'IRM cérébrale révèle une leucoencéphalopathie périventriculaire kystique.

A 19 ans, une cataracte, des troubles de l'oralité (dysarthrie, dysphagie), des troubles sphinctériens et des troubles cognitifs viennent alourdir le tableau clinique d'un jeune homme qui avait obtenu un CAP menuiserie et pratiquait le cyclisme en club pendant quelques mois avant la perte de la marche à 20 ans.

La biopsie musculaire montrait des fibres atrophiques, des aspects cellulaires de nécrose régénération, des internalisations nucléaires parfois multiples en faveur du diagnostic de maladie de Steinert.

L'IRM cérébrale à 24 ans rapporte un état criblé associé à la leucopathie périventriculaire.

Un variant pathogène par répétition de 500 à 600 CTG (mise en évidence par Southern-Blot) dans le gène *DMPK* est responsable du tableau neuromusculaire de ce patient, chez qui, de façon inhabituelle, existait également un syndrome cérébelleux.

1.2.3. La Sclérose Latérale Amyotrophique

On dénombre 7 cas de SLA génétiquement déterminées, soit 11,5% (7/61). Tous présentent un variant pathogène du gène *SOD1* et une histoire familiale en accord avec une transmission autosomique dominante pour 6 cas sur 7. En effet, un variant délétère est trouvé chez un cas a priori sporadique.

✓ Variants pathogènes *SOD1*

Les variants pathogènes de ce gène sont impliqués dans 10% des formes familiales de SLA [98]. Le gène *SOD1* code pour l'enzyme superoxyde dismutase impliquée dans le fonctionnement mitochondrial, en particulier l'équilibre des espèces oxydatives.

Essentiellement, les variations pathogènes missense de ce gène sont responsables de l'ALS. Les variants délétères perte de fonction génèrent un syndrome neurodéveloppemental pouvant associer une hypotonie axiale, des fasciculations, un retard global au développement avec absence de langage, une atteinte de la corne antérieure, une ataxie et tétraplégie spastique progressive.

Dans notre population, 6 cas sur 7 appartiennent à une seule et même grande famille martiniquaise qui arbore le variant délétère p.Gly94Val. De façon surprenante, on observe dans cette grande famille, un phénomène d'anticipation peu décrit dans cette pathologie, controversé, toujours associé à un variant *SOD1* mais jamais rapporté avec ce variant.

Le dernier cas, non rattaché à cette famille, est a priori sporadique et présente le variant pathogène p.Glu121Gly. De façon probablement anecdotique, l'histoire familiale rapporte une nièce atteinte d'amyotrophie spinale de type 2. Les délétions de ce gène avaient été recherchées et exclues sur ce cas SLA par le biais du conseil génétique.

1.2.4. Les paraparésies spastiques héréditaires (PPSH)

Dans les PPSH, un cas sur cinq est génétiquement identifié. Il s'agit du cas d'une patiente d'origine Guadeloupéenne, qui a été vue dans le service suite à l'apparition de trouble de la marche à 49 ans.

Au moment du diagnostic, à 55 ans, la démarche est spastique avec un frottement des cuisses, la marche sur les talons et les pointes est impossible, il existe un déficit moteur asymétrique des psoas, une sensibilité non altérée, des réflexes ostéotendineux pyramidaux et vifs aux quatre membres, tandis que les réflexes cutanés plantaires sont indifférents.

Des antécédents familiaux sont en faveur d'une transmission autosomique dominante.

L'analyse du gène *SPST* met en évidence une variation pathogène c.1099-7_1102del11insATAA à l'état hétérozygote qui entraîne théoriquement la synthèse d'une protéine tronquée. Cette variation délétère est associée à un variant c.865C>T p.Hist289Tyr dont la pathogénicité n'est pas connu.

L'étude de ségrégation familiale montre que les 2 variants sont situés sur le même allèle. Ce variant est donné pour responsable de la Paraparésie Spastique Héréditaire de type 4 (PSPH, SPG4).

1.3. MNDREM sans étiologie génétique

Nous présentons ci-après les phénotypes sans étiologie génétique de notre population. Ainsi que la proportion de cas pouvant faire évoquer une composante étiologique génétique de part, l'héritabilité, c'est-à-dire, la présence d'antécédents familiaux, ou par une particularité phénotypique : précocité de signes, anticipation ou atypie clinique.

1.3.1. Phénotypes concernés

Une proportion de 76% (234/309) des MNDREM de notre population sont sans étiologie génétique. Certains cas ont été testés en vain par les filières de diagnostic de biologie moléculaire, par séquençage de type Sanger ou NGS type panel de gènes.

- ✓ **Maladie de Parkinson/Syndromes Parkinsoniens, (PARK/SPARK)**

Les syndromes parkinsoniens (SPARK) ou maladies de Parkinson représentent 142 cas sur 234 sans étiologie, soient 60% des MNDREM sans étiologie.

Parmi lesquels, 130 cas sporadiques, 3 cas récessifs et 9 cas dominants. Il y a parfois des cas sporadiques avec autres antécédents familiaux de MND.

- ✓ **Sclérose Latérale Amyotrophique (SLA)**

Parmi les MNDREM sans étiologie, on dénombre 53 cas de SLA, soit 23% (53/234) : 5 censures ou non précisés, 40 sporadiques et 8 dominants. Ces-derniers ont été explorés sans succès pour les différents gènes de la SLA par la filière diagnostic habituelle.

- ✓ **Ataxies Spinocérébelleuses (CA)**

On dénombre 31 cas de CA parmi les MNDREM sans étiologie : 13% des MNDREM sans étiologie sont des CA (31/234).

Parmi lesquels : 74% (23/31) (présentent des antécédents familiaux, 16% (5/31) sporadiques dont 1 mitochondrial, 9% (3/31) censures.

Les formes familiales (23/31) se répartissent selon les modes de transmission suivants : 14 cas dominants, 8 cas récessifs et 1 cas non précisé.

✓ **Autres phénotypes**

Une proportion de 1,7% sont des PPSH (4/234) : dont 1 forme probablement dominante débutée à 15 ans, mais certainement syndromique, et 2 autres sporadiques. 3% (7/234) des autres cas sont en fait des MNDREM d'étiquetage clinique complexe.

Le diagramme ci-après (FIGURE 32) présente la répartition des MNDREM avec ou sans étiologies. Et la répartition des phénotypes de la catégorie sans étiologie.

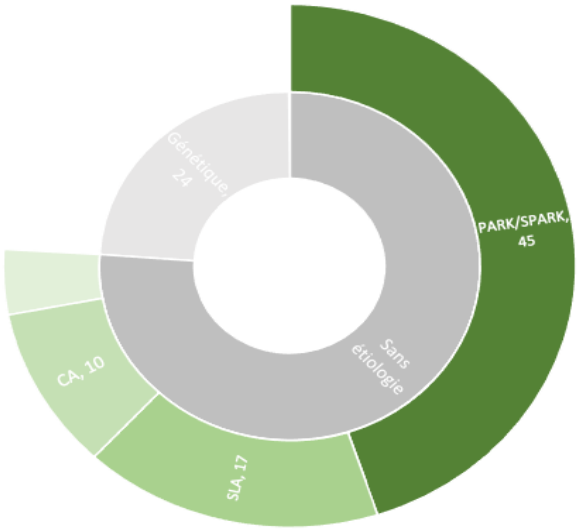


FIGURE 32 - Répartition des phénotypes avec ou sans étiologies

1.3.2. MNDREM avec antécédents familiaux

On considère un antécédent familial comme étant la présence d'un apparenté ou 1e ou 2nd degré atteint de la même maladie ou d'une pathologie allélique, ou encore de toute autre maladie neurodégénérative chronique (MND). 23% (50/234) des MNDREM sans étiologie, présentent un antécédent familial de pathologie neurodégénérative ou neurologique chronique.

Les phénotypes sans étiologie, avec antécédents familiaux sont :

- Huit cas, de SLA de transmission dominante. Ces huit cas ont été testés sans succès pour les gènes majeurs de diagnostic de SLA.
- Vingt-trois cas de CA compatibles avec une transmission dominante (14) ou récessive (8)
- Dix-huit cas de SPARK: 3 cas récessifs et 9 cas dominants, auxquels on ajoute 6 formes sporadiques avec autres antécédents familiaux de MND type (tremblement, démence etc…).

1.3.3. MNDREM avec précocité des signes, anticipation ou atypie clinique

Nous nous sommes intéressés également aux MNDREM présentant des atypies. Soit par la présence de :

- Symptômes inhabituels dans leur présentation clinique :
- Soit par un âge de début précoce ;
- Soit par une transmission inhabituelle de la pathologie, comme la présence d'un phénomène d'anticipation.
- Ou autres symptômes associés

Dans les SPARK, nous avons noté au moins un cas d'âge de début à 48 ans, ou encore 1 cas avec des troubles cognitifs associés à une maladie de Parkinson.

De même, un phénomène d'anticipation dans une famille de SLA avec variant pathogène *SOD1* est étonnant.

2. Présentation clinique de l'échantillon sélectionné

Compte tenu des critères de sélection de l'étude et de la disponibilité des échantillons d'ADN, nous avons testé 14 ADN dont 1 provenant d'un apparenté non atteint. Nous détaillons ci-après, la présentation clinique des 13 cas de l'échantillon sélectionné.

2.1. Cas de Sclérose Latérale Amyotrophique (SLA)

2.1.1. SLA atypique

Cette femme (17S353) présente des troubles dysarthriques depuis l'âge de 68 ans. L'examen clinique de cette même année met en évidence une dysphonie et une dysarthrie parétique, une amyotrophie des premiers espaces interosseux et fasciculations de la langue. Il semble également y avoir une atteinte cognitive et un très probable syndrome de Meige. On retient alors le diagnostic de SLA avec atteinte cognitive.

L'IRM cérébrale à 71 ans montre une leucopathie microvasculaire. L'ENMG à 72 ans révèle une activité spontanée tonique discontinue.

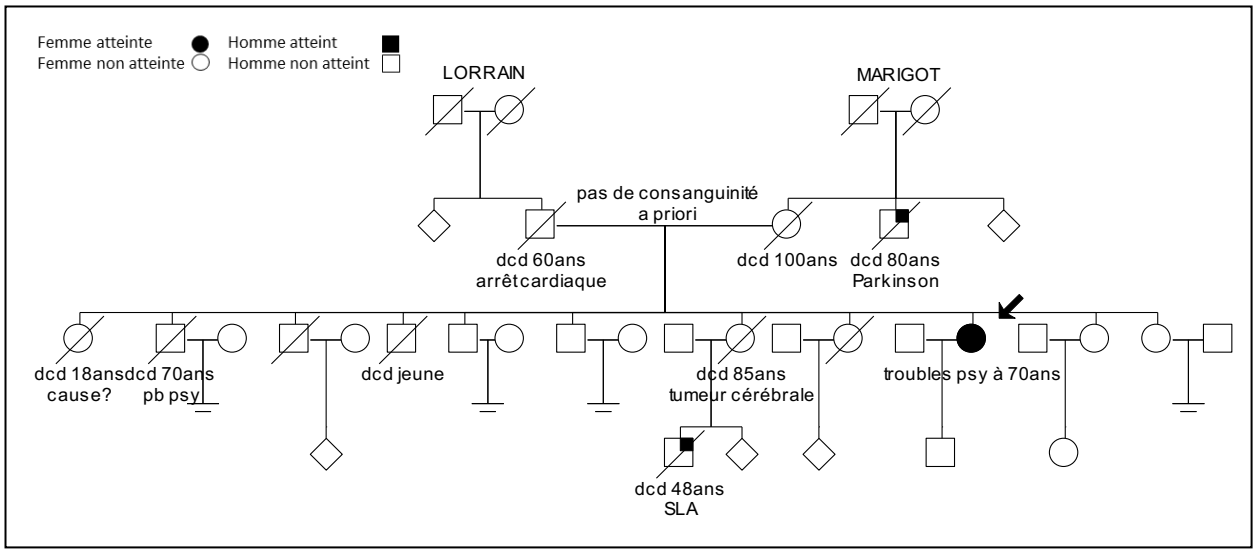


FIGURE 29 – Pedigree du cas 17S353 indiquée par une flèche.
Les apparentés atteints de MNDREM remarquables sont noircis au quart de leur symbole. Dcd : décédé ; psy : psychiatrique ; pb : problème

Dans les antécédents familiaux, on note l'existence d'un neveu, issu d'une sœur asymptomatique, décédé à 48 ans de SLA. Ce qui pourrait suggérer une transmission dominante de pénétrance variable.

Les variants pathogènes des gènes de SLA, classiquement explorés par la filière diagnostic, ont été exclus. Il n'y a pas de variants pathogènes dans les gènes suivants : *SOD1, C9ORF72, ALS2, ANG, CHMP2B, DCTN1, FIG4, FUS, OPTN, SETX, SPG20, SQSTM1, TARDBP, TBK1, UBQLN2, VAPB, VCP.*

2.1.2. Anticipation familiale à variants pathogènes *SOD1*

Dans cette grande famille d'origine martiniquaise, la SLA est associée au variant de classe 5 c.281G>T, p.G94V de l'exon 4 du gène *SOD1*. Le pedigree montre une transmission autosomique dominant de la SLA, sans expression de FTD compatible avec le diagnostic de GALS (FIGURE 34). Nous disposions d'informations médicales sur 9 individus atteints réparties dans le tableau 14 ci- après.

Le début était principalement d'expression motrice dans les membres inférieurs (5/9), à l'exception de 2 cas aux membres supérieurs et 2 cas de début pseudo polynévrétique. Les premiers symptômes étaient principalement moteurs, à l'exception d'un cas avec des symptômes mixtes bulbaires et moteurs. Parmi les signes initiaux, les crampes étaient fréquentes (4/9) tout comme l'atrophie (4/9), les fasciculations (3/9) ou les faiblesses musculaires (3/9).

Nous avons appliqué un test de Mann-Whitney pour comparer l'âge d'apparition, le délai de survie et l'âge de décès parmi les générations III et IV les plus informatives. Le test a révélé des différences statistiques d'âge d'apparition (p = 0, 03) et d'âge de décès (p = 0, 01), mais pas de délai de survie (p = 0, 05).

Les gènes présentant des variants à expansion de motifs pathogènes connus dans la SLA ont été explorés. On ne retrouve pas de variants pathogènes dans les gènes *ATXN2* ni *C9ORF72*. Nous avons effectué l'analyse CoDESeq sur 3 patients de cette famille : 13S001 (III.2), 10S030 (III.6), 16S079 (V.1).

Identification			Caractéristiques cliniques					
Id.	sexe	Age début des signes	Durée de survie (mois)	age au décès	TYPE ALS	Symptome inaugural	Signes Initiaux	ENMG
III.1	M	46	108	55	expression motrice (membre inférieur)	Moteur	crampes, atrophie distale	certain
III.2	M	54	144	66	pseudopolynevretic	Moteur	déficit	ND
III.3	M	50	72	53	expression motrice (membre inférieur)	Moteur	atrophie distale, fasciculations	Certain
III.4	M	NP	NP	50	NP	NP	NP	NP
III.5	M	NP	NP	46	NP	NP	NP	NP
III.6	M	61	108	70	expression motrice (membre inférieur)	Moteur, bulbaire	fasciculations, crampes, atrophie distale	ND
IV.1	M	30	72	36	pseudopolynevretic	Moteur	deficit, crampes	Certain
IV.2	M	23	72	26	expression motrice (membre inférieur)	Moteur	crampes distales	Certain
IV.3	M	37	NP	NP	expression motrice (membre inférieur)	Moteur	Atrophy, deficit	Certain
IV.4	M	33	60	38	expression motrice (membre supérieur)	Moteur	fasciculations, crampes, atrophie distale	certain
V.1	M	13	228	alive	expression motrice (membre inférieur)	Moteur	fasciculations	ND

TABLEAU 14 – Caractéristiques cliniques des membres de la famille SLA avec anticipation.

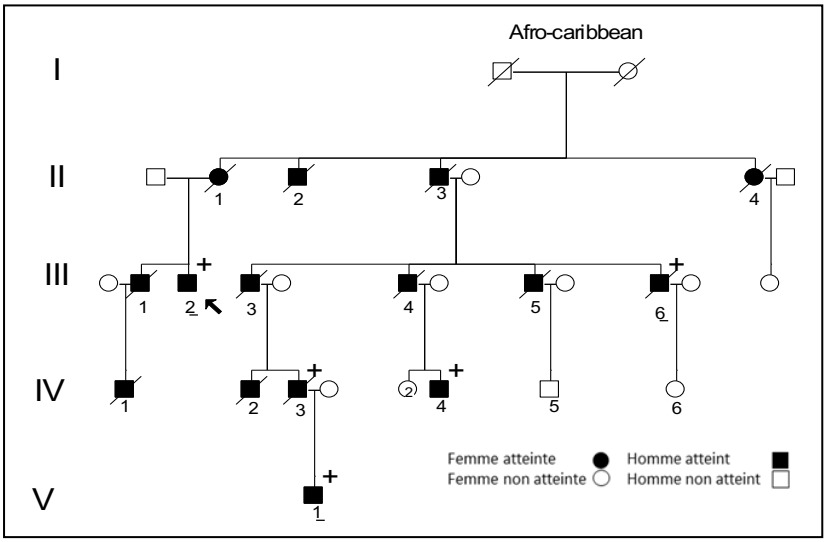


FIGURE 34 – Pedigree SLA familiale avec anticipation
Symboles : flèche pour le cas index ; «+»: Mutation SOD1 Gly94Val identifiée par analyse de routine ; le chiffre romain de la génération associé au numéro alphanumérique, indique l'identification du membre de la famille ; souligné, les cas étudiés.

2.2. Cas d'ataxies spinocérébelleuses (CA)

2.2.1. Transmission dominante

Les cas 18S231 et 17S162 sont respectivement mère et fille. Elles présentent toutes les 2, un syndrome cérébelleux statique apparu vers l'âge de 60 ans ; prédominant aux membres inférieurs pour la mère. Il y aurait une petite fille atteinte, avec un âge de début à 40 ans.

Chez le cas 18S231, l'IRM cérébrale et la ponction lombaire avaient été données pour normale ; Les anticorps anti-GAD étaient négatifs. Sur le plan génétique, les SCA 1, 2, 3, 6, 7, et 17 avaient été exclus.

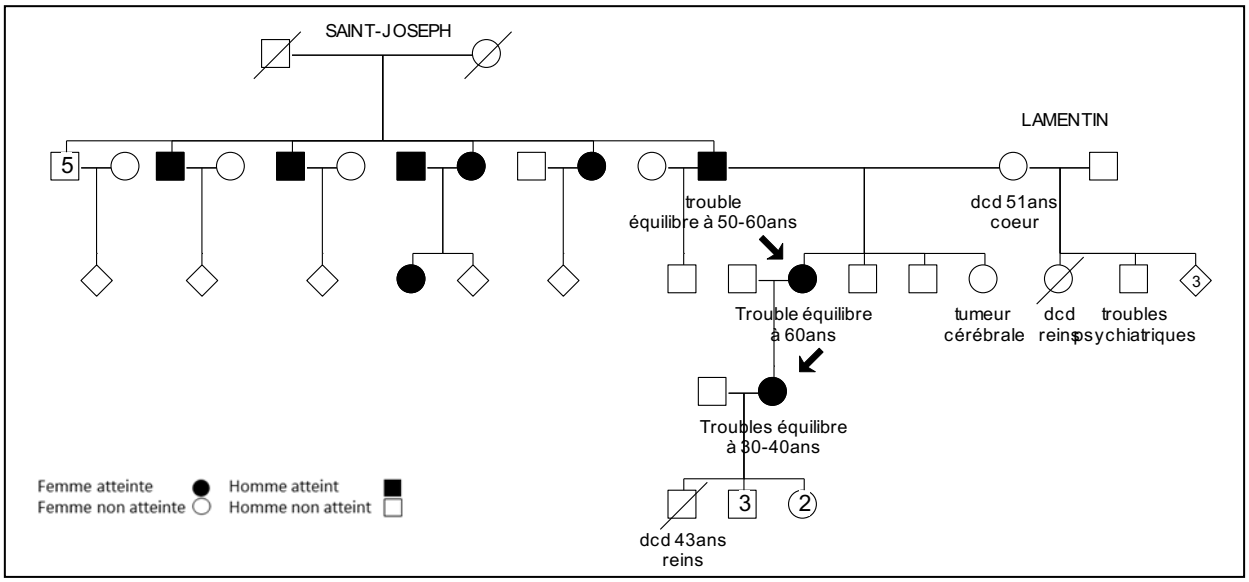


FIGURE 35 – Pedigree des cas 18S231 et 17S162 indiquées par une flèche.
Les apparentés atteints de MNDREM remarquables sont noircis au quart de leur symbole. Dcd : décédé

2.2.2. Transmission sporadique

L'analyse a été effectuée en comparant celle du cas index (10S103), une mère, et celle de son fils non atteint (18S235).

L'âge de début des signes est 28 ans avec : un syndrome cérébelleux statique et cinétique relativement symétrique, une ataxie cérébelleuse isolée et sporadique, un nystagmus horizontal prédominant à gauche, un syndrome pyramidal. Les réflexes

rotuliens étaient abolis et les réflexes achilléens peu marqués. A 29 ans, elle fait une crise convulsive, suite à l'accouchement de son 3e et dernier enfant. Elle présente depuis, un ptosis unilatéral intermittent.

Au moment du diagnostic, à 47 ans, il y a aussi des dysesthésies des mains et des pieds, une importante fatigabilité à l'effort et des troubles de la vision épisodiques. A 48 ans : l'IRM cérébrale montre une atrophie cérébelleuse isolée majeure prédominant sur le vermis ; la biopsie musculaire montre de discrètes anomalies pouvant orienter vers une atteinte métabolique ou mitochondriale ; l'ENMG montre une atteinte myélinique sensitive discrète des nerfs médians et une atteinte sensitive mixte des nerfs suraux. A 56 ans, l'évolution est toujours lentement progressive depuis l'enfance et le score SARA est à 20,5/40.

Les variants pathogènes du gène de l'ataxie de Friedreich ont été éliminés : il n'y avait pas d'expansion GAA dans le gène FRDA. Et il n'y avait pas non plus de variants pathogènes dans les gènes associés aux cytopathies mitochondriales fréquentes : il n'y avait pas de variations délétères dans les gènes : *POLG, MELAS, MERRF, NARP, Leigh, Twinkle PEO1*.

2.2.3. CA atypique

La patiente 14S121 présente en fait une probable atrophie multisystématisée dont l'évolution s'est faite de la façon suivante : syndrome cérébelleux débuté à l'âge de 39 ans, associé à des troubles psychiatriques (hallucinations). L'IRM cérébrale montrait alors une atrophie vermienne et cérébelleuse.

A 48 ans : Syndrome cérébelleux stable, sans déficit sensitivomoteur, pas de myoclonies, pas de troubles cognitifs. Apparition d'un tremblement au repos et à la marche, d'une rigidité axiale et segmentaire et d'une limitation des mouvements oculaires en particulier vers le haut. Antécédents de psychose et troubles psychologiques majeurs suivis pendant plus de 10 ans. Le discours fait apparaitre des thèmes de persécutions plutôt inquiétants. A 52 ans : Syndrome pyramidal sans signe de Babinski, nystagmus persistant, dysarthrie cérébelleuse, incontinence urinaire et constipation, dysmétrie axiale. Score SARA à 29,5.

Dans les antécédents familiaux, on suspecte la grand-mère maternelle d'avoir été atteinte (FIGURE 36). Les variants pathogènes *ATXN2*, *FMR1*, *NPC1* et *NPC2* ne sont pas retrouvés.

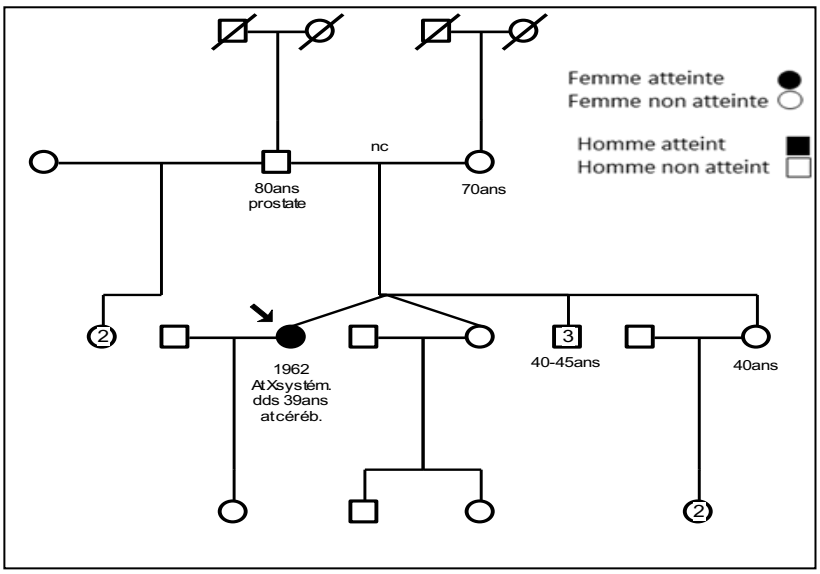


FIGURE 36 – Pedigree du cas 14S121 indiquée par une flèche.
AtXsystèm ; atrophie multisystématisée

2.3. Cas de syndromes Parkinsoniens (PARK/SPARK)

2.3.1. Transmission dominante

Chez le patient 09S079, la maladie a débutée vers l'âge de 57 ans. Le patient présentait une akinésie et une rigidité asymétrique, sans tremblement. La réponse était faiblement dopasensible. A 59 ans l'IRM cérébrale sans modification significative. A 60 ans, le bilan neuropsychologique est toutefois en faveur d'un léger déficit fronto-sous-cortical. Le patient présente par ailleurs, une hypertrophie ventriculaire gauche. La dopa-sensibilité est faible. Ce patient présente une maladie de Parkinson familiale, compatible avec un mode de transmission dominant : le père atteint avait été diagnostiqué probablement vers 70 ans. Après 10 ans d'évolution, la maladie s'était compliquée de dyskinésies, d'hallucinations, de pollakiurie, d'hypersalivation et une perte de la sensibilité des extrémités.

Chez le patient 09S079, aucune investigation moléculaire n'avait été menée.

2.3.2. Transmission récessive

Chez le cas 15S310, les troubles de la marche type petits pas (Parkinson) et modification du comportement sont apparus vers l'âge de 62 ans. Il a présenté des hallucinations visuelles, des troubles mnésiques, et des troubles du sommeil dans le même temps (entre 60 et 62 ans). A l'examen clinique, on retrouve : tremblement de repos côté droit, hypertonie extrapyramidale droite et un tableau anxiodépressif bien marqué. L'IRM cérébrale est normale à ce moment.

Dans les antécédents familiaux, on trouve une sœur atteinte avec un âge de début à 45 ans, le père d'origine Italienne est décédé à 82ans, il avait des tremblements apparus à l'âge de 70 ans.

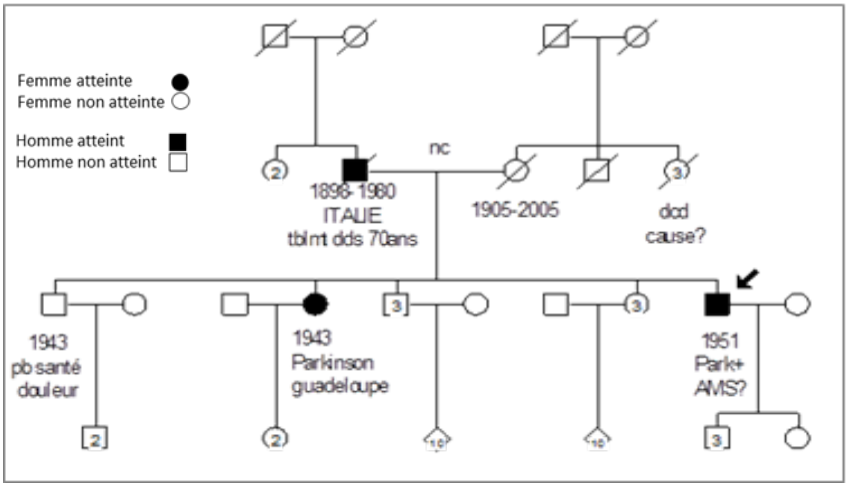


FIGURE 37 – Pedigree du cas 15S310 indiqué par une flèche.
tblmt : tremblement ; pb : problème

On évoque une transmission récessive ou dominante avec expressivité variable, et on note les âges de début plus précoce chez les atteints des jeunes générations. Dopa-sensibilité non recherchée**.** Les investigations génétiques excluaient, chez le patient 15S310, des variants pathogènes dans les gènes : *ATP13A2, ATP1A3, DNAJC6, FBXO7, GCH1, LRRK2, PARK2, PARK7, PINK1, PLA2G6, SNCA, SYNJ1, VPS13C, VPS35.*

2.3.3. Formes sporadiques

✓ Une forme atypique

Chez le patient 17S116 d'origine afro-caribéenne, une ataxie et une dysarthrie sont apparues à l'âge de 48 ans. Il présentait également des spasmes pharyngés et des troubles de la déglutition.

Après 10 ans d'évolution, à 58 ans, l'IRM cérébrale montre une atrophie cérébelleuse et médullaire ainsi qu'une atrophie corticale. Dans le même temps, l'ENMG met en évidence une polyneuropathie axonale sensitivo motrice aux 4 membres. Le patient souffre de syndrome dépressif et des troubles de l'humeur.

Après 19 ans d'évolution, à 67 ans, le tableau clinique associe : ataxie cérébelleuse, neuropathie axonale sensitivo-motrice chronique, fasciculations, amyotrophie des extrémités et syndrome frontal.

A 70 ans, la biopsie musculaire montre des Ragged Red Fibers et des amas mitochondriaux sous sarcolémniques en faveur d'une cytopathie mitochondriale. Mais l'analyse de la chaîne respiratoire ne rapporte pas de déficit d'activité des complexes. Et l'étude des grandes délétions de l'ADNmt est sans particularité.

L'examen des gènes impliqués dans les CA dominants par les méthodes de référence est non contributif, sauf une amplification de 35 CAG dans *ATXN1*. La dopa-sensibilité n'a pas été recherchée. Les variants pathogènes du gène *FRDA* ainsi que les délétions de l'ADNmt ont été exclus.

L'histoire familiale révèle que deux cousins germains sont atteints de pathologies neurodégénératives : l'un atteint de SLA et l'autre, de neuropathie amyloïde.

✓ 2 formes a priori classiques

Il s'agit de formes a priori classiques et sporadiques, dont les évolutions encouragent au diagnostic moléculaire.

Chez le cas 17S218, le symptôme inaugural a été une dépression à l'âge de 77 ans. Puis sont apparus des troubles cognitifs (désorientation), troubles praxiques, hypersalivation, faciès figé, hypotension orthostatique, dysarthrie, extinction de voix, dysphagie, trouble de la posture et de l'équilibre. On observait également une limitation verticale du mouvement des yeux, et des saccades rapides altérées.

Il existait également un important syndrome extrapyramidal : Akinésie faciale majeure. Un bavage, une marche rapide et ample mais sans balancement des bras qui restent semi fléchis. On notait également une attitude voutée, une rigidité axiale et segmentaire majeure, ainsi qu'un tremblement d'attitude.

Au total, trouble du comportement avec agitation et akathisie. La dopa-sensibilité n'avait pas été recherchée. Il n'y avait pas d'antécédents familiaux. Les variants pathogènes dans les gènes *HTT* et *JPH3* avaient été exclus.

Pour le cas 15S132 : il s'agit d'une femme qui a été vue pour la première fois à 79 ans. Elle présentait une incontinence urinaire, une instabilité à la marche, une instabilité avec rétropulsion, une hypoacousie gauche. Il y aurait probablement une épilepsie à crises partielles complexes. La patiente présentait par ailleurs un diabète non insulinodépendant.

L'IRM cérébrale montre une atrophie diffuse à prédominance temporale. Le bilan neuropsychologique réalisé à 80 ans rapporte un dysfonctionnement fronto-sous cortical avec un score MMS à 19/30.

A 85 ans, le tableau clinique était donc dominé par une agitation motrice et des mouvements involontaires. La dopa-sensibilité n'avait pas été recherchée. La patiente était retraitée de la fonction d'ouvrière agricole. Il y avait une histoire familiale de déficience avec pour certains une épilepsie. Les variants pathogènes dans les gènes *HTT* et *JPH3* avaient été exclus.

3. Analyse génomique de l'échantillon

Après application des critères de sélection et d'exclusion de l'étude, et selon la disponibilité des échantillons d'ADN, nous avons testé 13 cas par la méthode CoDE-seq. Un ADN d'individu asymptomatique a été testé dans le cadre d'une investigation comparative, avec son parent atteint.

Nous avons analysé les données de CoDE-seq, et interprété les variants SNVs et CNVs et parfois tandem repeats.

Conformément à la méthode de classification de variants proposée par l'ACMG, nous rapportons les variants de classe 3, 4 ou 5. Une fiche individuelle synthétique d'analyse est en annexe. Nous présentons ci-après la synthèse des résultats moléculaires obtenus.

3.1. Analyse qualité technique

Le nombre moyen de reads généré par échantillon est de 93 millions. La couverture effective après alignement atteint 91 à 95 Mb. La profondeur de séquençage est filtrée de façon automatique sur la base de plus de 5 reads. Toutes les données interprétées présentent une profondeur de séquençage au moins égale à 30X. L'analyse qualité des données montrent des taux de couverture et de profondeur de séquençage au moins égale à celle exigée dans la routine diagnostique.

3.2. Gènes candidats dans les phénotypes PARK/SPARK

Le TABLEAU 13, ci-après ce parapgraphe, résume les données génomiques obtenues de l'analyse des données CoDE-Seq des patients sélectionnés dans cette catégorie phénotypique.

3.2.1. *ABCA7* et *MAPT* dans une forme dominante

L'analyse CoDESeq du patient 09S079 révèle que le phénotype observé chez le patient ainsi que l'histoire familiale, pourraient être en rapport avec la conjonction de 2 variants probablement pathogènes :

1) Un variant STOP dans l'exon 9 du gène *ABCA7* c.853A>T p.Arg285*, probablement de classe 4 ou 5.

Il s'agit d'un variant très rare susceptible de provoquer un arrêt prématuré de la transcription du gène. *ABCA7* (ATP Binding Cassette Subfamily A Member 7) code pour une protéine de la superfamille des ATP-binding cassette (ABC) transporters, impliquées dans le transport de diverses molécules à travers les membranes intra- et extracellulaires.

2) Une délétion de 240 pb emportant les exons 7 et 8 du gène *MAPT* associée à un variant missense hétérozygote (c.689A>G, p.Gln230Arg) de classe 3 situé dans l'exon 6 de ce même gène.

MAPT code pour une proteine Tau associée au microtubule. Ce variant pourrait être un variant perte de fonction.

L'enquête familiale permettrait de vérifier si la ségrégation des variants mis en évidence confirme le caractère délétère de leur occurrence conjointe.

3.2.2. *GIGYF2* dans une forme récessive

L'analyse du cas 15S310 n'a pas permis de mettre en évidence de variants clairement pathogènes susceptibles d'être impliqués dans la pathologie. Un variant de signification incertaine de nature à interrompre le cadre de lecture a été mis en évidence dans l'exon 29 du gène *GIGYF2*. Un indel d'au moins 5 motifs CAG est trouvé : c.3626_3646delTGCCACAGCAGCAGCAGCAGC, p.Leu1209_Gln1215del.

Une étude de ségrégation familiale débutant par l'exploration de la sœur serait contributive. L'analyse des variants de type expansion de motifs répétés par l'outil GangSTR, confirme l'insertion de plusieurs motifs dont les répétitions de CAG. Une étude de la soeur serait également contributive.

3.2.3. *RARS2, NDUFAF1, ANO10, ATXN1*, ou *PLD3* insuffisants dans SPARK sporadique et atypique

Pour ce cas atypique 17S116, des variants délétères de classe 4, associés à l'ataxie spinocérébelleuse ont été mis en évidence. Mais aucun n'est suffisant à lui seul pour expliquer la maladie :

1) Le variant très rare missense hétérozygote c.422A>G, p.His141Arg de l'exon 6 du gène *RARS2*, est prédit délétère par tous les logiciels d'interprétation employés.

Ce gène nucléaire code pour une protéine qui se localise dans les mitochondries, où il catalyse le transfert de la L-arginine vers son ARNt apparenté, une étape importante dans la traduction des protéines codées par la mitochondrie. Les variants pathogènes de ce gène sont impliquées dans l'hypoplasie pontocérébelleuse de type 6 qui répond à un mode de transmission récessif et produit un phénotype plus sévère que le phénotype du patient.

2) Le variant très rare hétérozygote missense c.623T>C, p.Leu208Pro de l'exon 3 du gène *NDUFAF1* est également prédit délétère par tous les logiciels d'interprétation employés

NDUFAF1 code pour une protéine de facteur d'assemblage du complexe I mitochondrial et serait donc plus en lien avec la symptomatologie du patient. Mais il n'y avait pas de déficit de la chaîne respiratoire.

3) 2 variants de classe 3 ou 4 du gène *ANO10* : le variant hétérozygote missense, c.1133G>A, p.Arg378Gln de l'exon 6, probablement de classe 4 et le variant d'épissage c.1219-6delT de l'exon 7 de signification incertaine, classe 3.

ANO10 code pour une protéine transmembranaire qui appartient à la famille des anoctamines des canaux chlorure activés par le calcium.

L'analyse des CNVs révèle, une duplication interstitielle de 231 pb, comprenant intron 6 - exon 7 du gène *PLD3*. Ce gène est associé à SCA 46 mais il n'existe aucune description clinique en lien avec cette délétion interstitielle, il est probablement bénin.

Conformément au résultat de la filière « diagnostic », la méthode GangSTR confirme la présence d'expansion de motifs à la limite supérieure dans les gènes *ATXN1* et *3*, avec respectivement 35 et 32 répétitions CTG. Le screening des autres gènes de CA sujets aux expansions de motifs par cette méthode : *ATXN2, ATXN7, CACNA1A, KLHL1AS, ATXN10, PPP2R2B, TBP, ATXN1L, ATXN2L, ATN1, NOP56* et *BEAN*, est non contributif.

On relève seulement une amplification en zone normale haute de 25 CAG dans le gène *TBP.*

Pour aller plus loin dans l'interprétation des variants délétères chez ce patient, l'analyse des apparentés serait utile et par ailleurs, une analyse in vivo de l'effet du variant intronique de *ANO10* serait intéressante.

3.2.4. *LRRK2* et *ANKK1* dans les formes classiques

L'analyse de l'ADN du cas 17S218, révèle un variant hétérozygote missense rare, de classe 4 dans l'exon 34 du gène *LRRK2* : c.4937T>C, p.Met1646Thr.

LRRK2 fait partie de la famille des kinases riches en leucine répétés et code pour une protéine avec une région de répétition ankryine. La protéine est présente en grande partie dans le cytoplasme mais s'associe également à la membrane externe mitochondriale.

On trouve aussi 2 autres variants candidats de classe 3, missense hétérozygotes, dans les gènes :

1) c.4696G>C, p.Asp1566His exon 34 du gène *HLA-DRB1*, facteur de susceptibilité multifactoriel à l'atrophie multisystématisée.
2) c.239C>G, p.Thr80Arg exon 2 du gène *SORL1* impliqué dans la maladie d'Alzheimer de transmission dominante.

L'absence d'histoire familiale plaide en faveur de variant de novo qu'il conviendrait de pouvoir vérifier par une étude de ségrégation en testant les parents ou apparentés non atteints. Une analyse en trio serait contributive.

Pour le cas 15S132, l'analyse CoDE-Seq met en évidence un variant hétérozygote STOP c.1759G>T, p.Glu587* dans l'exon 8 du gène *ANKK1*. Ce variant est probablement pathogène, de classe 4. Le gène *ANKK1* code pour une Ser/Thr kinase étroitement liée au récepteur D2 de la dopamine (*DRD2*). L'arrêt prématuré de la transcription, induite par l'apparition d'un codon STOP est très fortement en faveur de l'implication du variant dans la symptomatologie.

Par ailleurs, on trouve également :

1) Un variant hétérozygote missense c.4229C>T, p.Thr1410Met dans l'exon 30 du gène *LRRK2.* Décrit dans la maladie de Parkinson 8 dominante, ce variant est rapporté avec des interprétations divergentes, tantôt de classe 3 ou 4.
2) Des variants de signification incertaine dans le gène *TRPM7*, de nature à décaler le cadre de lecture ou impactant le site d'épissage, susceptibles d'interagir dans la symptomatologie.

Enfin, on détecte une duplication Xq28 de 7Mb associée au retard mental lié à l'X type 72 (mrx72) avec ou sans syndrome parkinsonien. Cette duplication contient notamment le gène *RAB39B* dont les variants pathogènes sont associés à une déficience intellectuelle, épilepsie et parkinsonisme de révélation variable. Cette duplication du chromosome X est donc aussi susceptible de participer au phénotype.

L'histoire familiale orientait vers une épilepsie syndromique avec déficience intellectuelle pouvant être en rapport avec le CNV du chromosome X. Une étude de ségrégation familiale permettrait de confirmer l'implication des SNVs et des CNVs dans le phénotype.

N°	SEXE	AGE DIAGNOSTIC	AGE 1es SIGNES	DUREE EVOLUTION	ANTECEDENTS FAMILIAUX DE MND	TRANSMISSION	SYMPTOME INAUGURAL	ENMG	IRM cérébrale	NEUROPSY.	BIOPSIE MUSCULAIRE	Gènes déjà analysés	SNVs classe 4/5	CNVs MND
09S079	M	61	57	13	PARK	DOM	akinésie	NR	Sans modification significative (59ans)	Léger déficit fronto-sous-cortical (60ans)	NR	0	ABCA7 c.853A<T p.Arg285*	7q21.1del (240pb) exons 7 et 8 MAPT
15S310	M	63	62	6	PARK/ TREMB	REC	Trouble comportement	NR	Sans particularité (64ans)	Hallucinations visuelles, trouble mnésique, troubles sommeil depuis 60-62ans.	NR	ATP13A2, ATP1A3, DNAJC6, FBXO7, GCH1, LRRK2, PARK2, PARK7, PINK1, PLA2G6, SNCA, SYNJ1, VPS13C, VPS35	GIGYF2 c.3626_3646delTGCCACAG CAGCAGCAGCAGC p.Leu1209_Gln1215del	0
17S218	M	78	77	4	NON	SPO	Dépression	NR	Atrophie en rapport avec l'âge	Désorientation, mémoire	NR	HTT, JPH3	LRRK2 c.4937T>C p.Met1646Thr	0
15S132	F	79	78	9	NON	SPO	Céphalées bifrontales, rupture de contact	NR	Atrophie diffuse à prédominance temporale	Dysfonctionnement fronto-sous cortical, MMS 19/30	NR	HTT, JPH3	ANKK1 c.1759G>T, p.Glu587* LRRK2 c.4229C>T, p.Thr1410Met	Xq28dup (7Mb) loss of function 5'UTR-3'UTR RAB39B
17S116	M	50	48	27	SLA, NAF	SPO	Ataxie, dysarthrie	polyneuropathie axonale sensitivo motrice	Atrophie cérébelleuse et médullaire et atrophie corticale	Trouble de l'humeur, et atteinte cognitive	RRF/ chaîne respiratoire normale	ATXN1, 2, 3, CACNA1A, FRDA, ADNmt del	0	19q13.2dup (231pb) intron6-7 PLD3

TABLEAU 13. *Caractéristiques cliniques et moléculaires des PARK/SPARK testés*

FWI : French West Indies ; M : masculin/F : Féminin ; TREMB : tremblement ; REC : sœur atteinte/DOM : dominant : un parent atteint/SPO : sporadique ; RRF : Ragged Red Fibers.

3.3. Gènes candidats dans les Ataxies Cérébelleuses

Le TABLEAU 14, ci-après ce paragraphe, résume les données génomiques obtenues de l'analyse des données CoDE-Seq des patients sélectionnés dans cette catégorie phénotypique.

3.3.1. *SYNE1* et *CHMP1A* dans une forme sporadique

L'analyse CoDE-Seq du cas 10S103 effectuée en comparaison avec son fils 18S235 révèle plusieurs variants hétérozygotes dans le gène *SYNE1* dont deux de classe 4 et un de classe 3. Tous impliqués dans l'ataxie spinocérébelleuse de type 8 de transmission autosomique récessive et prédits délétères par plus de 3 logiciels.

Les variants c.11675T>C p.Leu3892Ser exon 71 et c.7345C>T, p.Leu2449Phe exon 49, de classe 4 sont de bons candidats. Le second est retrouvé chez le fils à l'état hétérozygote. On trouve également un variant c.6135T>G, p.Ile2045Met exon 42 de classe 3.

Chez cette patiente, on trouve une délétion de 631Kb en 16q24.3, contenant l'intron 3-4 du gène *CHMP1A* qui est associé à hypoplasie Ponto cérébelleuse de type 8 (OMIM 614961) de transmission autosomique récessive. Il s'agit d'un syndrome neurodéveloppemental sévère rendant peu probable l'implication de ce variant dans la symptomatologie.

3.3.2. polyAla et polyGln d'ATXN3 dans une forme dominante

L'analyse comparative des cas 18S231 et 17S162, apparentés au 1e degré, ne permet pas de mettre en évidence de variants non ambigus partagés par les 2 cas.

En revanche, on note des expansions multiples de motifs répétés dans les exons 3 et 6 du gène *ATXN3.* La méthode confirme ces variants :

- 17S162, polyAla c.209_232dup + c.407_430dup, polyGln c.892_915dup18S231
- 18S231, polyAla c.232_233ins + c.430_431ins, polyGln c.915_916ins

Les autres gènes : *ATXN1, ATXN2, ATXN7, ATXN10, CACNA1A, KLHL1AS, PPP2R2B, TBP, ATXN1L, ATXN2L, ATN1, NOP56* et *BEAN,* ne présentaient pas de particularités.

Ces tracts multiples pourraient être à l'origine du phénotype partagé par mère et fille.

On note également la présence d'un variant de classe 3 ou 4 dans le gène *SPTBN2* (c.5939C>T, p.Ala1980Val, exon28). Ce variant n'est retrouvé que chez la fille (17S162). Il s'agit d'un variant missense hétérozygote rare, impliqué dans l'ataxie spinocérébelleuse de type 5, de transmission dominante prédit pathogène par plus de 3 logiciels.

Des CNVs probablement bénins impliquants les gènes *PLD3* et *ATP2B3* ont été mis en évidence chez la patiente 18S231.

3.3.3. *CCDC88C* dans une forme atypique

Aucun variant de signification non ambigu n'est trouvé dans l'analyse de l'ADN du patient 14S121. On rapporte toutefois des variants hétérozygotes de signification incertaine (VUS, classe 3) dans le gène *GIGYF2*, facteur de susceptibilité à la maladie de Parkinson 11:

1) c.1331C>T, p.Pro444Leu exon 14, est un variant missense qui n'est pas très rare dans la population africaine : 0.4%.
2) c.3126_3128dupGCA, p.Gln1043dup exon 26, variant de nature à interrompre le cadre de lecture, sa fréquence n'est pas rapportée dans les populations africaines.

Mais ces variants ne peuvent expliquer à eux seuls le phénotype.

De même, on retrouve un VUS dans le gène *CCDC88C.* Il s'agit d'un variant hétérozygote intronique c.1050+6dupG, p. ? de l'exon10. Il est extrêmement rare dans la population africaine et est impliqué dans l'ataxie spinocérébelleuse 40.

Son impact fonctionnel pourrait être testé in sillico avec un outil du genre HSF_Pro analysis.

Origine	N°	SEXE	AGE DIAGNOSTIC	AGE 1es SIGNES	DUREE EVOLUTION	ANTECEDENTS FAMILIAUX DE MND	TRANSMISSION	SYMPTOME INAUGURAL	DOPA-SENSIBILITE	ENMG	IRM cérébrale	NEUROPSY.	BIOPSIE MUSCULAIRE	Gènes déjà analysés	SNVs classe 4/5	CNVs MND
FWI	17S162	F	65	60	10	SCA	DOM	Ataxie à 60ans	NR	NR	Atrophie cérébelleuse modérée	NR	NR	0	ATXN3 *polyAla* c.209_232dup + c.407_430dup, *polyGln* c.892_915dup	0
FWI	18S231	F	65	60	26	SCA	DOM	Ataxie à 60ans	NR	NR	Non disponible	NR	NR	ATXN1, ATXN2, ATXN3, CACNA1A, ATXN7, TBP	ATXN3 *polyAla* c.232_233ins + c.430_431ins, *polyGln* c.915_916ins	19q13.2dup (31Kb) PLD3 Xq28dup (7,6Mb) ATP2B3 XL
FWI	10S103	F	47	28	28	SCA	SPO	Ataxie à 28 ans	NR	atteinte myélinique sensitive discrète. Et atteinte sensitive mixte des nerfs suraux.	Atrophie cérébelleuse isolée majeure	NR	Discrètes anomalies mitochondriales , rapport lactate/pyruvate normal	FRDA, POLG, ARNt Leu, NTTK, TWNK, SDHA	SYNE1 c.11675T>C p.Leu3892Ser + c.7345C>T p.Leu2449Phe	16q24.3del (631Kb) intron 3-4 CHMP1A
FWI	14S121	F	45	43	14	NON	SPO	Trouble psychiatrique	NON		atrophie vermis et cérébelleuse	Hallucinations	NR	NPC1, NPC2, FMR1, ATXN3	CCDC88C c.1050+6dupG p. ?	0

TABLEAU 14 *des caractéristiques cliniques et moléculaires des CA testés*

FWI : French West Indies ; M : masculin/F : Féminin ; NR : non-recherché ; DOM : dominant, un parent atteint/SPO : sporadique.

3.4. Gènes candidats dans la SLA

Le TABLEAU 15 résume les données génomiques obtenues de l'analyse des données CoDE-Seq des patients sélectionnés dans cette catégorie phénotypique.

3.4.1. CNVs dans la SLA atypique

L'étude de l'ADN 17S353 ne révèle aucun variant clairement pathogène pouvant expliquer le phénotype.

On relève cependant la présence de VUS hétérozygotes de nature à interrompre le cadre de lecture :

1) Dans l'exon 29 du gène *TRPM7* (c.4500_4502delTAG, p.Ser1500del).

Une étude de l'impact de ce variant permettrait de conclure à son caractère délétère.

2) Dans l'exon 4 du gène *NEFH* (c.1973_1978delAAGAGG, p.Glu658_Glu659del), facteur de susceptibilité de l'ALS.

Leurs fréquences ne plaident pas réellement en faveur de variants causaux (0,005-7, GnomAD).

Par ailleurs, on relève la présence de petites délétions emportant tout ou partie des gènes *ALS2*, *VCP* et *FANCG*. Les variants pathogènes de ces gènes sont responsables de phénotype SLA +/- atteintes cognitives, syndromiques ou juvénile.

3.4.2. *SOD1/TIAM1*, *DCTN1*, *NEFH*, *UBQLN2* et *C21orf2* dans la SLA avec anticipation

Les patients 13S001, 10S030 et 16S073 sont des individus d'une même famille. Ils sont atteints de SLA génétique. Une variation délétère p.Gly94Val du gène *SOD1* est responsable de la pathologie familiale, où au moins 9 cas sont déjà connus.

L'analyse CoDESeq révèle que les patients de la 2e génération ont au moins un 2nd variant probablement pathogène dans un autre gène de susceptibilité à la SLA.

Chez le patient 13S001, on trouve un variant hétérozygote missense dans l'exon 27 du gène *DCTN1* : c.3517C>G, p.Pro1173Ala, variant rare prédit pathogène.

Chez le patient 10S030, on retrouve 2 variants missense dans l'exon 4 du gène *NEFH* : c.1941G>C, p.Lys647Asn et c.2601G>T, p.Lys867Asn. Le premier dix fois moins rare que le second. Ils restent des variants peu fréquents dans la population de descendance africaine et prédits pathogènes par plus de 3 logiciels.

Chez le plus jeune patient touché, on ne retrouve aucun variant de signification non ambigüe associé au variant *SOD1*. En revanche, on retrouve :

1) Une délétion de 3.2 Mb en Xp11.21 incluant le gène *UBQLN2* dont les variants pathogènes sont impliqués dans la SLA de type 15 avec ou sans démence.
2) Une duplication de 14,858Kb en 21q23.1, incluant le gène *C21ORF2*, dont certaines isoformes, à domaine leucine-riche, sont mitochondriales.

Par ailleurs, par la méthode Genfilter on détecte deux variants communs aux 3 cas dans le gène *TIAM1*. Il s'agit de variants missense hétérozygotes très rares de classe 3 : c.2464C>T p.Pro822Ser de l'exon 13 et c.3882C>T c.3882C>T de l'exon 24.

N°	SEXE	AGE DIAGNOSTIC	AGE 1es SIGNES	DUREE EVOLUTION	ANTECEDENTS FAMILIAUX DE MND	TRANSMISSION	SYMPTOME INAUGURAL	ENMG	IRM cérébrale	Gènes déjà analysés	SNVs classe 4/5	CNVs MND
17S353	F	70	68	6	SLA	Pénétrance variable	Dysphonie-dysarthie	Activité spontanée tonique discontinue	Normale / Scintigraphie cérébrale en faveur d'une apraxie progressive (69ans)	SOD1, C9ORF72, ALS2, ANG, CHMP2B, DCTN1, FIG4, FUS, OPTN, SETX, SPG20, SQSTM1, TARDBP, TBK1, UBQLN2, VAPB, VCP	0	2q33.1del (67kb) loss of function, ALS2 9p13.3del (20kb) loss of function VCP/FANCG
13S001	M	58	54	12	SLA	DOM	Déficit moteur, pseudopolynevretique	NR	NR	SOD1, C9ORF72, SCA2	SOD1, c.281G>T, p.Gly94Val DCTN1, c.3517C>G; p.Pro1173Ala TIAM1, c.2464C>T p.Pro822Ser + c.3882C>T c.3882C>T	0
10S030	M	61	61	13	SLA	DOM	Fasciculations, crampes, atrophies distale, motrice et bulbaire	NR	NR	SOD1	SOD1, c.281G>T, p.Gly94Val NEFH, c.1941G>C; p.Lys647Asn TIAM1 , c.2464C>T p.Pro822Ser + c.3882C>T c.3882C>T	0
16S073	M	13	13	SLA	DOM	Fasciculation, motrice	Fasciculations Motrice des membres inférieurs	Atteinte neuropathique modérée	NR	SOD1, C9ORF72, SCA2	SOD1, c.281G>T, p.Gly94Val TIAM1 , c.2464C>T p.Pro822Ser + c.3882C>T c.3882C>T	Xp11.21del (3,4 Mb) UBQLN2 21q22.3dup (14,858Kb) C21orf2

TABLEAU 15 *des caractéristiques cliniques et moléculaires des SLA testés*

FWI : French West Indies ; M : masculin/F : Féminin ; NR : non-recherché ; DOM : dominant, un parent atteint/SPO : sporadique

4. Données démographiques et variants génomiques nouveaux

Dans un premier temps, une collection de données cliniques, biologiques et moléculaires des MNDREM vus au CERCA a été réalisée. Elle a permis de dresser la cartographie actuelle des MNDREM. On distingue principalement 5 phénotypes de représentation démographique décroissant comme suit : syndromes parkinsoniens et maladie de parkinson (SPARK), les ataxies spinocérébelleuses (CA), la sclérose latérale amyotrophiques (SLA), les maladies de Huntington et Huntington-like (HDs) et la paraparésie spastique héréditaire (PPSH).

L'analyse générale des données d'une population d'environ 309 cas de MNDREM, montre des âges de début moyen de 61 ans 2 mois, avec un écart-type moyen de 21 ans 9 mois, quel que soit le sexe. Mais ce paramètre est probablement biaisé par les pathologies de transmission dominante avec anticipation. On trouve que 40% des patients de la population de MNDREM présentent une histoire familiale de MND. Parmi lesquels 56% ont une variation pathogène identifiée dans un gène. En fait, seulement 23% des phénotypes sont étiquetés sur le plan moléculaire. Comme les modes de transmission, la rentabilité du test génétique varie selon les pathologies. On trouve un variant causal dans 100% des HD, 53% des CA, 13% des SLA, 20% des PPSH, et aucun des SPARK.

L'étude démographique des MNDREM génétique montre que :

- Dans HD, le sexe ratio est de 2. On trouve 50% de patients atteints de HD et 50% atteints de HDL1, qui présentent globalement des symptomatologies très similaires. La cohorte de HDL2 représente à elle seule une dizaine d'individus répartis sur une seule grande famille martiniquaise.
- Seul le gène *SOD1* est impliqué dans la SLA. Deux variants pathogènes sont exclusivement retrouvés : p.Gly94Val et p.Glu121Gly. Le 1e est retrouvé chez 6 patients sur 7 appartenant tous à une grande famille martiniquaise dans laquelle un phénomène d'anticipation a été statistiquement prouvé et nouvellement associé à ce variant (PUBLICATION EN COURS CI-APRES).
- Dans la PPSH, c'est l'association d'un variant délétère (c.1099-7_1102del11insATAA) et un VUS (c.865C<T) dans le gène *SPST*, qui est responsable du phénotype.

- Dans les CA, se basant sur la classification selon les modes de transmission, on trouve :
    1. 77% dominant, associés aux gènes *ATXN1, ATXN2, ATXN3* et *CACNA1A* ; le phénotype SCA2 causé par les variants pathogènes *ATXN2*, comptant pour 86% de ces formes.
    2. 15% récessif, associés aux gènes *ATM, FRDA* et *POLG*
    3. 8% mitochondrial, associés aux gènes *ARNt Leu* et *MT-CO1*

Et nous avons pu observer un patient ataxique dans le cadre d'une maladie de Steinert infantile.

La stratégie de sélection des sujets à étudier par la méthode CoDE-Seq choisie pour ces travaux, est basée sur la connaissance de l'histoire naturelle des phénotypes et la probabilité de trouver une variation délétère. Cette probabilité est d'autant plus importante que les signes apparaissent de façon précoce, qu'il existe une histoire familiale, ou que les symptômes sont considérés comme atypiques. Par notre stratégie, nous avons pu investiguer treize patients, dont un apparenté non atteint.

L'analyse qualité des données montrent des taux de couverture et de profondeur de séquençage au moins égale à celle exigée dans la routine diagnostique. Avec, en moyenne, une profondeur de séquençage de 30X et une couvertuure de 91 à 95Mb.

Il n'a pas toujours été possible de trouver des variants de classe 5, formellement pathogène. Un variant probablement pathogène a été mis en évidence chez 7 patients sur 13. Nous avons détecté des SNVs et des CNVs impliquant les gènes : *ABCA7, GIGYF2, LRRK2, SORL1, SYNE1, CCDC88C, TRPM7, DCTN1, TIAM1, NEFH, et SOD1*.

La nouveauté qu'apporte ce travail dans l'utilisation du NGS dans les MNDREM, est la possibilité de détecter des CNVs. Nous proposons des CNVs comme pouvant être impliqués dans certaines symptomatologies, en particulier :

- La délétion interstitielle des exons 7 et 8 du gène *MAPT* associé à un variant probablement pathogène STOP dans le gène *ABCA7* dans une maladie de parkinson de transmission dominante ;
- La délétion perte de fonction des gènes *ALS2*, *VCP* et *FANCG* dans cette SLA de pénétrance variable ;
- La duplication *C21orf2* associée à une délétion qui contient *UBQLN2* chez le plus jeune cas de SLA familiale avec anticipation.

Les résultats de l'étude génomique des patients nous permettent de compléter le panel de gènes en cause dans les MNDREM vues au centre de référence. Ce panel constitue les premières données génomiques spécifiques à notre population. Mais également une base à la modélisation des réseaux protéiques impliqués dans les processus neurodégénératifs de ces pathologies.

Dans le chapitre suivant, nous proposons en effet, la construction de réseaux d'interaction au moyen des outils bioinformatiques, afin prioritairement, de soutenir l'hypothèse de voies de signalisation communes aux différents gènes de MNDREM.

Article en voie d'acceptation dans **ALS-DFT journal**...

SOD1-related ALS with anticipation in a large family from Martinique

Anna-Gaelle Giguet-Valard[1], Rémi Bellance[1], Séverine Jeannin[1], Sophie Duclos[1], Pascale Olive[1], Oriane Allard-Saint-Albin[1], Cécile Cazeneuve[2] OU Fabienne Clot[2], Pittion-Vouyovitch Sophie[3], Juliette Smith-Ravin[4], Cyril Goizet[5]

Affiliations

1: University Hospital Center of Martinique. Neurological and neuromuscular rare disorders department, CERCA.

2 : University Hospital Center of Paris. La Pitié Salpétrière Hospital.

3: University Hospital of Nancy, Hospital of Brabois. Genetic Department.

4 : University of Antilles, campus Schoelcher. Research group BIOSPHERES/AREBIO. 59, rue Ernest Desproges 97200 FDF.

5 : National Reference Center for Rare Diseases 'Neurogenetic' Department of Medical Genetics, Pellegrin Hospital, Bordeaux University Hospital; Rare Diseases Laboratory: Genetics and Metabolism (MRGM), INSERM U1211, Bordeaux University, Bordeaux, France.

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a rare neurological disorder cause degeneration of upper and lower motor neurons and their axons. ALS is mostly sporadic, but there are familial forms. In more than half of the familial forms, a pathogenic variant is found in one of the following genes: *C9ORF72*, *SOD1*, *TDP-43*, *FUS* and *VCP*. *SOD1* is the 2nd most common gene involved in genetic forms of ALS. Genotype-phenotype relationships are occasionally established in genetically forms of ALS associated with *SOD1* mutations pathogenic variants. The p.Gly94Ser variant in *SOD1* is associated with a rarely described and unexplained anticipation phenomenon.

We report a large family from Martinique in whom ALS is associated with a p.Gly94Val pathogenic variant in *SOD1* and a statistically proven anticipation. A whole-exome study and detection of CNVs (CoDESeq) from 3 affected members of this family revealed the presence of probably pathogenic variants in other ALS genes. Variants in DCTN1 and *NEFH* were present in patients of the 2nd generation, and CNVs involving *UBQLN2* and *C21orf2* were found in the youngest case of the family.

Introduction

Amyotrophic Lateral Sclerosis (ALS) is a rare neurological disorder characterized by progressive muscle wasting and weakness due to degeneration of lower motor neurons (LMN) and their axons, and loss of upper motor neurons and their corticospinal axonal tracts [1] ALS causes progressive paralysis of the entire striated musculature of the limbs, trunk (including the respiratory muscles) and the cephalic extremity. ALS affects both sexes and may be familial (FALS) or sporadic (SALS). FALS is used if a positive family history for either ALS or frontotemporal dementia (FTD) within three generations is documented. Hereditary ALS (HALS) is considered if at least one first or second degree relative suffers from ALS and/or FTD. The term genetic ALS (GALS) is reserved to situations in which a pathogenic variant in a disease-causing gene is found segregating in a family. [1] Although the vast majority of cases are sporadic, substantial gain has been made in the knowledge of the genetic causes of the disease, especially in HALS. [2] Pathogenic variants in five genes, *C9ORF72*, *SOD1*, *TDP-43*, *FUS* and *VCP*, account together for more than 50% of FALS cases. Age of onset of FALS is statistically earlier than sporadic forms: 46 years on average for FALS and 56 years on average for sporadic ALS. Variability in age of onset in FALS is known since 1995 and several studies tried to establish genotype-phenotype correlations [3]. Clear correlations and prognostic trends can be drawn only for few pathogenic variants, particularly in *SOD1*.

SOD1 is the second most frequently reported responsible gene in FALS after *C9ORF72* [4], [5]. *SOD1* pathogenic variants can be identified in 12 to 20% of FALS cases and in about 1-2% of apparently sporadic cases [4], [6]. To date, over 150 different pathogenic variants have been reported in *SOD1*, primarily consisting of missense pathogenic variants, with a smaller number of nonsense and deletion/insertion pathogenic variants [7]. Most *SOD1* pathogenic variants have autosomal dominant (AD) transmission with high-penetrance [8]. The most

common SOD1 variant in North America, p.Ala4Val, is responsible for an aggressive form of ALS, in which death occurs usually within a year after symptoms onset [5]. Other variants consistently associated with a poor survival are p.Gly41Ser, p.Gly94Ala (previously reported as p.Gly93Ala), and p.Arg115Gly. Homozygous p.Asp90Ala is associated with prolonged survival of over 10 years and may present with sensory involvement [9]. In 2002, Iwai et al reported the first description of a *SOD1* p.Gly94Ser pathogenic variant apparently associated with anticipation which corresponds to severity aggravation and earlier onset of symptoms through generations [10]. Thereafter, statistically proved anticipation has been described in several families with GALS caused by p.Leu84Phe, p.Leu84Val and p.Gly94Ser in *SOD1* respectively in Japanese, American and Italian families. In the family harboring the p.Gly94Ser mutation, early occurrence of death was observed in the last generations of affected relatives [9], [11]. Importantly, anticipation has already been published exclusively in *SOD1*-related GALS but never explained [3], [12], [13].

We report the first GALS Martinican family caused by a p.Gly94Val pathogenic variant in *SOD1* in whom apparent anticipation was statistically confirmed. We then developed for the first time in Martinique Next Generation Sequencing (NGS) capacities to search for additional molecular characteristics that may act as modifier genes or di/oligogenic inheritance in order to explain this anticipation. Whole Exome Sequencing (WES) is classically considered inadequate for detection of CNVs or expansions of repeated sequence. However, several studies have already shown that selective bioinformatics analyses applied to clinically or genetically orientated genetic regions may allow identification of CNVs [14], [15]. For this study, we selected a method that associates WES with detection of Copy Number Variants (CNV). CoDE-Seq combines the high-quality screening of DNA from the coding regions of the 20,000 human genes, with the sequencing of hundreds of thousands of 120 nucleotides probes that regularly cover the whole human genome for accurate analysis of

gene CNV [15]. We essentially focused our analyses on Single Nucleotide Variants (SNVs) and Copy Number Variants (CNVs) in genes already implicated in determinism of ALS.

METHODS

Patients and clinical presentation

A large family with many relatives who displayed ALS was followed in the Caribbean Reference Center for Rare Neurological and Neuromuscular Diseases (CERCA) in Martinique, since early 1990s (Figure 1). Altogether, fifteen relatives were affected. The oldest affected generation (generation II) had not been medically evaluated, although most of affected relatives of the 3rd generation have been followed (generation III). We collected detailed clinical data of eleven relatives; eight of them were directly followed in CERCA.

SOD1 genetic analyses

Written informed consent was obtained for all genetically tested patients. Genetic testing was performed in accordance with French legislation regarding genetic diagnostics tests (French bioethics law 2004-800). Patients' DNA was extracted from leukocyte pellets using conventional procedures. Complete routine molecular screening of the *SOD1* gene, which consisted in Sanger sequencing of the 5 exons, was performed in the index case as previously described [8]. The pathogenic variant was then searched in four relatives (Figure 1).

Statistical analysis of anticipation

We used the Mann-Whitney test to compare, age of onset, survival delay and age at death among generations III and IV (Figure 1). Remaining that the first generation (II) has not been medically evaluated and the last generation (V) comprises only one individual.

Targeted genes investigation

It has been shown that the size of the CAG repeat in normal alleles of *ATXN2* represents a risk factor for ALS [16]. In order to exclude nucleotide repeats amplification variants in known genes implicated in anticipation, we explored *C9ORF72* and *SCA2* genes by routinely PCR and Sanger methods.

Copy Number Variation Detection and Exome Sequencing (CoDE-Seq)

Patient-derived material was obtained through procedures approved by the Institutional Review Board (approval 2020/079). For this part of the study, DNA samples were available for only three out of those five patients (Figure 1).

Next generation sequencing (NGS) experiments were subcontracted to LIGAN© to benefit from CoDE-Seq method. The whole exome capture was performed with NimbleGenSeqCap EZ Enrichment combined with homemade design from NimbleGenSeqCap EZ Choice XL. This specific capture consisted of a backbone of one hundred thousand probes of 120 nucleotids regularly spaced throughout the entire genome. Probes were placed specifically every 10kb apart in the genomic regions known to be affected by CNVs (based on the literature and Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources, DECIPHER [17], which encompassed 881Mb (i.e about 30% of the human genome). For the about 70% remaining, probes were placed every 25kb. The probes targeting whole exome and backbone covered 65Mb of the genome. The effective coverage after NGS reached 91%. The enrichment preparation was carried out according to the manufacturer's protocol NimbleGenSeqCap EZ.

Bioinformatics analyses

The demultiplexing of sequence data was performed initially by LIGAN©, using bcl2fastq Conversion Software (Illumina; version 2.19.1). Raw data analyses were performed in

BIOSPHERES laboratory and with the support of Bourgogne and Lyon 1 universities. A home-made Bioinformatic pipeline (VarAP version 1.16b) was used to implement the following steps: 1. Sequence reads from FASTQ files were mapped to the human genome (hg38/GRCh38) using Burrows-Wheeler Aligner (version 0.7.15-r1140). On average, 93M reads were generated per sample; 2. Variant calling was performed using Genome Analysis ToolKit (version 3.8-0-ge9d806836); 3. the annotation of points variants including missense variants, nonsense variants, small insertions and deletions, splice-site variants was performed using snpEff version 4.3t 2017-11-24.

To support pathogenicity we used the genome aggregation database GnomAD browser especially data for afro population and several in silico prediction software: PROVEAN, PolyPhen-2, SIFT, MutationTaster, MutationAssessor, MetaSVM, LRT, CADD, FATHMM, GERP, PhyloP (100 & 20), and PhastCons (100 & 20). With a homemade program, we added a supplementary annotation concerning panel of genes tagged in NCBI gene database with neurodegenerative Disorders, NDs (387 genes) or ALS (436 genes) (appendix to this document). Detection of CNVs was performed using 2 methods: eXome Hidden Markov Model program (XHMM) version 1.0 [18], exomedepth v1.1.15 [19]. Results were filtered by cnvscan v0.2.0 and annotations have been done with AnnotSV v2.2 [20], OMIM (http://www.ncbi.nlm.nih.gov/omim/) [21] and Decipher databases. Data obtained by XHMM were crossed with the gene panels in order to target only the regions containing genes of interest.

Biological filtration methods

We performed analyses on individual data file and common variants file. We used standards and guidelines of the American College of Medical Genetics and Genomics (ACMG) for interpretation of rare point variants.

RESULTS

Clinical presentation

The pedigree showed clear autosomal dominant inheritance of ALS without FTD expression (Figure 1). Onset was predominantly associated with motor expression in lower limbs (5/9), except for 3 cases in whom pseudopolynevretic (2/9) and upper limbs signs (2/9) were first observed (Table 1). Initial symptoms were mixed bulbar and motor symptoms in one case. Among initials signs, cramps were frequent (4/9), as muscle atrophy (4/9), fasciculations (3/9) and muscular weakness (3/9). All documented patients shared classical clinical signs of ALS: cramps, fasciculations and muscle atrophy. Interestingly, survival delay and age at onset of the disease decreases were observed during transmission through generations and clinical course impaired more rapidly in younger generations.

Analysis of *SOD1, C9ORF72* and *ATXN2*

Direct sequencing of *SOD1* in the index patient led to identify a p.Gly94Val pathogenic variant which co-segregated with the disease in four relatives (Figure 1), leading to the diagnosis of GALS. The p.Gly94Val variant has never been described although other pathogenic variants at the same position are known (p.Gly94Arg, p.Gly94Cys, p.Gly94Ala). The p.Gly94Val substitution impacts a conserved amino acid and is predicted to be deleterious by all software used.

There were no intermediate repeat expansion in *ATXN2* nor pathogenic expansion in *C9ORF72* in the three studied cases.

Statistical analysis of anticipation

Mann-Whitney test was applied to compare age of onset, survival delay and age of death among the more informative generations III and IV. The required fifth value that has been use

for calculation is the mean of the four values available in family cases. The test revealed statistical differences in age of onset (p=0, 03) and age of death (p=0, 01), but not in survival delay (p=0, 05).

Individual CoDE-Seq results

All clinical and molecular data are summarized in Table 1.

Case III.2 (13S001) presented ALS at age 54 years. The survival delay was about 144 months. The occurrence of the disease was pseudopolynevretic, and first signs were motor deficit. We found a heterozygote missense variant in protein coding *DCTN1* gene (rs184147813) which is very rare in afro population (0,0001%) located in exon 27 (c.3517C>G; p.Pro1173Ala). This variant is predicted to be pathogenic by more than 3 softwares, and associated with autosomal recessive and dominant susceptibility to ALS and autosomal dominant distal hereditary motor neuropathy type VIIB. [22], [23]; According to ACMG, this variant is of uncertain significance (VUS) class IV. No CNV involved in ALS were identified by the two variants callers used.

Case III.6 (10S030) had an age of onset at age 61 years and survival delay was about 108 months. Initial symptoms were motor and bulbar. First signs were fasciculation, cramps and distal atrophy. We found two probably pathogenic heterozygous variants in exon 4 of *NEFH* predicted to be pathogenic by more than 3 softwares: one missense c.1941G>C; p.Lys647Asn variant (rs200634512), frequency of 0,001% in African population; and one missense c.2601G>T; p.Lys867Asn variant (rs138156220), frequency of 0,0005% in African population. This gene is associated with autosomal recessive or dominant susceptibility to ALS.[24] No CNV were identified by the two variants callers used.

Case V.1 (16S073) was followed since age 13 years when fasciculations of the tongue appeared. Needle electromyogram showed chronic neurogenic changes and fibrillation potentials without ongoing denervation in the studied muscles. This individual is still alive and followed once a year. To date, there is 228 months of evolution. We didn't find any non-ambiguous pathogenic SNVs supplementary to the *SOD1* variant. We found two CNVs containing genes which have previously been associated with ALS:

- A large hemizygote deletion of 3.2 Mb, ranging from position 43711851 to 77620567 on Xp11.21 including *UBQLN2* of which pathogenic variants are properly associated with ALS/ALS15 with or without FTD.[25] Pathogenic variants leading to proline substitutions in the PXX repeat region which is unique to *UBQLN2*, have functional consequences with an impairment of protein degradation
- A duplication of 14,858Kb on 21q22.3, containing *PFKL* and *C21orf2* genes. PFKL is not properly associated with any human disease, but *C21orf2* encodes for four alternatively spliced transcript variants producing four different isoforms containing leucine-rich repeats, three of these isoforms are mitochondrial proteins. *C21orf2* belongs to novel genes associated with susceptibility to ALS through genome wide association studies (GWAS) [26].

DISCUSSION

We described herein a large family of GALS caused by a novel p.Gly94Val pathogenic variant in *SOD1*. Clinical presentation was generally classical and unremarkable among relatives except apparent anticipation through disease transmission over successive generations. Very interestingly, statistical Mann-Whitney analysis confirmed statistical differences on age of onset and age of death over two successive generations, demonstrating true anticipation phenomenon. There was no difference concerning shorter survival delay for the 4th generation compared to the 3rd.

Anticipation is generally associated with dynamic mutations linked to germinal unstable amplification of nucleotides repeats in coding and non-coding regions of the genome [27]. However, our observations reinforce clinical anticipation evidences in GALS associated with variants affecting the p.Gly94 residue in *SOD1*, as previously described in Korean families [10]. In order to test the hypothesis of ascertainment bias to explain apparent anticipation, Vucic S et al. performed a statistical study on 112 individuals [13]. Clinical history was collected with the age of death and disease duration determined in successive generations (generation 1, grandparent; generation 2, parents; generation 3, children), from five large *SOD1*-related ALS families including 34 *SOD1*-related patients. Positive intergenerational differences were evident in 92% of parent-offspring transmissions in the *SOD1*-related FALS cohort. Three *SOD1* variants segregated in the five reported families (Val148Gly, Ile113Thr and Glu100Gly) Although no mutations were found in the known ALS genes including *TARDBP, FUS, OPTN, VCP, UBQLN2, ANG, FIG4, DCTN1* and *CHMP2B*. They concluded that anticipation should be a real phenomenon in FALS, possibly due to co-inheritance of modifier genes [13].

NGS made it possible to very quickly discover new genes involved in ALS. It is now evident that ALS is associated with a complex genetic determinism. More and more additional mutations are found by WES methods applied in SALS or FALS. There are many studies in favor of oligogenism and polygenism in SALS as in GALS [26], [28]. CNVs are also currently more studied in ALS and seem to play a role in pathogenesis. Blauw et al. identified CNVs predisposing to SALS and showed higher number of genes deleted specifically in ALS patients that strongly suggests that multiple rare deletions might have an important role in ALS pathogenesis [29].

In the family described here, analyses of *C9ORF72* and *ATXN2* did not bring elements to explain anticipation as those genes were normal in our three studied patients. However results

of the CoDE-Seq method applied for the first time in three relatives of a *SOD1*-related ALS family may argue for potential involvement of oligogenism to explain observed anticipation. Indeed, VUS class 4 were found only in two relatives of the generation III: one in *DCTN1* (rs184147813) in patient III.3, and two in *NEFH* (rs200634512 and rs138156220) in patient III.6. Moreover, we found two CNVs also previously associated with ALS susceptibility in GWAS in our younger case (patient V.1): one deletion of 3.2 Mb in Xp11.21 including *UBQLN2* and one duplication of intermediate length in 21q22.3 including *C21orf2*. Synergic effects of these different VUS with the familial inherited *SOD1* pathogenic variant may be hypothesized as all of them are located in genes previously implicated in ALS determinism, at least as susceptibility genetic factors in large GWAS [26]. Our results indicate the importance of a WES to search for supplementary genetic arguments helping to understand familial expression variability. Further studies are needed to validate our hypotheses of potential oligogenism explaining anticipation in *SOD1*-related GALS.

Acknowledgements

The authors thank the UMR 8199 LIGAN-MP Genomics platform (Lille, France) which belongs to the 'Federation de Recherche' 3508 Labex EGID (European Genomics Institute for Diabetes; ANR-10-LABX-46) and was supported by the ANR Equipex 2010 session (ANR-10-EQPX-07-01; 'LIGAN-MP'). The LIGAN-PM Genomics platform (Lille, France) is also supported by the FEDER and the Region des Hauts de France.

Declaration of interest statement

The authors declare they have no conflicts of interest to disclose

Figure 1.Pedigree of the G94V *SOD1* related family

Circle symbol for women; square symbol for men; filled symbols for patients; arrow for index patient. "+": SOD1 G94V mutation identified by routine analysis; underlined number indicate CoDE-Seq tested cases.

Above symbols, Roman numeral associated with alphanumeric number, indicate identification of family member.

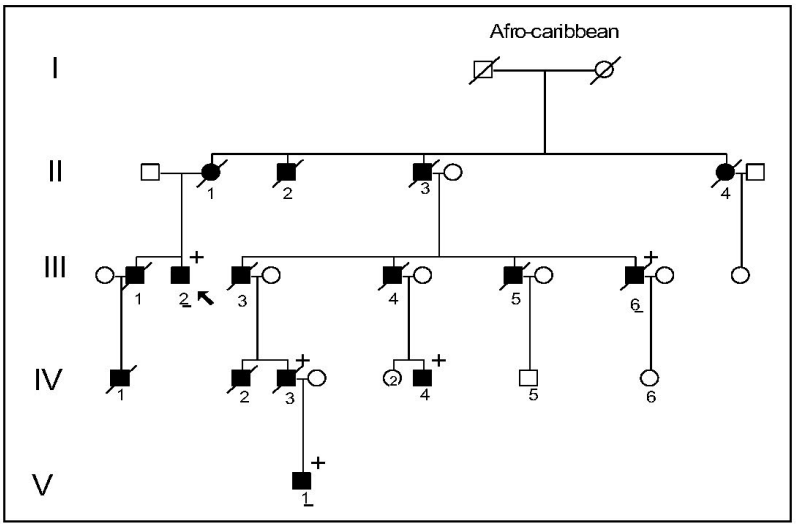


Table 1. Clinical features
Id.: Identification of individual; M, sex masculine; NP, non-precise; LL, Lower Limb; UL, Upper Limb; ND, not done

Identification			CLINICAL CHARACTERISTICS						GENOMICS DATA	
Id.	**sex**	**age of onset**	**Survival delay (months)**	**age of death**	**ALS TYPE of occurrence**	**Initial Symptoms**	**Initial Signs**	**ENMG**	**SNVs**	**CNVs**
III.1	M	46	108	55	Motor expression (LL)	Motor	cramps, distal atrophy	certain		
III.2	M	54	144	66	pseudopolynevretic	Motor	deficit	ND	*SOD1*, class 5 (c.281G>T, p.G94V) D*CTN1*, class 4 *(rs184147813)* TIAM1, class 4 (rs200667365; rs?)	No
III.3	M	50	72	53	Motor expression (LL)	Motor	distal atrophy, fasciculations	Certain		
III.4	M	NP	NP	50	NP	NP	NP	NP		
III.5	M	NP	NP	46	NP	NP	NP	NP		
III.6	M	61	108	70	Motor expression (LL)	Motor, bulbar	fasciculations, cramps, distal atrophy	ND	*SOD1*, class 5 (c.281G>T, p.G94V) *NEFH, class 4 (rs200634512;rs138156220)* TIAM1, class 4 (rs200667365; rs?)	
IV.1	M	30	72	36	pseudopolynevretic	Motor	deficit, cramps	Certain		
IV.2	M	23	72	26	Motor expression (LL)	Motor	distal cramps	Certain		
IV.3	M	37	NP	NP	Motor expression (UL)	Motor	Atrophy, deficit	Certain	*SOD1* class 5 *(*c.281G>T, p.G94V)	
IV.4	M	33	60	38	Motor expression (UL)	Motor	fasciculations, cramps, distal atrophy	certain	*SOD1* class 5 *(*c.281G>T, p.G94V)	
V.1	M	13	228	alive	Motor expression (LL)	Motor	fasciculations	ND	*SOD1*, class 5 (c.281G>T, p.G94V) TIAM1, class 4 (rs200667365; rs?)	DELXp11.21-3.4 Mb: *UBQLN2* DUP*21q22.3*-14,858Kb*: C21orf2*

REFERENCES

[1] A. Ludolph *et al.*, « A revision of the El Escorial criteria - 2015 », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 16, n° 5-6, p. 291-292, août 2015, doi: 10.3109/21678421.2015.1049183.

[2] S. Morgan et R. W. Orrell, « Pathogenesis of amyotrophic lateral sclerosis », *Br. Med. Bull.*, vol. 119, n° 1, p. 87-98, sept. 2016, doi: 10.1093/bmb/ldw026.

[3] M. Aoki *et al.*, « Variance of age at onset in a Japanese family with amyotrophic lateral sclerosis associated with a novel cu/zn superoxide dismutase mutation », *Ann. Neurol.*, vol. 37, n° 5, p. 676-679, mai 1995, doi: 10.1002/ana.410370518.

[4] S. Millecamps *et al.*, « SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations », *J. Med. Genet.*, vol. 47, n° 8, p. 554-560, août 2010, doi: 10.1136/jmg.2010.077180.

[5] P. V. S. de Souza, W. B. V. de R. Pinto, M. A. T. Chieia, et A. S. B. Oliveira, « Clinical and genetic basis of familial amyotrophic lateral sclerosis », *Arq. Neuropsiquiatr.*, vol. 73, n° 12, p. 1026-1037, déc. 2015, doi: 10.1590/0004-282X20150161.

[6] A. Chio *et al.*, « Prevalence of SOD1 mutations in the Italian ALS population », *Neurology*, vol. 70, n° 7, p. 533-537, févr. 2008, doi: 10.1212/01.wnl.0000299187.90432.3f.

[7] O. Abel, J. F. Powell, P. M. Andersen, et A. Al-Chalabi, « ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics », *Hum. Mutat.*, vol. 33, n° 9, p. 1345-1351, sept. 2012, doi: 10.1002/humu.22157.

[8] D. R. Rosen *et al.*, « Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis », *Nature*, vol. 362, n° 6415, p. 59-62, mars 1993, doi: 10.1038/362059a0.

[9] R. W. Orrell *et al.*, « Clinical characteristics of SOD1 gene mutations in UK families with ALS », *J. Neurol. Sci.*, vol. 169, n° 1-2, p. 56-60, oct. 1999, doi: 10.1016/S0022-510X(99)00216-6.

[10] K. Iwai, « Anticipation in familial amyotrophic lateral sclerosis with SOD1-G93S mutation », *J. Neurol. Neurosurg. Psychiatry*, vol. 72, n° 6, p. 819-820, juin 2002, doi: 10.1136/jnnp.72.6.819.

[11] W. Kim, J.-S. Kim, K.-S. Lee, Y.-J. Gwoun, J.-M. Kim, et K.-H. Lee, « Anticipation and Phenotypic Heterogeneity in Korean Familial Amyotrophic Lateral Sclerosis with Superoxide Dismutase 1 Gene Mutation », *J. Clin. Neurol. Seoul Korea*, vol. 3, n° 1, p. 38-44, mars 2007, doi: 10.3988/jcn.2007.3.1.38.

[12] H. Zhang *et al.*, « A rare Cu/Zn superoxide dismutase mutation causing familial amyotrophic lateral sclerosis with variable age of onset and incomplete penetrance in China », *Amyotroph. Lateral Scler.*, vol. 6, n° 4, p. 234-238, janv. 2005, doi: 10.1080/14660820510044478.

[13] S. Vucic, G. A. Nicholson, A. Chio, et M. C. Kiernan, « Apparent anticipation in SOD1 familial amyotrophic lateral sclerosis », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 14, n° 5-6, p. 452-456, sept. 2013, doi: 10.3109/21678421.2013.764569.

[14] C. Marelli *et al.*, « Mini-Exome Coupled to Read-Depth Based Copy Number Variation Analysis in Patients with Inherited Ataxias: HUMAN MUTATION », *Hum. Mutat.*, vol. 37, n° 12, p. 1340-1353, déc. 2016, doi: 10.1002/humu.23063.

[15] L. Montagne *et al.*, « CoDE-seq, an augmented whole-exome sequencing, enables the accurate detection of CNVs and mutations in Mendelian obesity and intellectual disability », *Mol. Metab.*, vol. 13, p. 1-9, juill. 2018, doi: 10.1016/j.molmet.2018.05.005.

[16] A. Chio *et al.*, « ATXN2 polyQ intermediate repeats are a modifier of ALS survival », *Neurology*, vol. 84, n° 3, p. 251-258, janv. 2015, doi: 10.1212/WNL.0000000000001159.

[17] H. V. Firth *et al.*, « DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources », *Am. J. Hum. Genet.*, vol. 84, n° 4, p. 524-533, avr. 2009, doi: 10.1016/j.ajhg.2009.03.010.

[18] M. Fromer et S. M. Purcell, « Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data », *Curr. Protoc. Hum. Genet.*, vol. 81, n° 1, avr. 2014, doi: 10.1002/0471142905.hg0723s81.

[19] V. Plagnol *et al.*, « A robust model for read count data in exome sequencing experiments and implications for copy number variant calling », *Bioinformatics*, vol. 28, n° 21, p. 2747-2754, nov. 2012, doi: 10.1093/bioinformatics/bts526.

[20] V. Geoffroy *et al.*, « AnnotSV: an integrated tool for structural variations annotation », *Bioinformatics*, vol. 34, n° 20, p. 3572-3574, oct. 2018, doi: 10.1093/bioinformatics/bty304.

[21] V. A. McKusick, « Mendelian Inheritance in Man and Its Online Version, OMIM », *Am. J. Hum. Genet.*, vol. 80, n° 4, p. 588-604, avr. 2007, doi: 10.1086/514346.

[22] C. Munch *et al.*, « Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS », *Neurology*, vol. 63, n° 4, p. 724-726, août 2004, doi: 10.1212/01.WNL.0000134608.83927.B1.

[23] P. Caroppo *et al.*, « *DCTN1* Mutation Analysis in Families With Progressive Supranuclear Palsy–Like Phenotypes », *JAMA Neurol.*, vol. 71, n° 2, p. 208, févr. 2014, doi: 10.1001/jamaneurol.2013.5100.

[24] D. A. Figlewicz *et al.*, « Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis », *Hum. Mol. Genet.*, vol. 3, n° 10, p. 1757-1761, 1994, doi: 10.1093/hmg/3.10.1757.

[25] H.-X. Deng *et al.*, « Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia », *Nature*, vol. 477, n° 7363, p. 211-215, sept. 2011, doi: 10.1038/nature10353.

[26] R. Chia, A. Chiò, et B. J. Traynor, « Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications », *Lancet Neurol.*, vol. 17, n° 1, p. 94-102, janv. 2018, doi: 10.1016/S1474-4422(17)30401-5.

[27] A.-S. Lebre et A. Brice, « Maladies par expansion de polyglutamine : données moléculaires et physiopathologiques », *médecine/sciences*, vol. 17, n° 11, p. 1149-1157, nov. 2001, doi: 10.1051/medsci/200117111149.

[28] H. Zhang *et al.*, « Screening for possible oligogenic pathogenesis in Chinese sporadic ALS patients », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 19, n° 5-6, p. 419-425, juill. 2018, doi: 10.1080/21678421.2018.1432659.

[29] H. M. Blauw *et al.*, « Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen », *Lancet Neurol.*, vol. 7, n° 4, p. 319-326, avr. 2008, doi: 10.1016/S1474-4422(08)70048-6.

SUPPLEMENTARY DATA

List of tagged NDs genes extracted from NCBI gene database the 01th jul 2019 (387 genes)

https://www.ncbi.nlm.nih.gov/gene/?term=(neurodegenerative)+AND+%22Homo+sapiens%22%5Bporgn%3A__txid9606%5D

List of tagged ALS genes extracted from NCBI gene database the 01th jul 2019 (436 genes)

https://www.ncbi.nlm.nih.gov/gene/?term=(amyotrophic+lateral+sclerosis)+AND+%22Homo+sapiens%22%5Bporgn%3A__txid9606%5D

CHAPITRE IV

Étude computationnelle préliminaire des interactions géniques dans les MNDREM

CHAPITRE IV – Etude computationnelle préliminaire des interactions géniques

Dans ce chapitre, nous proposons une méthode bioinformatique d'identification d'interactions géniques dans les MNDREM pour répondre à la question suivante : les gènes de MNDREM de notre population sont-ils en interaction ?

Les prédictions d'interactions entre protéines (IPPs) sont basées sur un ensemble de paramètres biologiques d'interactions géniques. Après avoir rappelé le cadre théorique de la détection et la prédiction d'IPPs, nous présenterons l'étude modèle des gènes rapportés dans notre population, puis l'analyse des IPPs impliquant JPH3 en particulier.

1. Rappels sur les Interactions Protéine-Protéine (IPPs)

Les processus biologiques sont le fruit de nombreuses et diverses interactions protéines-protéines (IPPs). Quand on parle d'interactions protéiques et de processus biologiques, il est indispensable de ne pas perdre de vue que 2 protéines peuvent avoir une « interaction fonctionnelle », sans avoir d'interaction « physique-directe ».

1.1. Définitions

Des décennies de recherche en biologie cellulaire, moléculaire, biochimie etc., ont permis de documenter les propriétés et fonctions moléculaires de nombreuses protéines. Des centres de ressources biologiques donnent accès à des bases de données (BDD) de structure, fonction et autres propriétés des protéines issues de diverses organismes [120].

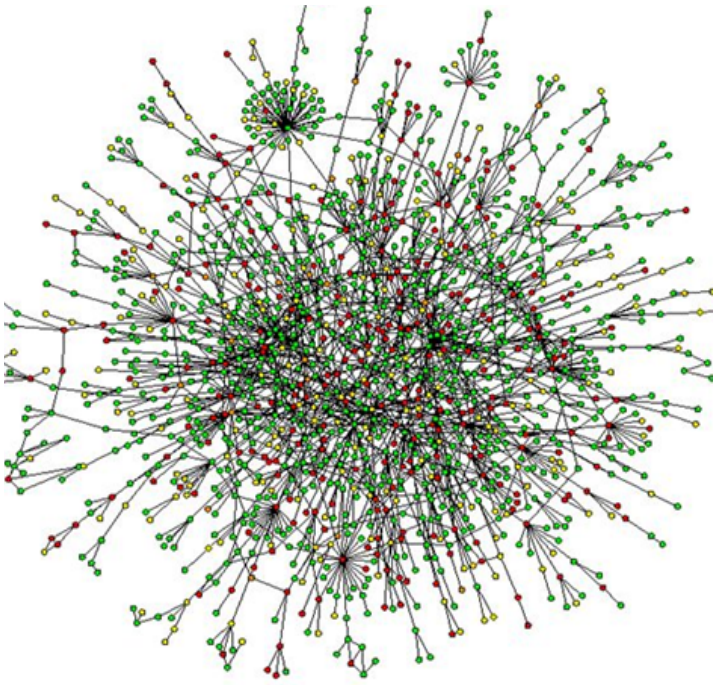


FIGURE 41 – *Représentation schématique des IPPs : les sphères sont les protéines, les arrêtes les intéractions.*

Grâce aux avancées récentes et rapides dans les technologies à haut débit, un nombre important de données sur l'interaction protéine-protéine de divers organismes sont disponibles et actuellement stockées dans plusieurs bases de données. Ainsi, la carte complète des interactions protéiques intervenant dans le développement d'un individu s'appelle « interactome ».

Les IPPs sont modélisées sous forme de nuage, on parle de réseau (FIGURE 41).

En règle générale, la modélisation respecte les mêmes règles : des nœuds (ou sommets) représentent les protéines, et des cordes (ou liens/arêtes) symbolisent les interactions. Parfois, on confère un graphisme particulier aux cordes selon la nature des interactions en question.

Il existe diverses BDD d'IPPs. Elles collectent aussi bien les données de détection expérimentales que les données prédictives. Certaines sont populaires et régulièrement utilisés dans des études publiées, parmi lesquels BioGrid, string, ou encore Apid. Ces ressources informatiques offrent en plus la possibilité de modéliser et analyser des réseaux d'IPPs d'intérêt.

1.2. Détection expérimentale

Alors que la quantité de données annotant la séquence génomique augmente de façon exponentielle, l'annotation des séquences de protéines semble prendre du retard, tant en termes de qualité que de quantité. Des approches génomiques fonctionnelles multifonctions à haut débit sont nécessaires pour combler l'écart entre l'information sur les séquences brutes et les informations biochimiques, physiologiques et médicales pertinentes.

Pour déterminer si 2 protéines sont en interaction physique, il existe plusieurs expérimentations :

- Spectroscopie de résonance magnétique nucléaire,
- Transfert d'énergie de résonance de fluorescence,
- Résonance de plasmon de surface,
- Microscopie à force atomique et microscopie électronique à haut débit qui font suite aux expériences double-hybrides avec la levure (Y2H) et la purification par affinité, réalisées in vivo ou in vitro. [121]

Dans le principe de modélisation, 2 protéines en interaction directe sont connectées par une corde. C'est l'interconnexion physique ou physiologique de proche en proche, qui permet de dessiner la modélisation en réseau.

1.3. Prédiction d'IPPs

Pour découvrir des interactions qui ne sont pas accessibles aux méthodes à haut débit des méthodes de calcul sont nécessaires. Ces calculs de prédictions peuvent ensuite être vérifiés en utilisant des méthodes expérimentales. Un certain nombre d'approches informatiques pour la découverte d'interactions protéiques ont été développées ces dernières années. Ces méthodes diffèrent dans les informations sur les caractéristiques utilisées pour la prédiction de l'interaction des protéines.

L'afférence d'une interaction est basée sur des arguments de vraisemblance mais aussi sur des méthodes statistiques de prédiction [121]–[124].

1.3.1. Arguments de vraisemblance

- ✓ **Les gènes voisins**

L'idée principale est que les gènes apparentés sont situés près les uns des autres dans le génome. Cette méthode de prédiction est donc basée sur le contexte génomique.

L'inconvénient de cette méthode est que, d'une part, elle ne parvient pas à reconnaître l'interaction entre des gènes liés mais distants, d'autre part, le choix des génomes de référence peut affecter la performance de la méthode.

- ✓ **La relation phylogénique**

Cette méthode est une version flexible de la méthode des gènes voisins qui peut détecter certaines interactions que la méthode des gènes voisins ne parvient pas à détecter. L'idée fondamentale est que les gènes liés fonctionnellement restent ensemble dans de nombreuses espèces éloignées pour jouer un rôle dans un processus biologique. L'interaction est donc afférée en fonction de la similitude phylogénétique.

Trois inconvénients à cette méthode :

1) Le fait que le nombre et la distribution des génomes utilisés peuvent influencer les résultats de manière spectaculaire.
2) Il ne peut pas être appliqué aux protéines exclusives qui se présentent dans presque tous les organismes.
3) Ne peut fonctionner que sur les génomes complets.

✓ **La fusion de gène**

Cette méthode complémentaire des méthodes de gènes voisins et relation phylogénétique a un avantage majeur : sa fiabilité. En effet, les événements de fusion de gènes existants sont très informatifs sur les relations fonctionnelles.

L'un des inconvénients de cette méthode est que l'événement de fusion n'est pas répandu, en particulier chez les procaryotes.

✓ **La comparaison de structure tridimensionnelle**

Prédit l'interaction protéine-protéine selon la séquence conservée du polypeptide court, comme les signatures ou la similarité des séquences et le nombre de k-let (sous-séquences avec la longueur k)

1.3.2. Arguments statistiques : analyse de modélisation de réseau

La modélisation des IPPs sous forme de réseau, offre en elle-même une puissance d'analyse permettant d'argumenter en faveur ou non des IPPs en question. Comme dans les réseaux du monde réel, les réseaux d'interactions protéines-protéines partagent des fonctions topologiques qui rendent ces réseaux différents des réseaux aléatoires.

Les caractéristiques topologiques sont utilisées comme preuve pour discerner les interactions qui représentent des vrais positifs et celles qui sont faussement positives. Ces caractéristiques permettent d'attribuer un score de confiance plus sensible à chaque interaction.

Les éléments topologiques qui définissent un réseau sont les suivants :

- Degré d'interaction : nombre de connections,
- Répartition des interactions/Hub : nœuds à haut degré d'interaction,
- k-core : sous-graphe du réseau où chaque sommet à un degré > k-1,
- Densité du lien : rapport entre le nombre d'arêtes du réseau et le nombre maximal possible de liens qui peuvent être calculés selon une équation mathématique,
- Coefficient de clustering,
- Longueur du chemin d'accès (d'un sommet à un autre du réseau),
- Longueur moyenne du chemin (sur tout le réseau),
- Diamètre : longueur maximale du chemin dans le réseau,
- Centralité : montre l'importance des nœuds et des bords du réseau, plusieurs mesures d'évaluation sont possibles,
- Motif : sous-graphique du réseau dont les occurrences sont significativement élevées.

Des outils informatiques comme *NetworkAnalyser*, réalisent l'analyse topologique automatisée des réseaux.

Des calculs basés sur les valeurs de ces paramètres permettent de discriminer les interactions en fonction de comparaison de scores mathématiques [122].

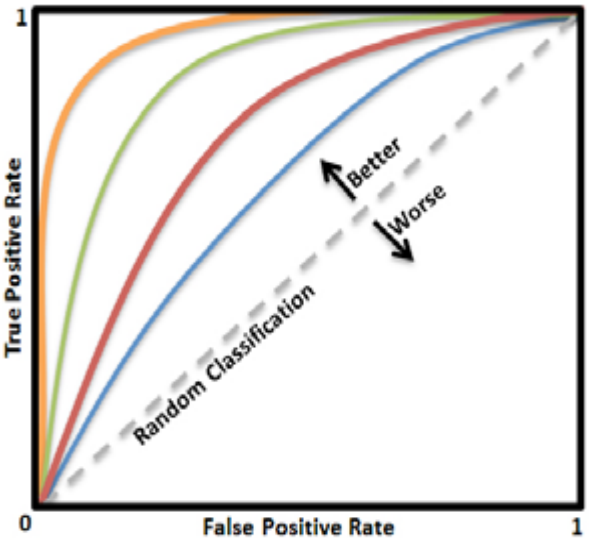


FIGURE 42 – Représentation de la courbe de ROC

En plus de ces mesures, un outil graphique populaire permet d'évaluer la performance de classification, c'est la courbe de ROC (receiver operating characteristic) qui reproduit la sensibilité (taux positif vrai) par rapport à la spécificité (taux de négation réelle), variant entre 0 et 1. (FIGURE 42)

Plus la courbe suit la bordure gauche et ensuite la bordure supérieure de l'espace ROC, plus la classification est sensible et spécifique. La zone sous la courbe ROC (ASC), est une autre mesure de la précision, plus l'ASC est proche de 1, plus la classification est sensible. L'ASC est une mesure de performance fiable pour les problèmes levés par la prédiction des PPIs.

Il existe de nombreux outils pour évaluer et visualiser les performances des classifications, comme *Cytoscape*, *Médusa* ou encore *NAViGaTOR.*

1.3.3. Apports de l'Intelligence Artificielle

- ✓ **Données de la littérature (text or littérature mining)**

Pour détecter des IPPs, l'analyse des données de la littérature demeure un puissant outil pouvant être véritablement fastidieux si elle doit être faite manuellement. L'automatisation de l'analyse des données extraites de la littérature est rendue possible par le développement de systèmes informatiques d'indexation des données : Named Entity Recognition or NER step, Computational natural language processing (*NLP*) et linguistics-based methods. Ces résultats automatisés de l'exploration de données peuvent ne pas être aussi fiables que les données organisées manuellement, mais la croissance rapide de la littérature biomédicale publiée peut rendre ces méthodes plus fiables.

- ✓ **Analyse de fonctions génomiques / protéomiques hétérogènes, par des algorithmes de machine learning**

Support vector machines (*SVM*), Artificial neural networks, Naïve Bayes, K-Nearest neighbors, Decision tree, Randomforest sont autant de méthodes d'intelligence artificielle permettant d'afférer une interaction entre 2 ou plusieurs protéines selon des variables définies.

Actuellement, données expérimentales et données de prédiction assistée par ordinateur, sont étroitement liées pour répondre à la question des IPPs, comme le montre le schéma de la FIGURE 43 qui suit [124].

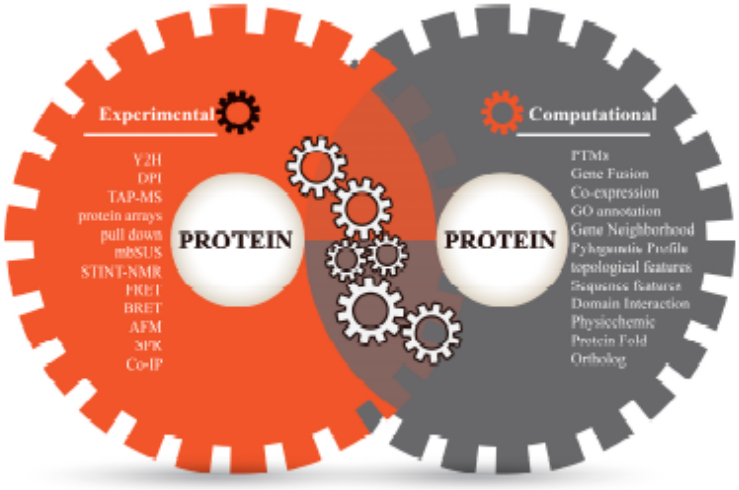


FIGURE 43 – Différentes méthodes de détection des IPPs tiré de la référence [5]

2. Modélisation et analyse de réseaux

Tout l'enjeu de l'utilisation des bases de données et des outils de prédiction d'IPPs est de détecter des partenaires protéiques physiques et/ou fonctionnels et de construire un réseau d'interaction, est de pouvoir discriminer les interactions faussement positives.

2.1. Outils de détection et de modélisation

Il existe 3 types de BDD d'IPPS, selon la méthode d'afférence d'interaction. [125], [126]. On distingue :

2.1.1. Bases de données d'interaction afférée par l'expérimentation

Ce sont les bases type *BioGRID* (Biological General Repository for Interaction Datasets). Des outils de curation et collection des données issues de la littérature permettent d'alimenter ces bases.

2.1.2. Métabases de données

Colligent les données expérimentales tirées de la littérature, issues de plusieurs BDD. *APID* (Agile Protein Interactomes DataServer) est une métabase populaire [127].

2.1.3. Bases de données de prédiction

Colligeant les données des bases précédentes mais aussi les données dites « non IPPs », c'est à dire, les IPPs afférées sur d'autres arguments que l'expérience : relation phylogénique, conservation de la co-expression au travers les espèces, contexte génomique et extraction automatique des données de la littérature. *STRING* est une référence dans ce type de BDD.[128]

GeneMANIA trouve d'autres gènes liés à un ensemble de gènes, en utilisant un très grand ensemble de données d'association fonctionnelle comme : les interactions protéiques et génétiques, les voies de signalisation, la co-expression, la colocalisation et la similarité des domaines protéiques. On utilise *GeneMANIA* pour trouver de nouveaux membres d'une voie ou d'un complexe, trouver de nouveaux gènes à fonction spécifique, comme les protéines kinases [129].

2.1.4. Outils de modélisation

Ces bases de données s'accompagnent d'outils de modélisation et de visualisation des IPPs en nuage de points liés par des cordes. Il existe d'autres outils informatiques dédiés à la modélisation, comme *Cytoscape* [130].

2.2. Paramètres de base

La performance de prédiction doit être évaluée par différentes mesures. Quatre paramètres de base sont importants pour l'analyse:

- TP (True Positive ou hit): le nombre d'interactions vraies prédit correctement.
- TN (True négatif ou véritable négatif): le nombre de non interactions prédites correctement.
- FP (False Positive ou erreur de type I): le nombre de non-interactions prédites de manière incorrecte en tant qu'interaction.
- FN (False Negative ou erreur de type II): le nombre d'interactions prédites de manière incorrecte en tant que non-interaction.

2.3. Élimination des Faux Positifs

Il existe principalement 2 méthodes d'élimination des faux positifs. Là encore, expérimentale ou statistique :

2.3.1. Bases de données d'expérimentation

Il existe des BDD générées sur la base d'expérimentation excluant des interactions entre 2 protéines. On parle de « négatome ». Les BDD type *Negatome* [131], recensent les non-interactions protéiques.

On peut aussi utiliser des méthodes d'analyse de vraisemblance type mesure de similarité sémantique, ou localisation tissulaire, ou système d'annotation par *Gene Ontology* [132] par exemple, etc…

Gene Ontology est un projet bioinformatique destiné à structurer la description des gènes et des produits géniques dans le cadre d'une ontologie commune à toutes les espèces.

2.3.2. Les outils statistiques d'exclusion d'interaction

On recommande d'utiliser des méthodes statistiques tels que l'exclusion de cluster d'interaction mutuelle/simultanée entre protéines d'un réseau [133].

3. Étude computationnelle

3.1. Rationnel de l'étude

Les MNDREM sont un ensemble de pathologies programmées aboutissant à la dégénérescence plus ou moins précoce et plus ou moins rapide de tout ou partie du SNC. Elles s'expriment par une atteinte motrice et ou une atteinte cognitive dont la caractérisation permet d'orienter l'étiologie. La montée en puissance de moyens d'étude biochimique, histologique et moléculaire, ont permis d'améliorer continuellement les classifications des MNDREM.

Il existe une forte hétérogénéité génétique dans les MNDREM. Certains gènes sont impliqués dans plusieurs phénotypes. On constate donc des chevauchements troublants aussi bien cliniques que moléculaires.

Les variants pathogènes par expansions de motifs répétés sont très représentés dans l'étiologie moléculaire des MNDREM. Des hypothèses physiopathologiques rapportent notamment que ces expansions gain de fonction créeraient des perturbations mitochondriales et des perturbations du système ubiquitine-protéasome, à l'origine d'inclusion cytoplasmique ou de dépôts de protéines mal repliées et non éliminées par les mécanismes cellulaires, parfois caractéristiques des certaines MNDREM.

Dans le champ des maladies neurodégénératives, les études de prédiction d'IPPs sont peu répandues. On retrouve cependant quelques études du genre, sur les maladies à expansions de motifs répétés.

L'étude de l'interaction protéines-protéines du set de gènes impliqués dans les MNDREM permettrait d'évaluer les voies de recrutement de ces gènes. Les résultats d'une étude comme celle que nous proposons pourraient permettre de détecter des partenaires protéiques candidats à des études in vivo sur l'impact de variants dans les gènes impliqués.

La limite de cette stratégie est l'hypothèse des variants gain de fonction délétère, qui par définition, entrainent le recrutement inadapté donc imprévisible et non prédictible de molécules. Dans le cadre d'étude in vitro, l'analyse biochimique d'IPPs permettrait de pallier cette limite.

3.2. Design

3.2.1. Identification d'IPPs dans le MNDREM d'intérêt

- ✓ **Set de gènes**

L'étude moléculaire de notre population a permis de montrer l'implication des gènes listés ci-après, dans l'expression de MNDREM (TABLEAU 16).

GENES	PROTEINES	FONCTIONS
FXN	Frataxin (210 AA)	Protéine mitochondriale appartenant à la famille FRATAXIN. Intervient dans la régulation du transport et de la respiration mitochondriale du fer Favorise la biosynthèse de l'hème, l'assemblage et la réparation des amas de fer-soufre en fournissant Fe (2+) aux protéines impliquées dans ces voies. Peut jouer un rôle dans la protection contre le stress oxydatif. In vitro, la forme oligomérique a une activité ferroxidase. L'expansion de la répétition trinucléotidique intronique GAA de 8 à 33 répétitions à> 90 répétitions entraîne une ataxie de Friedreich. L'épissage alternatif entraîne plusieurs isoformes.
ATXN2	Ataxin-2 (1313 AA)	Impliqué dans le trafic d'EGFR, agissant comme régulateur négatif de l'endocytose d'EGFR au niveau de la membrane plasmique
GIGYF2	GRB10-interacting GYF protein 2 (1320 AA)	Composant clé du complexe 4EHP-GYF2, un complexe multiprotéique qui agit comme un répresseur de l'initiation de la traduction. 4EHP-GYF2 agit comme un facteur qui relie EIF4E2 à ZFP36 / TTP, reliant la répression de la traduction à la désintégration de l'ARNm. Peut agir en coopération avec GRB10 pour réguler la signalisation des récepteurs de la tyrosine kinase, y compris les récepteurs de l'IGF1 et de l'insuline. Appartient à la famille GIGYF.
DCTN1	Dynactin subunit 1 (1278 AA)	Joue un rôle clé dans le transport rétrograde médié par la dynéine des vésicules et des organites le long des microtubules en recrutant et en attachant la dynéine aux microtubules. Se lie à la fois à la dynéine et aux microtubules, fournissant un lien entre des cargaisons spécifiques, des microtubules et de la dynéine. Essentiel pour cibler la dynéine sur le microtubule et les extrémités, recruter la dynéine vers les cargaisons membraneuses et améliorer la processivité de la dynéine (la capacité de se déplacer le long d'un microtubule sur une longue distance sans tomber de la piste). Peut également agir comme un frein pour ralentir le moteur de la dynéine pendant la motilité le long du microtubule.
ATXN3	Ataxin-3 (361 AA)	Enzyme de désubiquitination impliquée dans le maintien de l'homéostasie des protéines, la transcription, la régulation du cytosquelette, la myogenèse et la dégradation des substrats chaperons mal repliés. Lie de longues chaînes de polyubiquitine et les coupe, alors qu'elle a une activité faible ou nulle contre les chaînes de 4 ubiquitines ou moins. Impliqué dans la dégradation des substrats chaperons mal repliés
SOD1	Superoxide dismutase [Cu-Zn] (154 AA)	Détruit les radicaux qui sont normalement produits dans les cellules et qui sont toxiques pour les systèmes biologiques
POLG	DNA polymerase subunit gamma-1 (1239 AA)	Impliqué dans la réplication de l'ADN mitochondrial. S'associe à l'ADN mitochondrial; Appartient à la famille de l'ADN polymérase de type A
ABCA7	ATP-binding cassette sub-family A member 7 (2146 AA)	Joue un rôle dans la phagocytose par les macrophages des cellules apoptotiques. Se lie à APOA1 et peut fonctionner dans l'efflux de phospholipides médié par l'apolipoprotéine à partir des cellules. Peut également médier l'efflux de cholestérol. Peut réguler l'homéostasie cellulaire des céramides lors de la différenciation des kératinocytes.
ATM	Serine-protein kinase ATM (3056 AA)	Sérine / thréonine protéine kinase qui active la signalisation des points de contrôle lors des cassures double brin (DSB), de l'apoptose et des stress génotoxiques tels que la lumière ultraviolette A ionisante (UVA), agissant ainsi comme un capteur de dommages à l'ADN. Reconnaît la séquence consensus du substrat [ST] -Q. Phosphoryle 'Ser-139' du variant d'histone H2AX / H2AFX aux cassures double brin (DSB), régulant ainsi le mécanisme de réponse aux dommages à l'ADN.
JPH3	Junctophilin-3 (748 AA)	Les junctophilines contribuent à la formation de complexes membranaires jonctionnels (JMC) qui relient la membrane plasmique au réticulum endoplasmique ou sarcoplasmique dans les cellules excitables. JPH3 est spécifique au cerveau et semble avoir un rôle actif dans certains neurones impliqués dans la coordination motrice et la mémoire.
DMPK	Myotonin-protein kinase (639 AA)	Protéine kinase qui est nécessaire pour le maintien de la structure et de la fonction des muscles squelettiques. Peut jouer un rôle dans la différenciation et la survie des myocytes en régulant l'intégrité de l'enveloppe nucléaire et l'expression de gènes spécifiques du muscle. Peut également phosphoryler PPP1R12A et inhiber l'activité de la myosine phosphatase pour réguler la phosphorylation de la myosine. Également essentiel à la modulation de la contractilité cardiaque et au maintien d'une bonne activité de conduction cardiaque.
CCDC88C	Protein Daple (2028 AA)	Régulateur négatif de la voie de signalisation canonique Wnt, agissant en aval de DVL pour inhiber la stabilisation de CTNNB1 / bêta-caténine. Peut également activer la voie de signalisation JNK; Appartient à la famille CCDC88.
SORL1	Sortilin-related receptor (2214 AA)	Probablement un récepteur endocytaire multifonctionnel, qui pourrait être impliqué dans l'absorption des lipoprotéines et des protéases. Lie le LDL, la principale lipoprotéine plasmatique transportant le cholestérol, et le transporte dans les cellules par endocytose. Lie la protéine associée au récepteur (RAP). Pourrait jouer un rôle dans l'interaction cellule-cellule. Impliqué dans le trafic APP vers et depuis l'appareil Golgi. Il agit probablement comme un récepteur de tri qui protège l'APP du trafic vers l'endosome tardif et de la transformation en bêta amyloïde, réduisant ainsi la charge de peptide amyloïdogène.
TIAM1	T-lymphoma invasion and metastasis-inducing	Module l'activité des protéines de type RHO et relie les signaux extracellulaires aux activités du cytosquelette. Agit comme une protéine stimulatrice de dissociation du PIB qui stimule l'activité

	protein 1 (1591 AA)	d'échange GDP-GTP des GTPases de type RHO et les active. Active RAC1, CDC42 et dans une moindre mesure RHOA.
HTT	Huntingtin (3142 AA)	Peut jouer un rôle dans le transport médié par les microtubules ou la fonction vésiculaire; Appartient à la famille huntingtine.
LRRK2	Leucine-rich repeat serine/threonine-protein kinase 2 (2527 AA)	Régule positivement l'autophagie par une activation dépendante du calcium de la voie de signalisation CaMKK / AMPK. Le processus implique l'activation des récepteurs de l'acide nicotinique adénine dinucléotide phosphate (NAADP), l'augmentation du pH lysosomal et la libération de calcium par les lysosomes. Avec RAB29, joue un rôle dans la voie de trafic rétrograde pour le recyclage des protéines, telles que le récepteur du phosphate de mannose 6 (M6PR), entre les lysosomes et l'appareil de Golgi d'une manière dépendante du rétomère.
SPAST	Spastin (616 AA)	Régule positivement l'autophagie par une activation dépendante du calcium de la voie de signalisation CaMKK / AMPK. Le processus implique l'activation des récepteurs de l'acide nicotinique adénine dinucléotide phosphate (NAADP), l'augmentation du pH lysosomal et la libération de calcium par les lysosomes. Avec RAB29, joue un rôle dans la voie de trafic rétrograde pour le recyclage des protéines, telles que le récepteur du phosphate de mannose 6 (M6PR), entre les lysosomes et l'appareil de Golgi d'une manière dépendante du rétomère.
SYNE1	Nesprin-1 (8797 AA)	Protéine modulaire multi-isomère qui forme un réseau de liaison entre les organites et le cytosquelette d'actine pour maintenir l'organisation spatiale subcellulaire. En tant que composant du complexe LINC (LInker of Nucleoskeleton and Cytoskeleton) impliqué dans la connexion entre la lame nucléaire et le cytosquelette. Les interactions nucléocytoplasmiques établies par le complexe LINC jouent un rôle important dans la transmission des forces mécaniques à travers l'enveloppe nucléaire et dans le mouvement et le positionnement nucléaires. Peut être impliqué dans l'attachement noyau-centrososme et la migration nucléaire.
ATXN1	Ataxin-1 (815 AA)	Facteur de liaison à la chromatine qui réprime la signalisation Notch en l'absence de domaine intracellulaire Notch en agissant comme un corépresseur CBF1. Se lie au promoteur HEY et pourrait aider, avec NCOR2, répression médiée par la RBPJ. Lie l'ARN in vitro. Peut être impliqué dans le métabolisme de l'ARN. En concertation avec CIC et ATXN1L, impliqués dans le développement cérébral.
TRPM7	Transient receptor potential cation channel subfamily M member 7 (1865 AA)	Canal ionique essentiel et sérine / thréonine-protéine kinase. Canal cationique divalent perméable au calcium et au magnésium. A un rôle central dans l'homéostasie des ions magnésium et dans la régulation de la mort des cellules neuronales anoxiques. Impliqué dans la nécroptose induite par le TNF en aval de MLKL en médiant l'afflux de calcium. L'activité kinase est essentielle pour la fonction du canal. Peut être impliqué dans un processus fondamental qui ajuste les flux de cations bivalents de la membrane plasmique en fonction de l'état métabolique de la cellule.
CACNA1A	Voltage-dependent P/Q-type calcium channel subunit alpha-1A (2506 AA)	Les canaux calciques sensibles à la tension (VSCC) médient l'entrée des ions calcium dans les cellules excitables et sont également impliqués dans divers processus dépendant du calcium, notamment la contraction musculaire, la libération d'hormones ou de neurotransmetteurs, l'expression génique, la motilité cellulaire, la division cellulaire et la mort cellulaire.
NEFH	Neurofilament heavy polypeptide (1020 aa)	Les neurofilaments contiennent généralement trois protéines filamentaires intermédiaires: L, M et H, impliquées dans le maintien du calibre neuronal. NF-H a une fonction importante dans les axones matures qui n'est pas soutenue par les deux plus petites protéines NF.
MT-CO1	Cytochrome c oxidase subunit 1 (513 aa)	La cytochrome c oxydase est le composant de la chaîne respiratoire qui catalyse la réduction de l'oxygène en eau. Les sous-unités 1 à 3 forment le noyau fonctionnel du complexe enzymatique. CO I est la sous-unité catalytique de l'enzyme.

TABLEAU 16, *liste et fonctions des gènes impliqués dans les MNDREM de notre population*

✓ **Outils et méthode bioinformatiques**

Pour détecter les interactions possibles entre ces gènes, nous avons utilisé *STRING*. Cet outil permet de détecter et visualiser les IPPs prédits dans la liste de gènes soumise. Les prédictions d'IPPs sont tirées des interactions expérimentalement prouvées, ou issues de bases de données sélectives de ce type d'interaction ; les arguments de vraisemblance (gènes voisins, co-occurrence, fusion de gène) ; et d'autres arguments issus de l'intelligence artificielle (extraction et exploitation des données de la littérature, homologie protéique, expérience de co-expression).

Pour identifier des clusters, nous avons utilisé *MCODE*.

Pour éliminer les faux positifs sur la base d'arguments de vraisemblance, en particulier biologiques, nous avons ensuite employé *ClueGO*.

3.2.2. Réseaux d'IPPs dans HDL2

✓ **Les gènes *JPH3* et *HTT***

***HTT*, Huntingtin**

L'Huntingtin est codée par le gène HTT situé en 4p16.3.

Le gène *HTT*, connu également sous le nom de « HD; IT15; LOMARS » comporte 67 exons. Il code essentiellement pour un seul transcrit comprenant tous ces exons. Le gène présente des polymorphismes de séquences de type SNVs. Certains sont pathogènes et impliqués dans le syndrome de Lopes-Maciel-Rodan. Il s'agit d'un syndrome neurodéveloppemental de transmission autosomique récessive, associant aux symptômes neurologiques, des troubles du comportement et des troubles orthopédiques, parfois des difficultés respiratoires et des difficultés à s'alimenter.

Par ailleurs, la région 5' riche en Glutamine (Gln) est très polymorphe. Avec des expansions pouvant aller de 10 à 35 répétitions dans la population générale. Mais les expansions supérieures ou égales à 36 et jusqu'à 120 répétitions, sont pathogènes. L'expansion de la polyglutamine conduit à une susceptibilité élevée au clivage de l'apopaïne et entraîne probablement une apoptose neuronale accélérée. Les expansions pathogènes sont responsables de la maladie de Huntington. La longueur de répétition augmente généralement au cours des générations

successives, mais se contracte également à l'occasion. Il existe une région poly-Proline (Pro) adjacente, qui est également polymorphe avec des expansions qui varient entre 7 à 12 résidus.

La protéine Huntingtin jouerait un rôle dans la fonction vésiculaire ou le transport vésiculaire médié par les microtubules. Il s'agit d'une grosse protéine ubiquitaire de 3142 AA et de 347,603KDa. On trouve toutefois une surexpression cérébrale, principalement dans le cortex cérébelleux, le néocortex, le striatum et la formation hippocampique.

Elle présente 2 régions d'interaction protéique, avec *TPR* et *ZDHHC17*. 5 domaines type HEAT* et plusieurs motifs poly-AA répartis de l'extrémité N à C-terminal comme suit : poly-Gln et poly-Pro (N-ter), Poly-Thr et poly-Glu sur le reste de la séquence peptidique.

**Les HEAT sont des motifs structurels constitués de répétitions en tandem dont le nom est issu des 4 protéines possédant ce type de motif : Huntingtin (H), facteur d'Elongation eucaryotique 2 (E), phosphatase 2A (A) et kinase Tor1 (T).*

JPH3, Junctophilin-3

La Junctophilin-3 est codé par le gène *JPH3* situé en 16q24.2.

Le gène *JPH3*, connu également sous les noms de « JP3; HDL2; JP-3; TNRC22; CAGL237 », comporte 8 exons. Il code essentiellement pour un seul transcrit comprenant tous ces exons, mais des variants d'épissage beaucoup plus courts ont été décrits. Le gène est très peu polymorphe et n'est associé actuellement qu'à une seule pathologie humaine : Huntington-Like Disease 2 (HDL2). HDL2 est causée par l'expansion pathogène des répétitions CAG/CTG (40-59 répétitions) de la séquence située en 3'-UTR du gène (norme 6-28 répétitions).

La protéine JPH3 appartient aux protéines de la famille des junctophilines, super-famille de molécules impliquées dans les complexes membranaires jonctionnels (JMC) qui relient la membrane plasmique au réticulum endoplasmique ou sarcoplasmique dans les cellules excitables et assurent ainsi la médiation entre la surface cellulaire et les canaux ioniques intracellulaires. JPH3 est exprimé presqu'exclusivement dans le cerveau et semble avoir un rôle actif dans certains neurones impliqués dans la coordination motrice et la mémoire.

JPH3 est une protéine de 748 AA et d'environ 81KDa, composée d'un segment hydrophobe C-terminal (21 AA) ancré dans la membrane du réticulum endo- ou sarcoplasmique et d'un domaine cytoplasmique restant (727 AA) qui montre une affinité spécifique pour la membrane plasmique.

La protéine contient 8 domaines MORN (Membrane Occupation and Recognition Nexus). Ces domaines protéiques sont retrouvés dans les protéines de jonction membranaire mais leur fonction n'est pas encore clairement décrite.

✓ Outils et méthode bioinformatiques

Pour constituer un set de données de protéines en interaction avec *JPH3*, nous avons interrogé les bases *STRING, APID, Genemania*.

Pour la visualisation du réseau, nous avons opté pour l'outil *Cystoscape* pour sa popularité et ses performances. *Cytoscape* centralise les IPPs d'un grand nombre de bases de données, grâce à *PSICQUIC*, interface web de requête sur les IPPs [134]. Il permet de visualiser des réseaux de plusieurs centaines de milliers de nœuds et de cordes. [135]

Nous avons éliminé les FPs sur la base d'arguments de vraisemblance. Pour cela, nous avons annoté les gènes en utilisant *Gene Ontology*. Nous avons identifié les processus et fonctions biologiques, ainsi que les compartiments cellulaire correspondants à chaque gène [136]. Puis, avec *ClueGO,* nous visualisons ces annotations, puisque l'outil permet de mettre en évidence, les processus biologiques non redondants dans le réseau pour identifier les clusters de gènes.

Nous avons renforcé notre proposition par la recherche de clusters, basée sur la densité locale avec l'outil *MCODE* (Molecular Complexes Detection) [137].

3.3. Résultats

3.3.1. Les gènes de MNDREM sont en interaction

STRING permet de visualiser un réseau relativement dense d'IPPs entre les gènes de MNDREM mis en évidence dans nos travaux. La plupart des gènes semble être en interaction (FIGURE 44).

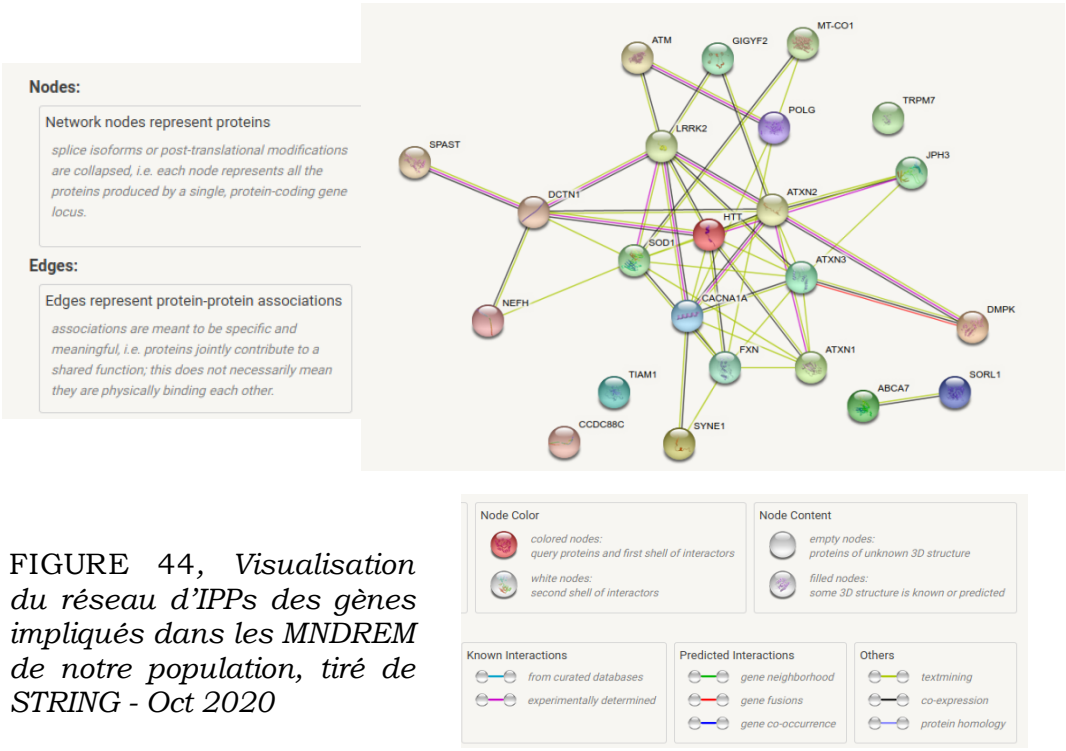


FIGURE 44, *Visualisation du réseau d'IPPs des gènes impliqués dans les MNDREM de notre population, tiré de STRING - Oct 2020*

L'outil ne prédit pas d'interaction impliquant les gènes suivants : *TIAM1*, *CCDC88C* et *TRPM7*. Tandis qu'*ABCA7* et *SORL1* semblent en interaction exclusive.

- ✓ **HTT, FXN, DCTN1, ATXN1, ATXN2, et ATXN3 forment un cluster**

MCODE renforce les IPPs vraies. Il permet de proposer un cluster unique de 6 nœuds et 11 cordes, impliquant les gènes : *HTT, FXN, DCTN1, ATXN1/2* et *3*. Comme le montre la FIGURE 45 ci-après.

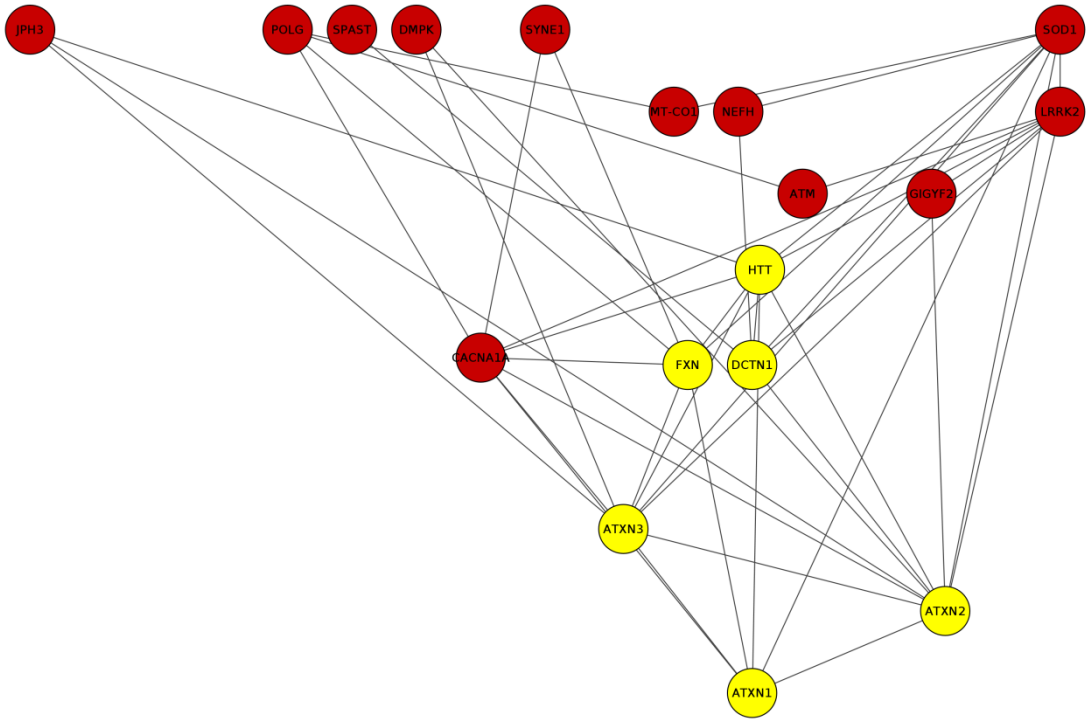


FIGURE 45, *Cluster d'IPPs entre HTT, FXN, DCTN1, ATXN1, ATXN2, et ATXN3 proposé par MCODE – Oct 2020 ;*

En rouge les gènes dont les IPPs prédites ne sont pas renforcer par la possibilité statistique de former un cluster ; en jaune le cluster proposé par MCODE. Les cordes entre les gènes ne sont pas toutes de même valeur

✓ Cytosquelette, microtubules et corps d'inclusion sont les structures associées aux IPPs de confiance

L'analyse des gènes par *ClueGO* révèle essentiellement 5 catégories de processus cellulaires dans lesquels les gènes d'intérêt sont impliqués :

- Corps d'inclusion,
- Processus de filament intermédiaire (organisation du cytosquelette à filament intermédiaire),
- Transport intracellulaire dépendant du cytosquelette,
- Transport des organites le long du microtubule,
- Mouvement des organismes multicellulaires (mouvement musculo-squelettique, processus neuromusculaire contrôlant l'équilibre).

ClueGO propose des clusters de gènes en fonction des processus biologiques (FIGURE 46). Le TABLEAU 17 ci-après, montre la répartition des gènes dans les

catégories. Certains gènes sont éliminés pour ne laisser que ceux pour lesquels des IPPs sont prédites avec un certain niveau de confiance.

Les gènes *HTT, FXN, DCTN1, ATXN1, ATXN2,* et *ATXN3* impliqués dans le cluster participent aux processus : corps d'inclusion, processus à base de filament intermédiaire (organisation intermédiaire du cytosquelette filamentaire), transport intracellulaire dépendant du cytosquelette, transport des organites le long du microtubule.

Nous avons vu précédemment que *ATXN1, 2 et 3*, ainsi que *FXN* sont responsables d'ataxies spinocérébelleuses. On note que *DCTN1* est responsable de pathologie de la corne antérieure, tantôt SLA ou neuropathie héréditaire distale motrice 7B ou un syndrome de Perry. Le syndrome de Perry est un trouble neuropsychiatrique caractérisé par une dépression mentale qui ne répond pas aux médicaments antidépresseurs. Un syndrome parkinsonien apparaît plus tard et une détresse respiratoire survient en phase terminale.

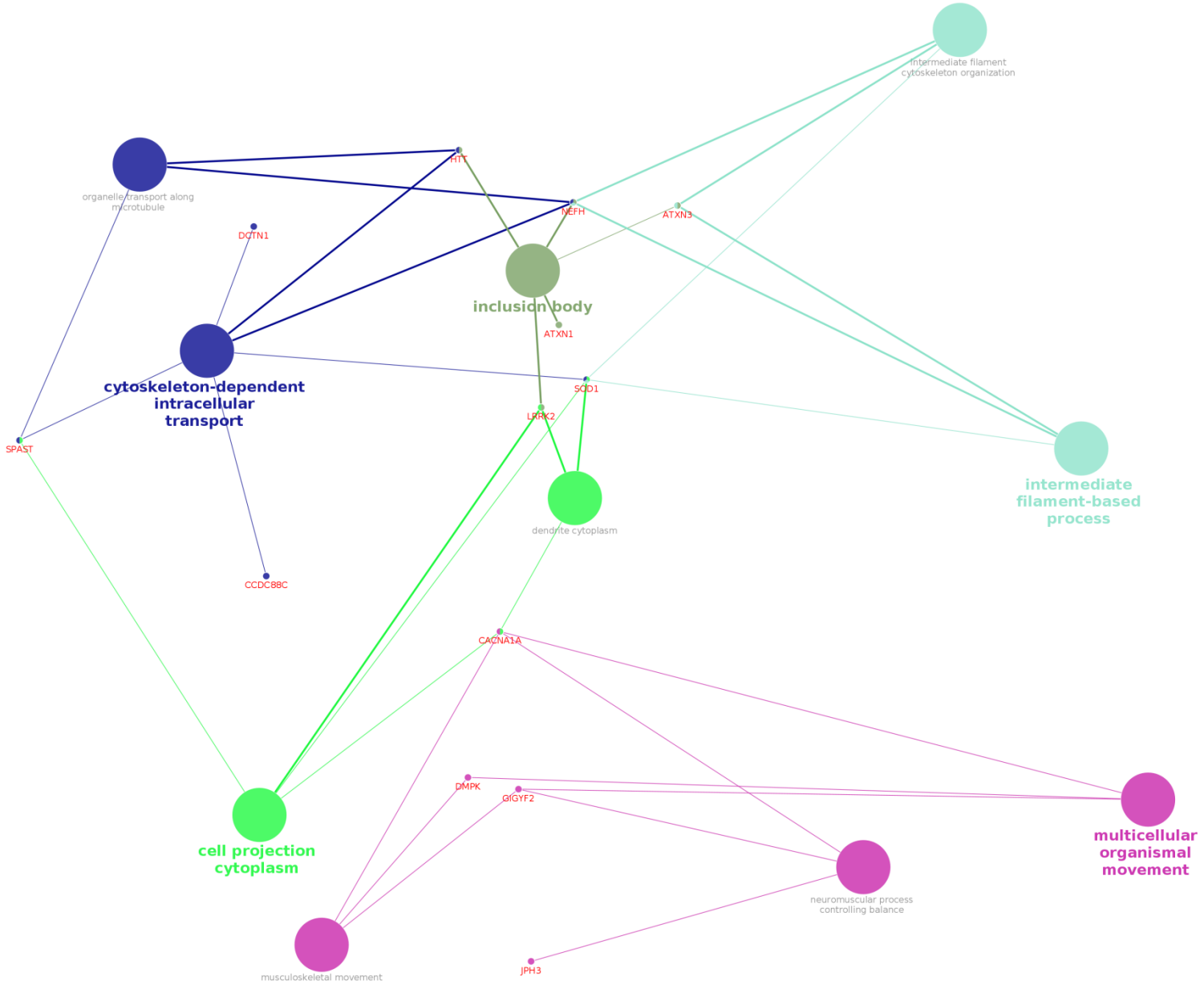


FIGURE 46, *clusters de gènes MNDREM proposés par ClueGO – oct 2020*

GO :ID	GO Term	Group PValue Corrected with Bonferroni step down	Nr. Genes	Associated Genes Found
GO:0016234	inclusion body	160,0E-9	5,00	[ATXN1, ATXN3, HTT, LRRK2, NEFH]
GO:0032838	cell projection cytoplasm	3,2E-6	4,00	[CACNA1A, LRRK2, SOD1, SPAST]
GO:0032839	dendrite cytoplasm	3,2E-6	3,00	[CACNA1A, LRRK2, SOD1]
GO:0045103	intermediate filament-based process	22,0E-6	3,00	[ATXN3, NEFH, SOD1]
GO:0045104	intermediate filament cytoskeleton organization	22,0E-6	3,00	[ATXN3, NEFH, SOD1]
GO:0030705	cytoskeleton-dependent intracellular transport	65,0E-9	6,00	[CCDC88C, DCTN1, HTT, NEFH, SOD1, SPAST]
GO:0072384	organelle transport along microtubule	65,0E-9	3,00	[HTT, NEFH, SPAST]
GO:0050879	multicellular organismal movement	12,0E-6	3,00	[CACNA1A, DMPK, GIGYF2]
GO:0050881	musculoskeletal movement	12,0E-6	3,00	[CACNA1A, DMPK, GIGYF2]
GO:0050885	neuromuscular process controlling balance	12,0E-6	3,00	[CACNA1A, GIGYF2, JPH3]

TABLEAU 17, *catégories de processus cellulaires associés aux gènes de MNDREM détectés par ClueGO – oct 2020.*

On relève que certains gènes sont impliqués dans des clusters de façon isolée, alors que d'autres interagissent dans plusieurs processus.

Les gènes *CACNA1A, LRRK2, SPAST, HTT* et *ATXN3* sont l'interface de 2 processus :

- *CACNA1A* est une jonction entre les processus qui impliquent le mouvement et les projections cytoplasmiques.
- *LRRK2*, entre les projections cytoplasmiques et les inclusions cellulaires.
- *SPAST* entre transport et projections.
- *HTT*, entre inclusions et transport.
- *ATXN3*, entre processus médiés par les filaments intermédiaires et inclusions.

Tandis que *SOD1* et *NEFH* sont les interfaces de 3 processus : *SOD1* : projections, transport et filaments ; *NEFH* : transport, filament et inclusion.

3.3.2. Proposition d'IPPs avec JPH3

APID, STRING et *Genemania*, proposent au total 166 interactants avec JPH3 (TABLEAU 18). On reconnaît certains gènes déjà décrits dans les ataxies, les paraplégies spastiques, les chorées en particulier acanthocytose et maladie de Huntington.

L'élimination des faux positifs par *ClueGO* permet de retenir 32 protéines dont les interactions sont statistiquement évaluées pour chaque catégorie d'argument, au seuil de significativité *p* (calculée avec correction de Bonferroni) très faible.

Ces 32 protéines sont ensuite filtrées sur la base d'arguments de vraisemblance suivants :

- Processus biologiques : régulation du transport des ions calcium dans le cytosol
- Co-localisation cellulaire : jonction membranaire du réticulum sarcoplasmique
- Fonction moléculaire : activité des canaux de libération de calcium

ACADSB	CORO2B	HMX1	NAT6	RAB15	TFF3
ACHE	COTL1	HSDL1	NELFCD	RBFOX1	TGM1
ACTN2	CPED1	HSPH1	NHLH2	RDH16	TMEM257
ADAM7	CPNE8	**HTT**	NR2E1	RET	TRDN
ADGRB2	CPNE9	IPO4	NUDT10	RFK	TRPC3
AKIRIN2	CUBN	ITIH3	OPCML	RIT2	TSPYL1
ALS2CL	DEC1	JPH1	OR10J3	RNF39	**VPS13A**
APOL5	DEFA4	JPH2	PACSIN1	RSPH1	**XK**
ARHGEF17	DHRS7C	**JPH3**	PANK1	RSPH10B	XPO7
ATN1	DLGAP1	JPH4	PANK2	RSPH10B2	ZBTB7B
ATP6V1G2	DNM3	**KCNC1**	PCDH13	RYR1	ZCCHC14
ATXN10	DUSP13	KCNE5	PCDH20	RYR2	ZDHHC7
ATXN2	ERC2	KCNIP1	PCDH8	SETD7	ZFX
ATXN3	ETV4	KCNN1	PCNX2	SLC25A14	ZNF343
AVP	EVX1	**KIF1A**	PDC	SLC25A30	ZNF48
BCHE	FAM131B	LBX1	PDIA2	SLC3A1	ZNF536
BGN	FAM222B	LIMK1	PGS1	SLC6A15	
C16orf78	FANCC	LRRC73	PI4KA	SLC9B2	
C16orf95	FKBP1B	MAL	PLD2	SMAD3	
C2CD2L	FOXI1	MALT1	POU3F3	SNCB	
CA11	FRMPD1	MAPK11	PP1A	SPAG6	
CA5A	FSD1	MARCH4	PPP2R2B	STIM1	
CASQ2	FXYD7	MBNL1	PPP2R5C	STK24	
CBR3	GALNT18	MORN1	PRKCG	STMN4	
CCK	GLP2R	MORN2	**PRNP**	SYPL2	
CDH8	GRIN1	MORN3	PRODH	SYT5	
CDK5R2	GSC	MORN4	PSD	TBP	
CECR6	HIBADH	MORN5	PTH	TBX3	
CHRDL1	HIST1H2BN	MRPS18C	PTPRN2	TDRD12	
CNNM2	HLCS	MTHFR	QDPR	TECR	

TABLEAU 18, *liste des interactants potentiels avec JPH3 (APID, STRING et Genemania, octobre 2020).*

La filtration sur la base de la **co-localisation cellulaire** ne révèle aucun cluster.

Tandis que celle sur les **fonctions moléculaires** montre d'une part, un cluster entre *JPH3, JPH2 et RYR1* ; d'autre part, *JPH3, BCHE, PRNP*. Respectivement impliqués dans l'activité du canal calcique et les liaisons bêta-amyloïde ou ion ammonium. *RYR1* code pour un récepteur à la ryanodine du muscle squelettique. Les variants pathogènes de ce gène sont associés à une susceptibilité à l'hyperthermie maligne, une myopathie à central core, et à une myopathie minicore avec ophtalmoplégie externe ; *BCHE* code pour une enzyme cholinestérase à large spécificité de substrat. Elle contribue à l'inactivation du neurotransmetteur acétylcholine et peut dégrader les esters organophosphorés neurotoxiques. Les variants pathogènes de ce gène sont responsables d'une maladie métabolique autosomique récessive caractérisée par une sensibilité accrue à certains médicaments anesthésiques, y compris les relaxants musculaires succinylcholine ou mivacurium. L'hydrolyse plus lente de ces médicaments entraîne un bloc neuromusculaire prolongé, conduisant à une apnée dont la durée varie en fonction de l'ampleur du déficit enzymatique ; *PRNP* code pour une protéine majeure à prion dont la fonction physiologique n'est pas claire. Toutefois, on pense qu'elle peut jouer un rôle dans le développement neuronal et la plasticité synaptique, l'entretien et l'homéostasie de la gaine de myéline, l'absorption et l'homéostasie du fer. Les oligomères solubles sont toxiques pour les cellules de neuroblastome en culture et induisent l'apoptose. Les variants pathogènes de ce gène sont impliqués dans divers phénotypes associés à des troubles moteurs mais sutout démentiels : la maladie de Creutzfeldt-Jakob (MCJ), l'insomnie familiale mortelle (FFI), la maladie de Gerstmann-Straussler (GSD), la maladie de Huntington-like de type 1 (HDL1) et le kuru.

Enfin, la filtration basée sur **les processus biologiques** porte à penser qu'il a un cluster d'IPPs entre *JPH1, JPH2, JPH3, JPH4* et *HTT*. Nous documentons également la nature des interactions prédites (TABLEAU 19). Ce qui permet de renforcer la confiance dans une interaction : les junctophilin, de même famille protéique sont co-exprimées, mais RYR1 est en fait le seul partenaire prédit qui est aussi co-exprimé. *STRING* propose cette interaction car il a été expérimentalement prouvé que les protéines orthologues de *Mus musculus* interagissent.

Protein	Interaction	Database
ACTN2	Co-expression	Genemania
AVP	Co-expression	Genemania
BCHE	Predicted	Genemania
CASQ2	Co-expression/database_annotated/automated_textmining	String
CCK	Co-expression	Genemania
CNNM2	Co-expression	Genemania
DHRS7C	Co-expression/automated_textmining	String
FKBP1B	Co-expression/Co-localization	Genemania
FXYD7	Co-localization	Genemania
GRIN1	Co-expression/automated_textmining	String
HTT	experimentally_determined_interaction/automated_textmining	String
JPH1	Shared protein domains – homology/Co-expression/database_annotated/automated_textmining	Genemania – String
JPH2	Shared protein domains	Genemania
JPH4	Shared protein domains – homology/Co-expression/database_annotated/automated_textmining	Genemania – String
KCNC1	Co-expression/automated_textmining	String
KCNE5	Co-expression	Genemania
KCNIP1	Co-expression/experimentally_determined_interaction/automated_textmining	String
KCNN1	Co-expression/Co-localization	Genemania
PDIA2	Co-expression	Genemania
PRKCG	Co-expression/automated_textmining	String
PRNP	automated_textmining	String
PTH	Co-expression	Genemania
RIT2	Co-expression/automated_textmining	String
RYR1	Co-expression/database_annotated/automated_textmining	String
RYR2	Co-expression/automated_textmining	String
SMAD3	in vitro	APID
STIM1	automated_textmining	String
SYPL2	Co-expression/automated_textmining	String
TRDN	Co-expression/database_annotated/automated_textmining	String
TRPC3	Co-expression/experimentally_determined_interaction/automated_textmining	String
XK	Co-expression/automated_textmining	String

TABLEAU 19, *nature des interactions et base de données référentes*

4. Interactions géniques et chevauchement clinique ?

Les MNDREM présentent des chevauchements cliniques et moléculaires impliquant probablement des voies de signalisation communes. Pour interroger les facteurs de sélectivité, l'étude expérimentale de partenaires protéiques fonctionnels serait informative mais bien trop fastidieuse tant les interactions protéines-protéines peuvent être complexes et nombreuses. La bioinformatique vient au secours de cette problématique en permettant la mise en œuvre d'outils de prédiction basés sur divers arguments moléculaires, biologiques, phylogéniques etc.. L'approche de prédiction d'IPPs est très utilisée en oncologie mais plus rare dans les champs de la neurogénétique.

Notre méthode de détection d'IPPs conforte l'idée que les gènes impliqués dans le MDREM d'intérêt, interagissent entre eux dans le cadre de processus biologiques impliquant le transport cytosquelette-dépendant, le mouvement des organismes multicellulaires ou encore les corps d'inclusion cellulaires. En particulier, on prédit un cluster entre les gènes *HTT, FXN, DCTN1, ATXN1, ATXN2*, et *ATXN3*, qui sont des gènes sujets aux expansions de motifs répétés, sauf *DCTN1*.

Pour comprendre la redondance phénotypique de HD et HDL2, pourtant causé par des variants pathogènes dans des gènes différents, et pour approcher la vulnérabilité sélective des cellules cérébrales qui dégénèrent, nous avons mis notre méthode à l'épreuve pour étudier les réseaux de protéines interagissant avec la protéine JPH3 par une analyse computationnelle. Nous avons montré que JPH3 est susceptible d'interagir de façon fonctionnelle avec une trentaine de protéines dont certaines sont impliquées dans la régulation du transport des ions calcium dans le cytosol, la jonction membranaire du réticulum sarcoplasmique ou encore l'activité des canaux de libération de calcium. Et nous avons trouvé que *RYR1, BCHE*, et *PRNP* pourraient être des partenaires de *JPH3* dans les fonctions de régulation de canaux calciques et liaisons bêta-amyloïde. L'identification des cascades de signalisation moléculaires impliquant JPH3 est un enjeu thérapeutique majeur puisque, contrairement à HTT, JPH3 est une protéine exprimée spécifiquement dans le cerveau. Cette sélectivité d'expression laisse entrevoir la possibilité d'IPPs restreint et envisager des cibles thérapeutiques dans le cadre des patients atteints de HDL2.

CHAPITRE V

Discussion générale

CHAPITRE V - Discussion générale

Les Maladies Neurodégénératives à Expression Motrice (MNDEM) sont un groupe très hétérogène de pathologies neurodégénératives. La revue de la littérature laisse penser que les MNDREM sont un véritable éventail de pathologies neurologiques dont les mécanismes moléculaires sont en fait étroitement liés. Nous avons donc proposé une représentation schématique en éventail des MNDREM basée sur les types d'atteintes motrices vis à vis des processus cellulaires impactés. La plupart des variants pathogènes impliqués dans les MNDREM sont des SNVs, mais on trouve également une forte représentation des variants pathogènes par expansions de motifs répétés.

En routine diagnostic moléculaire des MNDREM, on utilise le NGS pour rechercher des variants pathogènes par l'analyse de panel de gènes. La rentabilité diagnostic du NGS panel de gènes dans l'étiologie moléculaire des pathologies avec mouvements anormaux a été soulignée dans l'étude de Montaut et al. [138]. On emploie plus rarement l'exome (WES) qui est proposé plutôt dans les pathologies neurodévelopementales rares, non étiquetées et/ou néonatales. L'accès à l'exome dans les MNDREM est actuellement régulé par des filières de validation de prescription. Cependant, il est mis en œuvre pour le diagnostic des CA.

Les CNVs, en revanche, sont peu décrits et explorés dans les MNDREM. Seulement quelques articles référencés sur pubmed en juillet 2020 sur l'implication des CNVs dans les pathologies type Alzheimer, Huntington, Démence à corps de Lewy, Parkinson ou SLA [139]–[142]. Blauw et al., ont montré qu'ils sont particulièrement augmentés dans les SLA sporadiques [119]. La méthode de référence pour la recherche de CNVs est l'Analyse Comparative sur Puce à ADN, ACPA ou Comparative Genomic Hybridization (CGH). Mais de plus en plus d'outils bioinformatiques sont développés pour autoriser la détection de CNVs à partir des données de NGS. Les méthodes associant exome et détection de CNVs ont également montré leur efficacité dans l'identification de gènes responsables d'ataxies spinocérébelleuses [108].

Pour ce qui est de la détection de variants par expansions de motifs répétés, en principe, elle reste limitée dans les méthodes de NGS 2e génération, principalement pour des raisons techniques. Dans ce registre, il faut compter sur le NGS de 3e génération, qui offre la possibilité de séquencer sans l'étape d'expansion clonale et sur longs fragments. Cependant, l'adaptation d'outils bioinformatiques permet tout de même la détection de ce type de variants avec le séquençage 2e génération. Une récente étude sur l'analyse ciblée de panel de gènes associée à la détection de motifs répétés montre la rentabilité de cette méthode NGS dans le diagnostic des maladies neurodégénératives [143].

La méthode CoDE-seq est un exome complet (WES) augmenté qui a été mis au point par le CNRS UMR 8199 pour détecter de façon précise les CNVs et les variants pathogènes ponctuels. Elle a été éprouvé notamment dans l'obésité mendélienne et la déficience intellectuelle, les pathologies du développement et le diabète [109]. Dans le champ des pathologies neurologiques, elle a fait ses preuves dans l'Amyotrophie Spinale (SMA). L'équipe turque d'Ahmet Cevdet Ceylan, a validé son utilisation sur 80 patients atteints de SMA chez qui la méthode permettait de retrouver la délétion de l'exon 7 du gène SMN1 et confirmer les résultats obtenus par MLPA* (Multiplex Ligation-dependent Probe Amplification), méthode de référence [144].

L'absence de données génétiques et démographiques sur la population de notre territoire et plus largement dans la région des Antilles françaises, nous a conduit à mettre en œuvre une stratégie d'investigation des MNDREM en Martinique basée d'une part sur l'étude démographique des MNDREM, d'autre part, l'étude génomique au moyen de la méthode CoDE-seq, qui permet de détecter SNVs et CNVs dans l'exome de cas sélectionnés. Adaptée à notre problématique, nous avons expérimenté CoDE-seq pour la première fois dans les MNDREM, et recherché tantôt les variants par expansion de motifs répétés au moyen de GangSTR. Il s'agit d'un outil bioinformatique, choisi dans cette étude parmi d'autres méthodes de détection, pour sa grande capacité d'appel de variants [95].

* *La MLPA est une variante de la PCR multiplexe qui utilise une sonde divisée en deux, nécessitant la ligation des deux parties pour que l'amplification ait lieu. Cette méthode est utilisée pour dénombrer les copies d'une même séquence dans un ADN*

A propos de l'étude démographique et clinique…

L'étude démographique de 309 cas de MNDREM nous a permis d'apporter des précisions sur leurs caractéristiques : un âge de début moyen autour de 61 ans 2 mois quel que soit le sexe et une durée d'évolution moyenne de 15 ans. Ce qui est communément attendu dans ces pathologies.

Cinq phénotypes de représentation démographique décroissant comme suit, sont mis en évidence : syndromes parkinsoniens et maladie de parkinson (SPARK/PARK), les ataxies spinocérébelleuses (CA), la sclérose latérale amyotrophiques (SLA), les maladies de Huntington et Huntington-like (HDs) et la paraparésie spastique héréditaire (PPSH).

Quarante pour cent des MNDREM ont un antécédent familial de MND. Cette proportion importante pose la question de l'héritabilité de ces pathologies en règle générale. Tous les modes de transmission ont été observés parmi les phénotypes. Mais le mode de transmission lié à l'X est d'emblée exclu par notre méthode d'analyse puisque les antécédents familiaux rapportés ne concernent que les apparentés au 1^{e} degré. Or, pour confirmer une transmission liée à l'X, il faudrait rapporter des hommes atteints, apparentés au 2nd degré, c'est-à-dire, liés par des femmes (mère et grand-mère) non atteintes.

Seulement 23% de la population de MNDREM ont une étiologie moléculaire identifiée. La rentabilité du test génétique varie selon les pathologies. Dans notre cohorte aucune variation pathogène n'avait jamais été mise en évidence dans les syndromes parkinsoniens/maladie de Parkinson alors que ce phénotype représente 47% de la population et que nous avons trouvé 7% de forme familiale de transmission dominante ou récessive. Tandis qu'il y a une variation causale dans 100% des HD, 53% des CA, 13% des SLA, 20% des PPSH.

Certaines observations originales sont à relever de l'étude démographique des formes génétiques, notamment :

- Un sexe ratio de 2 dans la maladie de Huntington, sans qu'il y ait de raison biologique connue. En effet, HD est une MNDREM de révélation tardive se transmettant sur le mode autosomique dominant avec anticipation. Ceci dit, de plus en plus d'études portent à considérer les microenvironnements spécifiques liés au sexe comme pouvant être une variable étiologique dont la connaissance permettrait notamment d'améliorer les stratégies de traitement

[145]. Par ailleurs, la présence de patients atteints de HD est bien documentée en Amérique Latine mais seulement 9 cas répartis sur 3 familles trinidadiennes sont publiés pour le bassin caribéen [146], [147][147]. Nous avons donc actuellement a priori la plus grande cohorte de patients HD de la Caraïbe.

- Un phénomène d'anticipation dans une famille de SLA avec variant pathogène *SOD1*. L'anticipation dans la SLA est un phénomène décrit depuis le début des années 2000 par une équipe japonaise [76]. Mais elle a été longtemps controversée. L'étude de Vucic sur 112 familles a finalement conclu à une anticipation statistique [107]. L'anticipation dans la SLA a toujours été associée aux variants pathogènes du gène *SOD1*, en particulier le variant pathogène p.Gly94Ser. [76]
- Une prédominance inhabituelle de SCA2 par rapport à SCA3 qui est la forme d'ataxie spinocérébelleuse dominante la plus fréquente dans les populations Européennes [31]. Comme à Cuba où on trouve un cluster familial de SCA2 dans la région de Holguin [148] ;
- Et par ailleurs, une ataxie chez un patient atteint de maladie de Steinert juvénile causée par une expansion pathogène importante, comprise entre 500 et 600 CTG dans le gène DMPK.

A propos de l'étude génomique…

L'étude CoDE-seq n'a pas toujours permis de trouver un variant pathogène de classe 5 dans l'échantillon de patients testés. Au-delà de la difficulté que pose l'interprétation des variants en elle-même, une des limites de la stratégie d'étude génomique mise en œuvre, est l'analyse en solo. En effet, actuellement, le NGS en trio est de plus en plus populaire voire de principe, car il permet, par comparaison des variants parentaux, de trouver plus facilement des variants pathogènes *de novo* ou des variants compatibles avec une transmission récessive.

…des SNVs

Nous avons toutefois pu mettre en évidence un variant SNVs probablement pathogène de classes 4 ou 5 chez 7 patients sur 13.

C'est dans les **syndromes parkinsoniens** que la méthode s'est révélée le plus productif. On propose un variant pathogène dans les gènes:

1. *ABCA7* et *MAPT*, dans la forme dominante de SPARK. Le variant STOP dans l'exon 9 du gène ABCA7 c.853A>T p.Arg285* est un variant très rare qui provoque un arrêt prématuré de la transcription du gène. Les variants pathogènes de ce gène sont associés à la maladie d'Alzheimer autosomique dominante et précoce. Mais Karen Nuytemans et al. ont montré que les variants fonctionnels dans *ABCA7* sont responsables de chevauchement entre la maladie de Parkinson et la maladie d'Alzheimer [98]. La variation pathogène STOP retrouvée dans ce gène est très probablement de fort impact fonctionnel même si aucune prédiction n'est disponible. Elle est possiblement en lien avec le phénotype observé chez le patient. D'autant plus qu'elle est associée à une délétion interstitielle 240pb emportant les exons 7 et 8 du gène *MAPT*, lui-même porteur d'un variant missense hétérozygote (c.689A>G, p.Gln230Arg) de classe 3 dans l'exon 6. Les variants pathogènes de ce gène ont été associés à plusieurs troubles neurodégénératifs tels que la maladie d'Alzheimer, la maladie de Pick, la démence frontotemporale, la dégénérescence cortico-basale et les copains supranucléaires progressifs. Des délétions plus larges, des exons 6 à 9, ont été décrites comme étant un facteur de susceptibilité à la maladie de Parkinson de transmission dominante ou multifactorielle [99]. En plus, elle est associée à un variant de signification incertaine de l'exon 6 de ce même gène, qui pourrait être un facteur de susceptibilité de la maladie de Parkinson, comme l'a montré, l'étude d'une grande famille du sud de la République tchèque [100]. Cette délétion partielle est probablement impliquée dans le phénotype.

2. *NDUFAF1, RARS2, ANO10* dans la forme atypique avec ataxie et perturbations histologiques musculaires mitochodriales. Aucun variant n'est suffisant à lui seul pour conclure à leur implication dans le phénotype. En effet : Le variant très rare missense hétérozygote c.422A>G, p.His141Arg de l'exon 6 du gène *RARS2*, est prédit délétère par tous les logiciels d'interprétation employés. Mais il répond à un mode de transmission récessif et le phénotype qu'il produit, l'hypoplasie pontocérébelleuse de type 6, est plus sévère que le phénotype du patient ; Le variant très rare hétérozygote missense c.623T>C, p.Leu208Pro de l'exon 3 du gène *NDUFAF1* est également prédit délétère par tous les logiciels d'interprétation employés et serait plus en lien avec la symptomatologie du patient. Mais son dysfonctionnement causerait une atteinte du complexe 1 de la chaîne respiratoire. Or les analyses de celle-ci sont données pour normales chez

ce patient. Le variant hétérozygote missense, c.1133G>A, p.Arg378Gln de l'exon 6, probablement de classe 4 du gène *ANO10*, associé variant d'épissage de signification incertaine (classe 3) c.1219-6delT de l'exon 7 serait un bon candidat. En effet, les variants pathogènes de ce gène sont responsables de formes récessives d'ataxie avec déficit en co-enzyme Q10 et fasciculations [103].

3. *LRRK2*, dans une forme classique de maladie de Parkinson sans antécédents familiaux a priori. Le variant hétérozygote missense c.4937T>C, p.Met1646Thr dans l'exon 34 d'interprétation discutée dans les bases de données, tantôt classe 3 ou 4 [104], [105]. Ce gène est impliqué dans la maladie de Parkinson 8 de transmission dominante.

4. *ANKK1,* dans une forme classique de maladie de Parkinson sans antécédents familiaux a priori. Le variant hétérozygote STOP c.1759G>T, p.Glu587* dans l'exon 8 du gène *ANKK1* est probablement délétère. Les variants pathogènes de ce gène ont été beaucoup étudiés dans les comportements addictifs notamment à l'alcool mais aussi la schizophrénie. Aujourd'hui, leur implication dans les phénotypes tels que dystonie myoclonique et parkinsonisme sont discutés [106]. Le variant hétérozygote missense c.4229C>T, p.Thr1410Met dans l'exon 30 du gène *LRRK2*, de pathogénicité controversée mais probablement impliquée [105]

Dans les **ataxies cérébelleuses**, la méthode permet d'argumenter sur l'ataxie sporadique, par la mise en évidence de 2 variants probablement synergiquement délétères dans :

5. *SYNE1*. Les variants c.11675T>C p.Leu3892Ser exon 71 et c.7345C>T, p.Leu2449Phe exon 49, sont prédits délétères par plus de 3 logiciels. Les variants pathogènes de ce gène ont été mis en évidence dans une cohorte franco-canadienne où il représentait le gène responsable d'ataxie spinocérébelleuse pure héréditaire et récessive [107].

Dans les cas de **Sclérose Latérale Amyotrophique**, la méthode permet de proposer des variants de classe 4 uniquement chez les patients présentant une SLA familiale causée par la variation délétère *SOD1* précédemment identifiée. On trouve des variants suivants dans les gènes de susceptibilité à l'ALS :

6. *DCTN1* : variant hétérozygote missense c.3517C>G, p.Pro1173Ala dans l'exon 27, prédit pathogène par plus de 3 logiciels.

7. *NEFH* : 2 variants missense dans l'exon 4 du gène: c.1941G>C, p.Lys647Asn et c.2601G>T, p.Lys867Asn. Variants rares et prédits pathogènes par plus de 3 logiciels. Le premier dix fois moins rare que le second, ils restent des variants peu fréquents dans la population de descendance africaine.

Dans la famille SLA, nous avons prouvé qu'il existe une anticipation statistique et nous avons pu proposer un variant probablement pathogène supplémentaire dans un autre gène de SLA chez 2 atteints testés sur 3. L'analyse des variants communs aux cas familiaux testés a permis de trouver des variants candidats dans le gène *TIAM1* associé au variant causal *SOD1*, argumentant un éventuel effet modificateur, sans toutefois permettre d'expliquer l'anticipation en propre. *TIAM1* est situé à proximité du gène *SOD1* sur la bande 21q22.11 et est considéré comme un facteur de susceptibilité de la SLA par les études d'association génome entier [103]. Nous avons décrit pour la première fois, une anticipation dans la SLA associée au variant p.Gly94Val.

On trouve également des VUS, de classe 3 dont l'interprétation reste ambigüe. En particulier :

- ce variant de nature à interrompre le cadre de lecture mis en évidence dans l'exon 29 du gène *GIGYF2* chez ce patient atteint de SPARK récessif. Un indel d'au moins 5 motifs CAG est trouvé : c.3626_3646delTGCCACAGCAGCAGCAGCAGC, p.Leu1209_Gln1215del. Une étude d'association avait montré que les variants pathogènes ponctuels du gène *GIGYF2* étaient impliqués dans PARK11 dans les populations italiennes et françaises [101]. Mais c'est l'étude menée par des chinois, qui montre que

les indels de ce gène sont impliqués dans la maladie de Parkinson d'une population issue du nord de la Chine [102].

- Le variant hétérozygote c.4500_4502delTAG, p.Ser1500del de l'exon 29 du gène *TRPM7* de nature à décaler le cadre de lecture, chez ce patient atteint de SLA atypique. Les variants pathogènes du gène *TRPM7* ont été associés à l'ALS de l'île de Guam [117]. Leur implication est finalement controversée et sont à considérer comme facteurs de susceptibilité, notamment dans les populations de l'Ouest Pacific. [118]
- Le variant hétérozygote intronique c.1050+6dupG, p. ? de l'exon10 du gène *CCDC88C*. Il est extrêmement rare dans la population africaine et impliqué dans l'ataxie spinocérébelleuse 40. Les variants pathogènes du gène *CCDC88C* sont rares. L'un est décrit dans une famille polonaise qui présente une ataxie avec tremblement et démence [149]

… des CNVs

La nouveauté qu'apporte ce travail dans l'utilisation du NGS dans les MNDREM, est la possibilité de détecter des CNVs. Et nous proposons la participation de CNVs impliquant les gènes :

- *MAPT* (del7q21.1) dans une maladie de Parkinson dominante. Les délétions partielles de ce gène sont associées à la FTD17 [150].
- *C21orf2* (dup21q22.3) et *UBQLN2* (Xp11.21) chez le plus jeune cas de SLA familiale avec anticipation ; Les variants pathogènes conduisant à des substitutions de proline dans la région de répétition PXX qui est unique à *UBQLN2*, ont des conséquences fonctionnelles. Notamment une altération de la dégradation des protéines [116]. Des récentes études d'association génome entier rapportent *C21orf2* comme un nouveau facteur de susceptibilité à la SLA [117].
- *RAB39B* (dupXq28) chez ce patient avec déficience intellectuelle, épilepsie et syndrome parkinsonien [151]
- Petites délétions emportant tout ou partie des gènes *ALS2, VCP* et *FANCG.* Les variants pathogènes de ces gènes sont responsables de phénotype SLA +/- atteintes cognitives, syndromiques ou juvénile. Des études sur les CNVs dans la SLA tendent à montrer l'impact de ce type de variants dans ce phénotype [102].

… des tandem repeats

Pour ce qui est de la recherche de répétitions en tandem, elle a été mise en œuvre notamment chez ce patient atteint de syndrome parkinsonien avec ataxie spinocérébelleuse pour lequel le test génétique diagnostic avait souligné la présence de variant d'expansion de taille à la limite supérieure de la zone grise dans le gène *ATXN1*. Nous avons retrouvé cette variation de taille limite par l'analyse GangSTR. Fort de cela, nous avons employé la méthode pour vérifier la présence d'expansion de motifs répétés dans les gènes de CA qui y sont sujets, lorsque des SNVs pathogènes n'étaient pas trouvé. Par cette méthode, nous avons ainsi exclu, a priori, des variants d'expansion pathogène dans les gènes : *ATXN2, ATXN7, CACNA1A, KLHL1AS, ATXN10, PPP2R2B, TBP, ATXN1L, ATXN2L, ATN1, NOP56* et *BEAN* chez les patients évalués. En revanche, on note des variations d'expansion dans les motifs répétés des exons 3 et 6 du gène ATXN3. Des insertions multiples de motifs CAG semblent concourir à la génération d'expansion de motifs polyAla dont la présence est toxique [108]. Ces tracts pourraient être à l'origine du phénotype partagé par mère et fille.

Au total, nous avons détecté des SNVs et des CNVs impliquant les gènes : *ABCA7, GIGYF2, LRRK2, SORL1, SYNE1, CCDC88C, TRPM7, DCTN1, TIAM1, NEFH,* et *SOD1*. On pourrait qualifier certains d'entre eux de gènes de chevauchement car ils peuvent être impliqués dans plusieurs symptomatologies. Par exemple : *SYNE1* est impliqué à la fois dans l'ataxie spinocérébelleuse de type 8, et la myopathie de type Emery-Dreifuss autosomique dominante ; *CCDC88C* est impliqué à la fois dans SCA40 et l'hydrocéphalie de type 1 ; *NEFH* est impliqué à la fois dans la SLA et dans la maladie de Charcot-Marie-tooth de type 2CC ; *SOD1* est impliqué à la fois dans la SLA et dans un syndrome neurodéveloppemental associant tetraplégie spastique progressive et hypotonie axiale.

Notre méthode a offert un rendement non négligeable dans l'identification de variations causales dans les MNDREM d'intérêt non étiquetées. Presque la moitié des cas testés (7 cas sur 12, le 13^{e} étant un apparenté sain) trouve une proposition d'étiologie moléculaire, soit une rentabilité diagnostique de 58%. Ce qui est deux fois plus que le rendement proposé par Montaut dans l'étude des maladies à expression motrice par l'étude de panel de gènes en NGS (22%) [138].

A propos de la détection des Interactions Protéines-Proteines dans les gènes de notre population, en particulier JPH3

Dans l'optique d'interroger les interactions géniques de notre population, nous avons mis en œuvre une méthode bioinformatique basée sur la prédiction d'Interaction Protéine-Protéines (IPPs). En effet, cette approche de prédiction d'IPPs est très utilisée en oncologie mais plus rare dans les champs de la neurogénétique. Cependant, on peut référencer Meena Kumari Kotni et al., qui, en 2016, avait étudié l'expression différentielle de gènes (DEG) chez des patients atteints de Sclérose Latérale Amyotrophique (SLA) par variation pathogène de *C9ORF72*, par rapport à des témoins sains. L'analyse bioinformatique visait à construire un réseau de partenaire protéique de ces DEG, en intégrant les données ontologiques et des voies de signalisation. De nouvelles cibles thérapeutiques avaient donc été proposées [152]. De même, l'étude bioinformatique de Schaefer sur le rôle des répétitions CAG/polyglutamine dans le génome, suggère que la recherche en la matière devrait se concentrer sur l'amélioration de la caractérisation de l'influence des protéines à polyQ de type sauvage sur les interactions protéiques [153].

Notre méthode de détection d'IPPs conforte l'idée que les gènes impliqués dans le MDREM d'intérêt, interagissent entre eux dans le cadre de processus biologiques impliquant le transport cytosquelette-dépendant, le mouvement des organismes multicellulaires ou encore les corps d'inclusions cellulaires. En particulier, nous prédisons un cluster entre les gènes *HTT*, *FXN, DCTN1, ATXN1, ATXN2*, et *ATXN3*, qui sont des gènes sujets aux expansions de motifs répétés, sauf *DCTN1*. Ce sont aussi des gènes d'ataxie, sauf *HTT* et *DCTN1*. En particulier, *DCTN1* est responsable de pathologie de la corne antérieure, tantôt SLA, tantôt neuropathie héréditaire distale motrice 7B, tantôt un syndrome de Perry. Le syndrome de Perry est un trouble neuropsychiatrique caractérisé par une dépression mentale qui ne répond pas aux médicaments antidépresseurs. Un syndrome parkinsonien apparaît plus tard et une détresse respiratoire survient en phase terminale.

Pour confirmer les IPPs prédites et améliorer la méthode, nous pourrions utiliser le Negatome ou mettre en œuvre des études expérimentales. L'identification des cascades de signalisation moléculaires impliquant *JPH3* est un enjeu thérapeutique majeur puisque, contrairement à HTT, JPH3 est une protéine exprimée spécifiquement dans le cerveau. Cette sélectivité d'expression laisse entrevoir la possibilité d'IPPs restreint au niveau de JPH3 par rapport à HTT, et envisager des

cibles thérapeutiques dans le cadre des patients atteints de HDL2. Nous avons montré que JPH3 est susceptible d'interagir de façon fonctionnelle avec une trentaine de protéines. Et nous avons trouvé que *RYR1, BCHE*, et *PRNP* pourraient être des partenaires de *JPH3* dans les fonctions de régulation de canaux calciques et liaisons bêta-amyloïde.

RYR1 semble être un partenaire inattendu puisqu'il ne s'agit pas d'un gène de neurodégénérescence évoqué jusqu'ici. Il est impliqué dans une susceptibilité à l'hyperthermie maligne et une myopathie minicore avec ophtalmoplégie externe. Mais le concours des gènes *BCHE* et *PRNP* impliqués dans des pathologies neurodégénératives avec atteinte motrice et surtout démentielle, est moins surprenant. Ce qui suscite l'intérêt de cette méthode dans le challenge de la détection de partenaires physiopathologique de JPH3.

CONCLUSION ET PERSPECTIVES

Conclusion et perspectives

Le travail mené pour l'obtention du grade de docteur en sciences et présenté dans ce manuscrit avait deux objectifs principaux :

- Dresser la cartographie des Maladies Neurodégénératives à Expression Motrice (MNDREM)
- Trouver des variants géniques pathogènes responsables de MNDREM sans étiologies proposées par la filière de diagnostic médical.

La cartographie des MNDREM issue de ces travaux, a vocation de permettre par la suite, de proposer un support de registre de données locales, comme ceux du cancer ou des malformations congénitales qui existent en Martinique. Les résultats du travail démographique jettent les bases d'études cliniques et les prémices d'études épidémiologiques de certaines MNDREM.

Pour la recherche de variations causales dans les MNDREM et l'exploration d'éventuelles spécificités géniques locales, nous avons mis en œuvre une étude génomique plus large et plus complète que les panels utilisés en routine. Nous avons utilisé une méthode de séquençage haut débit exome entier associée à la possibilité de détecter des variants de nombre de copies (CNVs) : la méthode dite Copy Number Variation DEtection Sequencing, CoDE-Seq. Nous avons sous-traité le séquençage à une plate-forme spécialisée, LIGAN-GENOSCREEN basée à Lille, mais développé localement les outils informatiques indispensables à l'analyse et l'interprétation des données de NGS. Le centre de calcul de l'Université des Antilles a hébergé de façon sécurisée, les données de ces travaux. Nous avons tantôt développé des programmes d'analyse « maison ». Et nous avons pu vérifier la possibilité de détecter des expansions de motifs répétés au moyen de la bioinformatique. Ce qui est encourageant et prometteur quant à la puissance de détection de divers variants par l'utilisation du NGS. La méthode CoDE-seq est une méthode de choix pour interroger l'oligogénisme et la diversité des variants (SNVs, CNVs et expansion de motifs). Elle s'est révélée efficace dans notre entreprise.

Pour aller plus loin, nous avons par ailleurs fait l'essai de méthode bioinformatique à la recherche d'Interaction entre gènes de MNDREM car la question des chevauchements symptomatiques et de l'hétérogénéité clinique les interroge. Nous avons montré que les gènes associés aux MNDREM sont en interaction.

Ces travaux devraient par la suite faire l'objet de publications sur les données d'observations cliniques, les gènes impliqués dans notre population, ou encore les CNVs proposés dans la participation de certains phénotypes. Mais il semble toutefois primordial de pouvoir confirmer la pathogénicité des variants de signification ambigus ou incertaine par l'étude de ségrégation familiale dans un premier temps. Dans les perspectives, on imagine l'approfondissement de la méthode prédictive de détection des IPPs dans les MNDREM, et la validation de variants au moyen d'étude in vitro.

La bioinformatique se révèle une discipline incontournable dans l'analayse, la prédiction et la modélisation. Notre équipe a d'ailleurs porté sa pierre à l'édifice de la résistance et de la lutte contre la pandémie de COVID19, en montrant, par la bioinformatique, l'affinité non négligeable d'une molécule flavonoïde issue de la pharmacopée locale envers la protéase activatrice du virus SARS-COV-2 [154].

Au final, dans ces problématiques de spécificités génétiques populationnelles et de chevauchement clinique et moléculaire, l'avenir me semble pangénomique et intégratif. Dans ce registre, les techniques « –OMICS », permettant de tirer conclusion d'univers moléculaires entiers (peptides, gènes, ARNs etc…), bien que complexe de prime abord, sont des méthodes qui se révéleront cruciales dans les années à venir dans les questions de recherche, et a fortiori diagnostic, classification et traitement. Et la bioinformatique ainsi que l'intelligence artificielle nous aideront probablement dans les gestions des données massives pour l'interprétation de processus biologiques…

Tan pou tan, nou ni tout' tan !…
Littéralement du créole : *le temps c'est le temps, on a tout notre temps !*
Signifie que nous avons tout notre temps

Journal of Biomedical Research & Environmental Sciences
Subject Area(s): INFECTIOUS DISEASES | ANTIVIROLOGY | ANTIRETROVIROLOGY | BIOLOGY
RESEARCH ARTICLE

Gossypetin Derivatives are also Putative Inhibitors of SARS-COV 2: Results of a Computational Study

Anna-Gaelle Giguet-Valard[1,2*], Kevin Raguette[1], Stephanie Morin[1], Remi Bellance[2] and Juliette Smith Ravin[1]

[1]Biospheres, research group campus Schœlcher, University of Antilles / AREBIO, 59 rue Ernest Desproges 97200 FDF-Martinique
[2]University Hospital Center of Martinique

ABSTRACT

SARS-CoV-2 is the third most highly virulent human coronavirus of the 21st century. It is linked with fatal respiratory illness. Currently, there are still no effective treatments of Covid-19. Among many drugs evaluated, few have proven conclusive clinical efficacy. Furthermore, the spread of the disease mandates that ideal medications against Covid-19 be cheap and available worldwide. Therefore, there is a rationale to evaluate whether treatments of natural origin from aromatic and medicinal plants have the ability to prevent and/or treat COVID-19. We evaluated in this study the inhibition of COVID-19 protease by natural plants compounds such as Gossypetin-3'-O-glucoside (G3'G). G3'G has been isolated from the petals of Taliparitі *elatum Sw.* found almost exclusively in Martinique. It has no crystallography or modelisation studies. Antifungal and antioxidant properties are already published. We study its binding affinity so potential inhibition capability against SARS-CoV2 3CLpro mean protease as compared to other previously tested natural or pharmacological molecules by molecular docking. We propose Gossypetin derivatives as good tropical natural compounds candidate that should be further investigated to prevent or treat COVID19.

INTRODUCTION

Severe Acute Respiratory Syndrome (SARS) was the first new infectious disease identified in the twenty-first century. This acute and often severe, respiratory illness is caused by a new coronavirus, SARS-Coronavirus (SARS- CoV). Between 2002-2003, the first outbreak of SARS reached all five continents by air-travel routes.

SARS-CoV is a zoonotic virus that resides in hosts that form its natural reservoir, such as bats. It can also infect intermediate hosts, such as small animals (for example, palm civets), before being transmitted to humans. SARS-CoV can infect and replicate in several cell types in humans and cause serious organ injuries [1].

A novel coronavirus strain linked with fatal respiratory illness was reported in late 2019. At the beginning of 2020, the World Health Organization (WHO) permanently named the 2019-nCoV pathogen as SARS-CoV-2 and the causing disease as Coronavirus Disease 2019 (COVID-2019). SARS-CoV-2 is the third most highly virulent human coronavirus of the 21st century followed by the SARS-COV and MERS-COV. SARS-COV-2 is the seventh strain of human coronavirus. It is taxonomically placed in the Genre Beta coronavirus, and exhibits 89.1% and 60% nucleotide sequence similarity with SARS and MERS coronaviruses, respectively [2,3].

***Corresponding author**

Anna Gaelle Giguet Valard, Biospheres, research group campus Schœlcher, University of Antilles / AREBIO, 59 rue Ernest Desproges 97200 FDF-Martinique, France

Tel: +059-655-2264
E-mail: anna-gaelle.giguet@chu-martinique.fr

DOI: 10.37871/jbres1144

Submitted: 16 September 2020

Accepted: 08 October 2020

Published: 09 October 2020

Copyright: © 2020 Giguet Valard AG, et al. Distributed under Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects: Biology

Topic(s): Infectious Diseases; Antivirology; Antiretrovirology; Biology

Keywords

- 3CL Protease
- SARS-COV2
- Gossypetin-3'-O-glucoside
- Molecular docking
- Free binding energy
- Natural polyphenols compounds

VOLUME: 1 ISSUE: 6

How to cite this article: Giguet Valard AG, Raguette K, Morin S, Bellance R, Ravin JS. Gossypetin Derivatives are also Putative Inhibitors of SARS-COV 2: Results of a Computational Study. J Biomed Res Environ Sci. 2020 Oct 09; 1(6): 201-212. doi: 10.37871/jbres1144, Article ID: JBRES1144

Coronaviruses are large enveloped positive single strand RNA viruses. They are highly evolving, with a high frequency of gene recombination and mutation [4,5]. Like other beta-coronaviruses, the genomic organization of SARS-CoV-2 consists in a 5'-Untranslated Region (UTR), a replicase complex ORF1AB encoding Non-Structural Proteins (NSPS), a Spike Protein (S) gene, an Envelope Protein (E) gene, a Membrane Protein (M) gene, a Nucleocapsid Protein (N) gene, and a 3'-UTR containing several unidentified non-structural open reading frames (Figure 1).

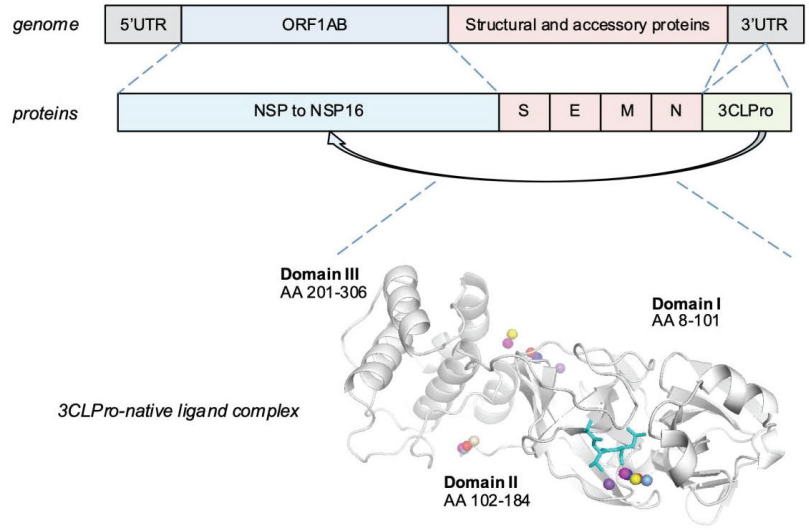
Usually, beta-coronaviruses genome transcribes an approximately 800 kDa polypeptide. This polypeptide is proteolytically cleaved to generate various proteins. The proteolytic process is mediated by a Papain-Like Protease (PLpro) and a 3-Chymotrypsin-Like Protease (3CLpro). 3CLpro gene is located at the 3'-end and exhibits a high-degree variability [5]. The 3CLpro cleaves the polyprotein at 11 distinct sites, generating various non-structural proteins involved in the virus replication. The functional importance of 3CLpro makes it an attractive target for the development of effective antiviral drugs [6,7]. 3CLpro is considered as a validated potential drug non-toxic target for anti-coronaviruses inhibitors screening [2]. Screening and repurposing of Food and Drug Administration approved antiviral drugs to inhibit 3CLpro may provide an alternative fast-track approach for identifying and developing new treatments for SARS-COV-2 infection [5].

Like other coronaviruses, the Three-Dimensional (3-D) structure of SARS-Cov-2 3CLpro consists of three domains. Domain I (8-101 amino acid residues) and II (102-184 amino acid residues) are essentially beta-barrels, and bear a resemblance to chymotrypsin. Domain III (201-306 amino acid residues) is mainly comprised of alpha-helices. The substrate binding region is located at the cleft of Domain I and II. In addition to the catalytic center, amino acid residues of subsites designated as S1 to S5, play a key role in natural substrate binding: Thr25, His163, Glu166, Cys145, Gly143, His172, Phe140, Met49, His41, Met165, Asn142 and Gln189. His41 and Cys145 3CLpro residues are considered to be critical for the virus activity. 3CLpro/ligand dyad is crucial for the catalytic activity and essential in viral replication [5].

Currently, there are still no effective treatments of Covid-19. Development of new treatments may require months to years. So far, clinical trials have been quite disappointing, and drug toxicity issues have been raised, notably for chloroquine [8,9]. Therefore, there is a rationale to evaluate whether treatments of natural origin from aromatic and medicinal plants have the ability to prevent and/or treat COVID-19 [4,10,11].

A joint research team of the Shanghai Institute of Materia Medica and Shanghai Tech University performed drug screening in silico, and enzyme activity test.

They identified thus far 30 agents with potential efficacy against COVID-19, including Western medicines, natural



ORF1AB : Open Reading Frame 1AB replicase complex ; spike protein (S), envelope protein (E), membrane protein (M), nucleocapsid protein (N), 3-chymotrypsin-like protease (3CLPro); colored balls represent alternative binding sites, native ligand (cyan color) bound the bottom right site.

Figure 1 Schematic representation of genome sequence organisation, encoding proteins and 3CPro-ligand structure.

products, and traditional Chinese medicines. They also found that Chinese herbal medicines such as Rhizoma Polygoni Cuspidati and Radix Sophorae Tonkinensis may contain active ingredients against SARS-COV-2 [12].

Computational prediction by molecular docking is a reliable method to study interactions between candidate molecules and proteins. It is widely used in sillico experiments to screen drugs inhibition efficacy. Molecular docking is a kind of bioinformatics modelling which involves the interaction of two or more molecules to give the stable adduct. Docking involves the placement of a ligand within a binding site and the prediction of the free energy of binding for such poses. In molecular docking, the most important aspect is the calculation of binding energy so as to fit a ligand in a binding site. It is actually a reference method used to find agent against nCoV-2019 [13].

Very few data is available on the clinical use of medicinal plants to prevent or cure COVID-19. However, scientists' teams worldwide (India, China, Iran, Morocco, sub-Saharian African) have published computer-based prediction studies of molecules from their own pharmacopoeia, mostly targeting the binding affinity of the main SARS-CoV-2 3CLprotease [5,14-16]. Therefore, we evaluated in this study the inhibition of COVID-19 protease by herbal plants compound found in Martinique.

Martinique is a tropical Caribbean island enjoying a large biodiversity. The population has a large use of local medicinal plants. Some proteins from Caribbean medicinal plants have already been identified in protein database for molecular docking tests. Many others have not, such as Gossypetin-3'-O-glucoside (G3'G). This molecule, isolated from the petals of Taliparitі *elatum Sw*. Found, almost exclusively in Martinique. It has no crystallography modelisation studies, despite reported antifungal and antioxidant properties [17].

In this study we searched to model G3'G, and study its binding affinity so potential inhibition capability against SARS-CoV2 3CLpro mean protease as compared to other previously tested natural or pharmacological molecules.

MATERIALS OF METHODS

Molecules

We selected pharmaceutical molecules included in the Guidelines (version 6) for treatment of COVID19 and molecules from medicinal plants. We defined a list of molecules allowing us to conclude about the effectiveness of our method and classify new candidate molecules. It is the reason why we also selected compounds that have already been computationally tested. We tested some drugs and medicinal plants molecules for the first time by molecular docking (Table 1).

Three-Dimensional (3D) dataset

3CLpro and ligands: The majority of biological macromolecules crystal structures is collected in PDB worldwide archive [18].

SARS-CoV-2 3-chymotrypsin-like protease (3CLpro) (PDB ID: 6LU7) structures were obtained from PDB (https://www.rcsb.org/). The native ligand for 6LU7 is n-[(5-methylisoxazol-3-yl) carbonyl] alanyl-l-valyl-n~1~-((1r,2z)-4-(benzyloxy)-4-oxo-1-{[(3r)-2-oxopyrrolidin-3-yl]methyl}but-2-eny-l-leucinamide [19,20].

Other three-dimensional structures of molecules of interest were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/). It is a chemical substance and biological activities repository consisting of three databases, including substance, compound, and bioassay databases [21].

3-D structures of the following molecules of interest were obtained in pdf format:

- **Natural molecules:** Kaempferol (CID 5280863), Quercetin (CID 5280459), Luteolin-7-Glucoside (CID 5280637), Aloenin (CID 162305), Cannabidiol (CID 644019), Curcumin (CID 969516), Nictoflorin (CID 5318767), Astragalin (CID 5282102), Digitoxigenin (CID 4369270), β-eudesmol (CID 91457), Gossypetin-8-glucoside (CID 5281621), Gossypetin-7-O-glucoside (CID 90658329), Gossypetin-3-glucoside (CID 44259979), Piceid (CID 5281718), Emodin-8-glucoside (CID 99649), Crocin (CID 5281233), and Gossypetin3'-O-glucoside.
- **Pharmaceutical drugs:** Nelfinavir (CID 64143), Hydroxychloroquine (CID 3652), Chloroquine (CID 2719), Amiodarone (CID 2157), Remdesivir (CID 121304016), Ribavirin (CID 37542), Arbidol (CID 131411), Lopinavir (CID 92727), Quercitrin (CID 5280459).

Modelisation: Gossypetin-3'-O-glucoside (G3'G) is not present in PDB norPubChem database. Using Opsin [22] we determined that G3'G international chemical name is 3,5,7,8-tetrahydroxy-2-(4-hydroxy-3-{[(3S,4S,5R,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}phenyl)-4H-chromen-4-one,allowing us to convert it in SMILES. We used also the "SMILES" of Crocin, and Lopinavir, which have no 3D model proposition. We use "*-h et -gen3D*" options of Openbabel software to obtain a 3-D model of these 3 molecules in pdf format [23].

In silico method

Lupinski's rule: According to Lupinski's rule, efficient drugs should have good permeation and oral absorption depending on respect of the following 5 physical and chemical properties: molecular weights > 500, C logP> 5, more than 5 hydrogen-bond donors, and more than 10

Table 1: Dataset of molecules. Shows PubChem Id, chemical formula of molecules of interest, usual trade name of drugs and plants species origin. P means medicinal plants origin and D means drugs.

Name	Type	Class	Trade name/species	PubChem Id	Chemical formula	Reference
Aloenin	P	Anthraquinones	Aloe vera	CID 162305	C19H22O10	[1]
Amiodarone	D	Class III Antiarrhythmic	Cordarone	CID 2157	C25H29I2NO3	[1]
Arbidol	D	Antiviral	Umifénovir	CID 131411	C22H25BrN2O3S	[2]
Astragalin	P	Flavonol	Harsingar	CID 5282102	C21H20O11	[3]
Beta-Eudesmol	P	Anti-Angiogenic/Anti-Tumor, Sesquiterpenoid	Laurel (*lauris nobilis L*)	CID 91457	C15H26O	[4]
Cannabidiol	P	Cannabinoid	Cannabis	CID 644019	C21H30O2	
Chloroquine	D	Antimalarial Of The 4-Aminoquinoline Family	Axemal, Dolquine, Quensyl, Plaquenil	CID 2719	C18H26ClN3	[5]
Crocin	P	Carotenoid	saffron (*crocus sativus*)	CID 5281233	C44H64O24	[4]
Curcumin	P	Polyphenol	Turmeric	CID 969516	C21H20O6	[1]
Digitoxigenin	P	Cardenolide	laurel (*nerium oleander*)	CID 4369270	C23H34O4	[4]
Gossypetin-3-Glucoside	P	Flavonoid	Hibiscus (*Hibiscus sabdariffa L/ tiliaceus*)	CID 44259979	C21H20O13	[6]
Gossypetin-3′-O-Glucoside	P	Flavonoid	Hibiscus (*Abelmoschus manihot/ Hibiscus elatus Sw*)	CID 99649	C21H20O13	[6]
Gossypetin-7-O-Glucoside	P	Flavonoid	Hibiscus (*sabdariffa L./tiliaceus/ elatus Sw.*)	CID 90658329	C21H19O13-	[6]
Gossypetin-8-Glucoside	P	Flavonoid	Hibiscus (*vitifolius/sabdariffa L*)	CID 5281621	C21H20O13	[6]
Hydroxychloroquine	D	Antimalarial Of The 4-Aminoquinoline Family	Axemal, Dolquine, Quensyl, Plaquenil	CID 3652	C18H26ClN3O	
Kaempferol	P	Flavonol	spinach, cabbage, dill, chinese cabbage, katuk	CID 5280863	C15H10O6	[1]
Lopinavir	D	Antiviral	Kaletra (ritonavir + lopinavir)	CID 92727	C37H48N4O5	
Luteolin-7-Glucoside	P	Flavones	olive, star fruit, chili pepper, welsh onion, leek	CID 5280637	C21H20O11	[1,5]
Native Ligand					C35H53N6O8	[7]
Nelfinavir	D	Antiretroviral	Viracept	CID 64143	C32H45N3O4S	[8]
Nictoflorin	P	Flavonoid	Harsingar	CID 5318767	C27H30O15	[1]
Piceid	P	Polyphenol	Picea sitchensis/ Reynoutria japonica	CID 5281718	C20H22O8	[9]
Quercetin	P	Flavonoid	dill, fennel leaves, onion, oregano, chili peppe	CID 5280459	C15H10O7	[3]
Quercitrin	P	Flavonoid	Wheat (*Fagopyrum tataricum*)/Oak (*Quercus alba/robur*)	CID 5280459	C21H20O11	[6]
Remdesivir	D	Antiviral	Remdesivir	CID 121304016	C27H35N6O8P	[5]
Ribavirin	D	Antiviral	Ibavyr	CID 37542	C8H12N4O5	[2,5]

acceptor groups. An orally active drug has no more than one violation of the following criteria [24]. Based on Lupinski's rules, we checked druglikeness of best candidate. We use RDKIT (open-source cheminformatics; http://www.rdkit.org) Lipinski's rule of five.

Determination of 3CLpro active sites: The detection of binding sites on the surface of the protein was done with MetaPocket 2.0 (MPT) [25]. This meta-server uses a consensual method, in which binding sites are predicted from eight methods: LIGSITEcs (LCS), PASS (PAS), Q-SiteFinder (QSF), SURFNET (SFN), Fpocket (FPK), GHECOM (GHE), ConCavity (CON) and POCASA (PCS). The results are combined to maximize the rate of successful prediction.

Molecular Docking: Pymol version 1.7.2.1 was used for protein optimization, by removing ligand, water and other atoms, and then adding a polar hydrogen group. Protonation at pH7.4 using amberff99 force field has been done with pdb2pqr version 2.1.0 and PROPKA algorithm [26,27]. Optimal structure of the protein was then saved as .pdf format. We converted all ligands in .mol2 format at pH7.4 with Openable software [23]. We generated several conformers and minimized theirs structures using MMFF94 force field. We used only the lowest energy conformers. Native ligand has not been minimized. For both ligands and protein, allocation of Gasteiger atomic partial charges and polar hydrogens were done with scripts from Auto Dock Tools Mgl tools 1.5.6: prepare_receptor4.py and prepare_

ligand4.py. The docking site on protein target was defined by establishing a grid box with a default grid spacing centered on the position of site native ligand. The docking analyses were performed with Autodock 4.2.6, AutodockVina version 1.1.2, and Smina version Oct 15 2019. AutoDock uses a Lamarckian genetic algorithm and an empirical free energy force field scoring function [28]. AutoDockVina employs an iterated local search global optimizer and an hybrid scoring function (empirical + knowledge based function) [29]. Smina is a fork of AutoDockVina that uses a scoring function called Vinardo, with enhanced features based on AutoDockVina [30]. To obtain free binding energy we made several per methods. We performed fifty docking runs per ligand with Autodock 4.2.6 and Smina method. Only twenty runs with AutodockVina version 1.1.2 which cannot do more [28,31].

Consensus scoring (C-score): To evaluate docking efficacy, we measured the number of favorable intermolecular interactions such as hydrogen bonds and hydrophobic contacts. But to classify which ligands are most likely to interact strongly with the protein 3CLpro, we rank them according to a consensus scoring method [32]. First, for each method we determined a score (S) such as: $S = \phi - \mu / \sigma$. Where ϕ is the value of free binding energy of a molecule; μ is the mean value of the set of molecules; σ is the standard deviation observed for the set of molecules. Then, we normalized the score S to S', where S'= (S-Smin)/(Smax – Smin). Where, Smax and Smin, are the maximum and the minimum S-score value of the set of molecules of a method. Finally, we attribute a Consensus-Score (C-score) for each molecule. This value is the sum of the three S'-score obtained for each molecule. Our ranking is based on the sorting of molecules respecting to their consensus-score.

RESULTS

Binding site determination

The determination of Amino Acids (AA) in the active site is used to analyze the Grid box and docking evaluation results. We identify residues of the active site by visualization of 3CLpro/native ligand linkages. We found several kinds of atomic links with critical AA of 3CL/Pro:

- H-bonds: Gly143, Phe140, His163, His164, Glu166, Gln189 and Thr190
- Amide Pi-stackedwithLeu141
- pi-alkyle bounds with His41, Met49, Leu167, Pro168 and Ala191
- Carbones-hydrogen bounds with Met165 and His172
- Van der Walls interactions on 11 AA including Asn142, Asp187, Arg188 and Tyr54.

Molecular docking analyses

Three-dimensional models of G3'G, Crocin and Lopinavir, which were not previously available, could be obtained using Openbabel software, and were subsequently used in the docking analysis.

Table 2 presents free binding energy (ΔG) scores found with each molecular docking method, and the T-value. Crocin has the longest carbon chain of the set. Free binding energy of the Crocin is not calculated by Autodock4 because it has too many twists. Autodock4 accepts a maximum number of 32 twists per molecules.

The consensus-score of ΔG values range from 0,111 to 2,347. G3'G is found with one of the lowest values which sign a best affinity to link 3CLPro.

We evaluated our products regarding Lupinski rules compliance and others Autodock 4 docking data (Table 3). The affinity of drug compounds is evaluated through the number and type (Hydrogen or hydrophobic type) of bonds, which occur with the active site of the protein. The lower is inhibition constant the more is affinity. While desolvation energy and intermolecular energy are proportional to affinity. Nictoflorin and Crocin present 3 violations. Chloroquine / Hydroxychloroquine and Ribavirin are the only drugs that show no violation. Among the natural compounds, Aloenin, Beta-eudesmol, Curcumin, Digitoxigenin, Kaempferol, Quercetin show no violation either. These are proteins from various classes: anthraquinones, sesquiterpenoid, polyphenol, cardenolide, flavonol, and flavonoid.

G3'G/ Gossypetin-3-Glucoside (G3G)/ Gossypetin-8-Glucoside (G8G), Digitoxigenin, Quercitrin, and Cannabidiol among natural products; Remdesivir, Chloroquine and Ribavirin among synthetic products, had the best inhibition constants. All Gossypetins derivatives, Nictoflorin and Crocin presented each more than 10 H-bonds with the protein. Remdesivir has 13 H-bonds acceptor and only 4 H bonds donor.

By matching ranking of the best intermolecular energy and desolvation energy ligands, we find that G3'G/G8G, Cannabidiol and Chloroquine are all time present in the top 10 ranking.

Finally, the products analysed were presented according to their origin (natural origin or chemical synthesis), and their rank in consensus score of free binding energy.

The compounds with the 6 more favourable free binding energy were natural molecules from medicinal plants (Table 4). All of them are polyphenols: flavonoids, flavonol or flavones. The best synthetic products in this ranking were Lopinavir and Nelfinavir (rank 7 and 10).

Table 2: Free binding energy (ΔG) by docking methods and consensus scores.

Name	Type	Autodock4	Autodock Vina	Smina	Consensus Score
Aloenin	P	-7,47	-7,2	-8,01	1,476
Amiodarone	D	-9,99	-6,8	-7	1,619
Arbidol	D	-9,38	-6,9	-7,01	1,634
Astragalin	P	-10,39	-8,1	-9,49	0,609
Beta-eudesmol	P	-7,3	-5,9	-6,07	2,347
Cannabidiol	P	-7,63	-6	-6,2	2,257
Chloroquine	D	-7,64	-6	-5,88	2,326
Crocin	P	NA	-6,8	-7,94	2,251
Curcumin	P	-8,57	-6,8	-7,06	1,725
Digitoxigenin	P	-7,79	-6,8	-7,23	1,753
Gossypetin-3-glucoside	P	-11,93	-8	-9,69	0,784
Gossypetin-3'-O-glucoside	P	-11,78	-7,5	-9,07	0,47
Gossypetin-7-O-glucoside	P	-10,66	-8,2	-9,74	0,499
Gossypetin-8-glucoside	P	-11,62	-7,9	-9,66	0,536
Hydroxychloroquine	D	-6,79	-6,2	-6,23	2,255
Kaempferol	P	-8,87	-7,7	-8,49	1,088
Lopinavir	D	-8,59	-8,9	-9,1	0,578
Luteolin-7-glucoside	P	-10,66	-8,4	-9,73	0,434
NATIVE LIGAND		-8,23	-7,6	-8,1	1,26
Nelfinavir	D	-9,44	-8,2	-8,66	0,836
nictoflorin	P	-11,4	-8,7	-10,47	0,111
Piceid	P	-9,12	-7,7	-8,78	1,004
Quercetin	P	-9,68	-7,3	-8,32	1,19
Quercitrin	P	-11,88	-8,4	-9,81	0,315
Remdesivir	D	-8,03	-8,1	-8,55	1,012
Ribavirin	D	-7,34	-6	-6,49	2,219

Table 3: Other docking data regards to Lupinski rules violations.

		Lupinski Rules					Docking DATA		
Name	Type	Molecular weight	LogP	H-Bond donor	H-bond acceptor	Violation	Inhibition Constant	Intermolecular Energy	VDW-H Bond Desolvation Energy
Aloenin	P	410,375	-0,492	5	10	0	3,32	-8,36	-8,36
Amiodarone	D	645,319	6,936	0	4	2	47,76	-12	-12
Arbidol	D	477,424	5,177	1	6	1	132,95	-9,4	-9,4
Astragalin	P	448,38	-0,245	7	11	2	24,24	-11,05	-11,05
Beta-Eudesmol	P	222,372	3,92	1	1	0	4,43	-7,91	-7,91
Cannabidiol	P	314,469	5,847	2	2	1	2,53	-9,89	-9,89
Chloroquine	D	319,88	4,811	1	3	0	2,53	-9,86	-9,86
Crocin	P	976,972	-5,225	14	24	3			
Curcumin	P	368,385	3,37	2	6	0	523,22	-10,11	-10,11
Digitoxigenin	P	374,521	3,604	2	4	0	1,96	-8,68	-8,68
Gossypetin-3-Glucoside	P	480,378	-0,833	9	13	2	2,31	-10,84	-9,61
Gossypetin-3'-O-Glucoside	P	480,378	-0,833	9	13	2	1,8	-9,61	-10,84
Gossypetin-7-O-Glucoside	P	479,37	-1,465	8	13	2	15,38	-8,66	-8,66
Gossypetin-8-Glucoside	P	480,378	-0,833	9	13	2	3,04	-8,84	-8,84
Hydroxychloroquine	D	335,879	3,783	2	4	0	10,5	-8,48	-8,48

Kaempferol	P	286,239	2,282	4	6	0	317,27	-9,89	-9,89
Lopinavir	D	628,814	4,328	4	5	1	507,05	-10,88	-10,88
Luteolin-7-Glucoside	P	448,38	-0,244	7	11	2	15,22	-10,64	-10,64
Native Ligand		684,835	2,099	6	10	2	919,78	-14,03	-14,03
Nelfinavir	D	567,796	4,748	4	6	1	119,44	-11,79	-11,79
Nictoflorin	P	594,522	-1,393	9	15	3	4,41	-10,76	-10,76
Piceid	P	390,388	0,447	6	8	1	207,24	-10,66	-10,66
Quercetin	P	302,238	1,988	5	7	0	80,71	-9,38	-9,38
Quercitrin	P	448,38	0,489	7	11	2	1,96	-10,86	-10,86
Remdesivir	D	602,585	2,312	4	13	2	1,3	-10,81	-10,81
Ribavirin	D	244,207	-3,011	4	8	0	4,18	-6,34	-6,34

Table 4: Rank of compounds regard to Lupinski's rules violation.

Name	Type	Rank	Violation
Nictoflorin	P	1	3
Quercitrin	P	2	2
Luteolin-7-Glucoside	P	3	2
Gossypetin-3'-O-Glucoside	P	4	2
Gossypetin-7-O-Glucoside	P	5	2
Gossypetin-8-Glucoside	P	6	2
Lopinavir	D	7	1
Astragalin	P	8	2
Gossypetin-3-Glucoside	P	9	2
Nelfinavir	D	10	1
Piceid	P	11	1
Remdesivir	D	12	2
Kaempferol	P	13	0
Quercetin	P	14	0
Natural Ligand		15	2
Aloenin	P	16	0
Amiodarone	D	17	2
Arbidol	D	18	1
Curcumin	P	19	0
Digitoxigenin	P	20	0
Ribavirin	D	21	0
Crocin	P	22	3
Hydroxychloroquine	D	23	0
Cannabidiol	P	24	1
Chloroquine	D	25	0
Beta-Eudesmol	P	26	0

G3'G and derivatives docking

G3'G and other gossypetin derivates link at least one of the essential AA of binding site domains I and II of the catalytic center: Cys145 or His41. All others AA of the active site are brought into play according to the ligand but not all together. G3G and G8G are the only ligands that bind critical conserved His41 as well as Cys145. Glu166 is a recurrent linker through all gossypetin and derivatives.

This residue belongs to the sub-sites S1 to S5

Gossypetin-8-Glucoside (G8G) and Gossypetin-7-O-Glucoside (G7G) presents the most bonds to the protein. They established thirteen bonds of several types. All the residues involved in the links with G8G are known residues of the active site. G7G links probably new and non-critical residues: Thr24, Thr26, Thr45, Cys44, Ser46 and His163.

Quercitrin, G3'G and Gossypetin-3-Glucoside (G3G) establish respectively 9, 8 and 7 bonds with known residues of active site. We found that Quercitrin links 3 times with Met165. Figure 2 summarized all of those data.

DISCUSSION

We develop a computational method of prediction of best candidate inhibitors of SARS-CoV-2 3-Chymotrypsin-Like Protease (3CLpro) that plays an essential role in replication of the virus, and represents an effectiveness and nontoxic drug target. This computer-based analysis demonstrates for the first time, a potential benefit of G3'G and derivatives to fight COVID-19.

Covid-19 is a viral infection which fastly spread worldwide, and whose treatment remains an unresolved challenge so far. Among many drugs evaluated, few have proven conclusive clinical efficacy. Furthermore, the spread of the disease mandates that ideal medications against Covid-19 be cheap, and available worldwide. Therefore, there is a rationale to evaluate whether treatments of natural origin from aromatic and medicinal plants have the ability to prevent and/or treat COVID-19. Also, computation-based studies, which have proven efficacy in the screening of new active medications, fulfill the needs for fast analysis of any potential benefit.

Our demonstration of the benefit of Gossypetin derivatives follows a comprehensive, step by step process. First, we targeted 3CLpro, a protein-enzyme which plays a critical role in virus replication. Then, along with putative medicinal plants derived products, we included in our analysis molecules proposed in the Guidelines-version 6 for treatment of COVID-19. We also included compounds that have already been computationally tested. We experiment molecular docking methods which are widely used for this purpose. We used 3 methods: Autodock 4.2.6, AutodockVina version 1.1.2, and Smina version Oct 15 2019. Each method uses a different and inaccurate approximation for a given problem. Either program can provide a better result. For example, AutoDock is better in discriminating ligands and decoys in more hydrophobic, poorly polar and poorly charged pockets, while Vina tend to give better results for polar and charged binding Pockets [33]. So we compiled those results using a consensus scoring approach.

The computational study is based on a number of repetitions of tests and the compilation of 3 prediction methods. In order to propose the most representative score docking of molecules, we perform fifty runs per ligand with Autodock 4.2.6 and Smina method and the maximum of twenty runs performed for AutodockVina. The ranking of candidates on the basis of free binding energy with respect to violations of Lipinski's rules and other physicochemical binding properties is common but probably not sufficient to offer ready-to-use molecules.

But a deep inside comparative study of Traditional Chinese Medecine versus conventional medicine tend to prove the efficacy of using natural compounds by traditional habits [4,34]

Quercitrin and some of its derivatives such as Nictoflorin, Luteolin-7-glucoside or Astragalin, and all Gossypetin derivatives have the best free binding energy. According to our study, they are good therapeutic candidates. However, Nictoflorin presents 3 violations of Lupinski's rules whereas all the others present only 2. This reduces Nictoflorin claims of oral absorption. So therefore, Gossypetins remain the most likely natural compounds candidate for treatment of COVID-19. Their inhibition constant, their dissolving energy and the quantity of predicted H-bonds are among the best of our dataset.

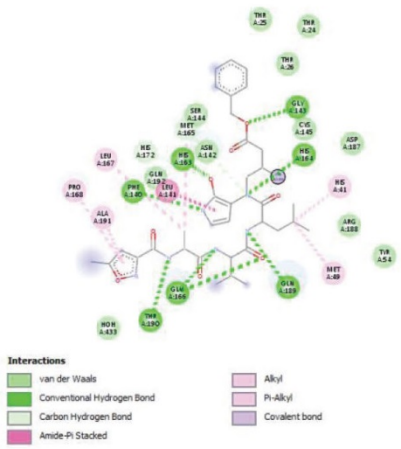
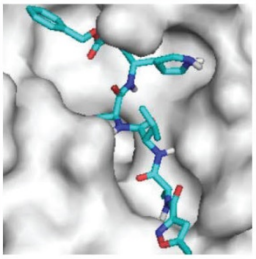
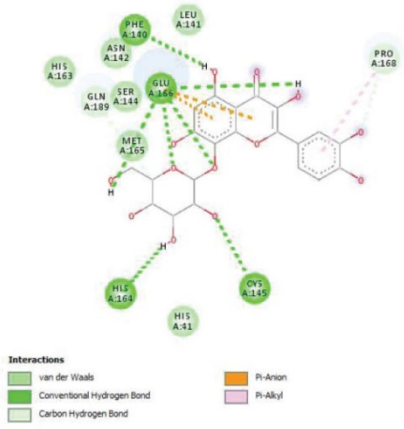
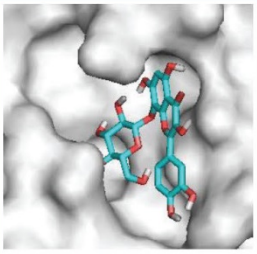
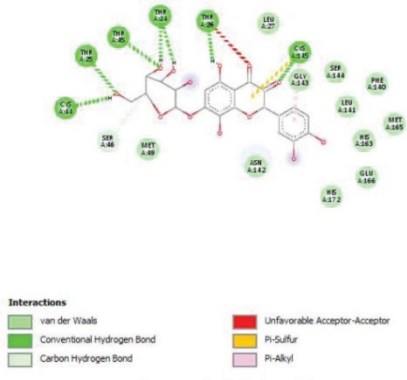
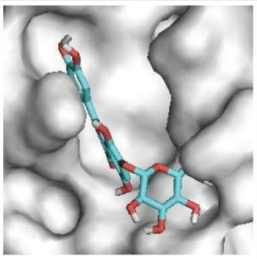
Potential of G3'G and derivatives as a Covid-19 treatment, is reinforced by the visualization of molecular interactions between the active site of 3CLpro/all Gossypetin derivatives. Those interactions, rendered possible by our successful 3-D modeling of Gossypetin-3'-O-glucoside (G3'G), shows that all these molecules interact with AA of the catalytic center. Glu166, which belongs to the S1 to S5 sub-sites, is systematically bounded. And G3G and G8G bind the critical conserved His41 and Cys145 residues of the catalytic center.

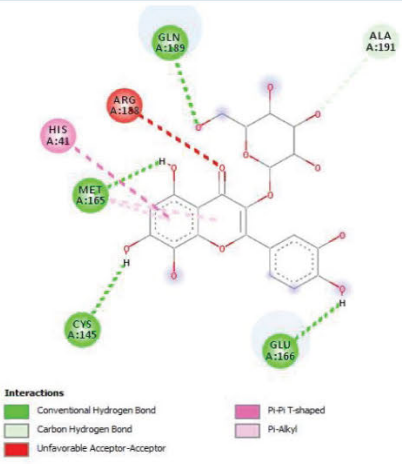
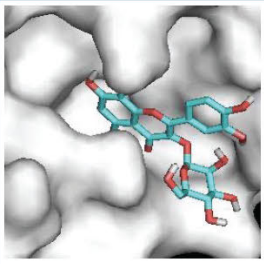
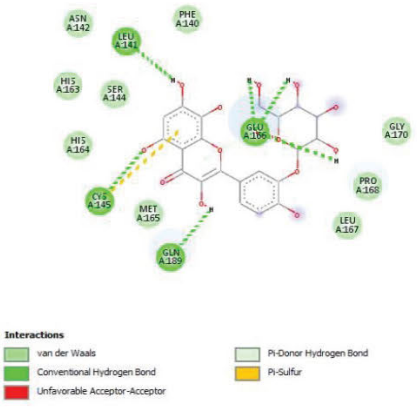
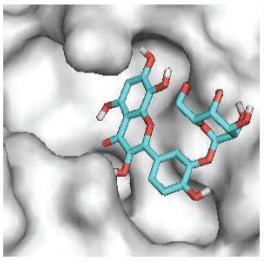
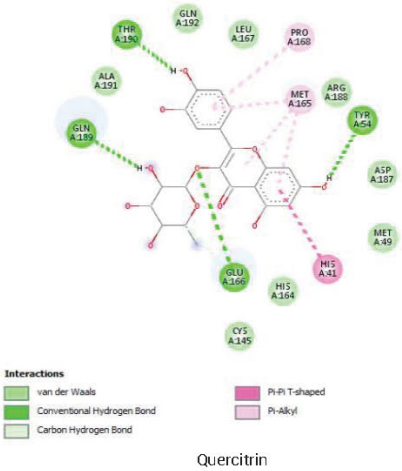
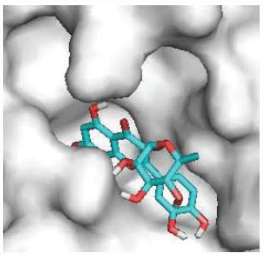
For all of those reasons we propose G8G, G3G, G3'G and finally G7G as good tropical natural compounds candidate that should be further investigated to prevent or treat COVID19. We were able to show that molecules derived from medicinal plants have better physico-chemical interaction properties and potential for inhibition of 3CLpro, as compared to some candidate drugs proposed for the treatment of COVID19.

Our computation analysis matches with previous analyses. We were able to verify that according to their ΔG, Digitoxigenin>Beta-eudesmol, as suggested by Aanouz et al. Our method does not allow us to compare Crocin with them. Similarly, in accordance with the experience of Khaerunnisa et al., Nelfinavir/Lopinavir>Kaempferol/Quercetin. However our classification methods propose a lower consensus-score of ΔG for Luteolin-7-Glucoside. Several reasons can explain this difference: 1. we worked with protonated protein and ligands, 2. The coordinates of the grid are generally not communicated by others, and probably significantly different from one experiment to another, 3. Our method for minimizing the ligand are different and the choice of conformer used for experiments may vary from one experimenter to another. Martinique is a tropical caribbean island. Biodiversity is a pride and a wealth. The population still practices traditional medication with the use of local medicinal plants.

Only four flavonoid glycosides have been found in nature derivatives from gossypetin: G8G or gossypin (H. *vitifolius* and H. *sabdariffa*); G7G or gossypitrin/gossypetrin

Figure 2: Binding residues and docking poses: Native ligand and G3'G/derivatives.

Compound	Bindingresidue	Pose
 Native ligand	**H-bonds** : Gly143, Phe140, His163, His164, Glu166, Gln189 and Thr190 **Amide Pi-stacked** : Leu141 **pi-alkyle** : His41, Met49, Leu167, Pro168 and Ala191 **Carbones-hydrogen** : Met165 and His172 **Van der Walls** : 11 AA including Asn142, Asp187, Arg188 and Tyr54	
 Gossypetin-8-glucoside	**H-bonds** :Phe140, His164, Glu166, et Cys145 **Pi-anion** : Glu166 **Pi-alkyle** : Pro168 **Carbones-hydrogen** : Gln189 et Pro168 **Van der Walls** : Asn142, His163, Leu141, Met165, et His41	
 Gossypetin-7-O-glucoside	**H-bonds** :Thr24, Thr25, Thr26, Thr45, Cys44 et Cys145 **pi-Sulfur** :Cys145 **Pi-alkyle** :Gly143 **Carbones-hydrogen** : Ser46 **Van der Walls** : 10 AA including Gly143, Phe140, His163, Glu166	

 Gossypetin-3-glucoside	**H-bonds** :Gln189, Met165, Glu166 et Cys145 **Pi-Pi T-shaped** :His41 **Pi-alkyle** :Met165 **Carbones-hydrogen** :Ala191	
 Gossypetin-3'-O-glucoside	**H-bonds :** Gln189, Glu166, Leu141 and Cys145 **Pi-Sulfur**Cys145 **Van der Walls :** 9 AA including Phe140, His163 and His164.	
 Quercitrin	**H-bonds** : Gln189, Glu166, Thr190 and Tyr54 **Pi-Pi T-shaped** : His41 **Pi-alkyle** : Pro168, Met165 **Van der Walls** : 8 AA including Gln192, Arg188, and His164.	

(H. *sabdariffa*, T. *elatum* and T. *tiliaceum*); G3G or gossytrin (H. *sabdariffa* and T. *tiliaceum*) and G3'G (A. *manihot* and T. *elatum*). For the first time G3'G was extracted and characterized from petals of flowers of Taliparitі *elatum Sw.* of Martinica [17]. The high solubility of G3'G in the aqueous phase makes it, like many other molecules derived from medicinal plants, a good candidate for treatments. Its antifungal and antioxidant properties are already known. His anti-SARS-COV-2 properties have never been explored.

CONCLUSION

The computational study that we have carried out confirms the interest of polyphenols in the treatment of SARS-COV2 by inhibition of 3CLpro. In particular, the Gossypetin derivatives present in tropical Hibiscus varieties are for the 1st time classified according to a predictive in sillico method, on the same level as Quercitrin or luteolin-7-Glucoside proposed by Indian research teams, and present a priori a better inhibitory potential than Lopinavir / Nelfinavir treatments against 3CLpro. The computational study that we have carried out confirms the interest of flavonoids in the treatment of SARS-COV-2 by inhibition of 3CLpro. It encourages research on local medicinal plants for the treatment of COVID-19. In particular, the secondary metabolites of Gossypetin extracted from tropical varieties of Hibiscus appear promising. Like all computational studies, our study should be supplemented with in vivo experiments to refine the therapeutic proposals. It could be the support for proposing local therapeutic or preventive trials and epidemiologic studies.

To go further in proposing new treatments, the molecules derived from medicinal plants seem interesting. We could broaden our screening of local medicinal plants compounds candidates. We could target another SARS-COV2 element such as other enzymes or specific structural proteins. Molecular dynamic study could also be the next step to confirm results or evaluate combination of therapeutics. In vivo experiments remain essential to confirm therapeutic efficacy.

ACKNOWLEDGMENT

Jocelyn INAMO, giving us valuable reviewing on discussion and Max MONAN advises for information about Gossypetins molecules and molecular chemistry.

References

1. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. Nat Rev Microbiol. 2009 Mar;7(3):226-36. doi: 10.1038/nrmicro2090. Epub 2009 Feb 9. PMID: 19198616; PMCID: PMC2750777.
2. Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. J Biomol Struct Dyn. 2020 Apr 13:1-10. doi: 10.1080/07391102.2020.1751298. Epub ahead of print. PMID: 32238094.
3. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, Wang MH, Cai Y, Wang W, Yang L, He D. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis. 2020 Apr;93:211-216. doi: 10.1016/j.ijid.2020.02.058. Epub 2020 Mar 4. PMID: 32145465; PMCID: PMC7102659.
4. Chen Z, Nakamura T. Statistical evidence for the usefulness of Chinese medicine in the treatment of SARS. Phytother Res. 2004 Jul;18(7):592-4. doi: 10.1002/ptr.1485. PMID: 15305324.
5. Tahir Ul Qamar M, Alqahtani SM, Alamri MA, Chen LL. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. J Pharm Anal. 2020 Aug;10(4):313-319. doi: 10.1016/j.jpha.2020.03.009. Epub 2020 Mar 26. PMID: 32296570; PMCID: PMC7156227.
6. Anand K, Yang H, Bartlam M, Rao Z, Hilgenfeld R. Coronavirus main proteinase: target for antiviral drug therapy. Coronaviruses with Special Emphasis on First Insights Concerning SARS. 2005:173–99. doi: 10.1007/3-7643-7339-3_9. PMCID: PMC7123552.
7. Yang Y, Zhang L, Geng H, Deng Y, Huang B, Guo Y, Zhao Z, Tan W. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. Protein Cell. 2013 Dec;4(12):951-61. doi: 10.1007/s13238-013-3096-8. Epub 2013 Dec 8. PMID: 24318862; PMCID: PMC4875403.
8. Naksuk N, Lazar S, Peeraphatdit TB. Cardiac safety of off-label COVID-19 drug therapy: a review and proposed monitoring protocol. Eur Heart J Acute Cardiovasc Care. 2020 Apr;9(3):215-221. doi: 10.1177/2048872620922784. Epub 2020 May 6. PMID: 32372695; PMCID: PMC7235441.
9. Ruamviboonsuk P, Lai TYY, Chang A, Lai CC, Mieler WF, Lam DSC; for Asia-Pacific Vitreo-Retina Society. Chloroquine and Hydroxychloroquine Retinal Toxicity Consideration in the Treatment of COVID-19. Asia Pac J Ophthalmol (Phila). 2020 Mar-Apr;9(2):85-87. doi: 10.1097/APO.0000000000000289. PMID: 32349115; PMCID: PMC7227199.
10. S Khaerunnisa, H Kurniawan, R Awaluddin, S Suhartati, S Soetjipto. Potential Inhibitor of COVID-19 Main Protease (M^{pro}) From Several Medicinal Plant Compounds by Molecular Docking Study. Medicine & Pharmacology, preprint, mars. 2020. doi: 10.20944/preprints202003.0226.v1.
11. AK Srivastava, A Kumar, N Misra. On the Inhibition of COVID-19 Protease by Indian Herbal Plants: An In Silico Investigation. arXiv. 2020. https://tinyurl.com/y5u8umsd
12. Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov Ther. 2020;14(1):58-60. doi: 10.5582/ddt.2020.01012. PMID: 32147628.
13. Wang G, Zhu W. Molecular docking for drug discovery and development: a widely used approach but far from perfect. Future Med Chem. 2016 Sep;8(14):1707-10. doi: 10.4155/fmc-2016-0143. Epub 2016 Aug 31. PMID: 27578269.
14. Aanouz I, Belhassan A, El-Khatabi K, Lakhlifi T, El-Ldrissi M, Bouachrine M. Moroccan Medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations. J Biomol Struct Dyn. 2020 May 6:1-9. doi: 10.1080/07391102.2020.1758790. Epub ahead of print. PMID: 32306860; PMCID: PMC7212546.
15. Joshi T, Joshi T, Sharma P, Mathpal S, Pundir H, Bhatt V, Chandra S. In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking. Eur Rev Med Pharmacol Sci. 2020 Apr;24(8):4529-4536. doi: 10.26355/eurrev_202004_21036. PMID: 32373991.
16. GA Gyebi, OB Ogunro, AP Adegunloye, OM Ogunyemi, SO Afolabi. Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CL pro): An in silico screening of alkaloids and terpenoids from African medicinal plants. J Biomol Struct Dyn. 2020;1-13. doi: 10.1080/07391102.2020.1764868.
17. Frantz François-Haugrin, Max Monan, Emmanuel Nossin, Juliette Smith-Ravin, Odile Marcelin. Antioxidant activity of an isomer of gossypitrin (gossypetin-3'-O-glucoside) isolated in the petals of Taliparitі elatum Sw., and determination of total phenolic content of the total flower. J Pharmacogn Phytochem 2016;5(5):200-208. https://tinyurl.com/yybspabz
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235-42. doi: 10.1093/nar/28.1.235. PMID: 10592235; PMCID: PMC102472.
19. Xu Z, Peng C, Shi Y, Zhu Z, Mu K, Wang X, Zhu W. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. bioRxiv; 2020. doi: 10.1101/2020.01.27.921627.

20. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H. The crytal structure of 2019-nCoV main protease in complex with an inhibitor N3. Nature. 2020;582:289-293. https://tinyurl.com/yyw4p86s

21. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. Nucleic Acids Res. 2016 Jan 4;44(D1):D1202-13. doi: 10.1093/nar/gkv951. Epub 2015 Sep 22. PMID: 26400175; PMCID: PMC4702940.

22. Lowe DM, Corbett PT, Murray-Rust P, Glen RC. Chemical name to structure: OPSIN, an open source solution. Journal of Chemical Information and Modeling. 2011 Mar;51(3):739-753. DOI: 10.1021/ci100384d.

23. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011 Oct 7;3:33. doi: 10.1186/1758-2946-3-33. PMID: 21982300; PMCID: PMC3198950.

24. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001 Mar 1;46(1-3):3-26. doi: 10.1016/s0169-409x(00)00129-0. PMID: 11259830.

25. Huang B. MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS. 2009 Aug;13(4):325-30. doi: 10.1089/omi.2009.0045. PMID: 19645590.

26. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W665-7. doi: 10.1093/nar/gkh381. PMID: 15215472; PMCID: PMC441519.

27. MHM Olsson, CR Sondergaard, M Rostkowski, JH Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. *J Chem Theory Comput. 2011*;525-537. doi: 10.1021/ct100578z.

28. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec;30(16):2785-91. doi: 10.1002/jcc.21256. PMID: 19399780; PMCID: PMC2760638.

29. Xu W, Lucke AJ, Fairlie DP. Comparing sixteen scoring functions for predicting biological activities of ligands for protein targets. J Mol Graph Model. 2015 Apr;57:76-88. doi: 10.1016/j.jmgm.2015.01.009. Epub 2015 Feb 1. PMID: 25682361.

30. Quiroga R, Villarreal MA. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. PLoS One. 2016 May 12;11(5):e0155183. doi: 10.1371/journal.pone.0155183. PMID: 27171006; PMCID: PMC4865195.

31. Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J Chem Inf Model. 2013 Aug 26;53(8):1893-904. doi: 10.1021/ci300604z. Epub 2013 Feb 12. PMID: 23379370; PMCID: PMC3726561.

32. Wang R, Wang S. How does consensus scoring work for virtual library screening? An idealized computer experiment. J Chem Inf Comput Sci. 2001 Sep-Oct;41(5):1422-6. doi: 10.1021/ci010025x. PMID: 11604043.

33. TF Vieira, SF Sousa. Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. Appl Sci. 2019;4538. doi: 10.3390/app9214538.

34. Y Yang, MS Islam, J Wang, Y Li, X Chen. Traditional Chinese Medicine in the Treatment of Patients Infected with 2019-New Coronavirus (SARS-CoV-2): A Review and Perspective. Int J Biol Sci. 2020;1708-1717. doi: 10.7150/ijbs.45538.

How to cite this article: Giguet Valard AG, Raguette K, Morin S, Bellance R, Ravin JS. Gossypetin Derivatives are also Putative Inhibitors of SARS-COV 2: Results of a Computational Study. J Biomed Res Environ Sci. 2020 Oct 09; 1(6): 201-212. doi: 10.37871/jbres1144, Article ID: JBRES1144

BIBLIOGRAPHIE

[1] « NEURODEGENERESCENCE ». [En ligne]. Disponible sur: http://www.neuroland.biz/initiativeMNM/.
[2] R. Bellance, « Le Centre de Référence Caribéen des maladies neuromusculaires et neurologiques rares », *médecine/sciences*, vol. 31, p. 45‑47, nov. 2015, doi: 10.1051/medsci/201531s313.
[3] P. Delobel, O. Leroy, M. Hamdane, A. v. Sambo, A. Delacourte, et L. Buée, « Proteasome inhibition and Tau proteolysis: an unexpected regulation », *FEBS Lett.*, vol. 579, nº 1, p. 1‑5, janv. 2005, doi: 10.1016/j.febslet.2004.11.018.
[4] A. S. Vieira, D. B. Dogini, et I. Lopes-Cendes, « Role of non-coding RNAs in non-aging-related neurological disorders », *Braz. J. Med. Biol. Res.*, vol. 51, nº 8, juin 2018, doi: 10.1590/1414-431x20187566.
[5] G. K. Tofaris et N. J. Buckley, « Convergent molecular defects underpin diverse neurodegenerative diseases », *J. Neurol. Neurosurg. Psychiatry*, p. jnnp-2017-316988, févr. 2018, doi: 10.1136/jnnp-2017-316988.
[6] A. Lannuzel *et al.*, « Atypical parkinsonism in Guadeloupe: a common risk factor for two closely related phenotypes? », *Brain*, vol. 130, nº 3, p. 816‑827, mars 2007, doi: 10.1093/brain/awl347.
[7] D. Caparros-Lefebvre *et al.*, « Guadeloupean parkinsonism: a cluster of progressive supranuclear palsy-like tauopathy », *Brain*, vol. 125, nº 4, p. 801‑811, avr. 2002, doi: 10.1093/brain/awf086.
[8] A. Lannuzel *et al.*, « Clinical varieties and epidemiological aspects of amyotrophic lateral sclerosis in the Caribbean island of Guadeloupe: A new focus of ALS associated with Parkinsonism », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 16, nº 3‑4, p. 216‑223, avr. 2015, doi: 10.3109/21678421.2014.992026.
[9] A.-G. Valard-Giguet *et al.*, « Capacity building in clinical and molecular characterization of rare neurodegenerative diseases through the development of a local register and improved use of bioinformatics data analysis in Martinique », *J. Glob. Health Rep.*, mars 2020, doi: 10.29392/001c.11877.
[10] H. Liu *et al.*, « Next-generation sequencing in a series of 80 fetuses with complex cardiac malformations and/or heterotaxy », *Hum. Mutat.*, p. humu.24132, nov. 2020, doi: 10.1002/humu.24132.
[11] J.-Y. Le Gall et R. Ardaillou, « Biologie du vieillissement », *Bull. Académie Natl. Médecine*, vol. 193, nº 2, p. 365‑404, févr. 2009, doi: 10.1016/S0001-4079(19)32590-7.
[12] K. Usdin, « The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases », *Genome Res.*, vol. 18, nº 7, p. 1011‑1019, juill. 2008, doi: 10.1101/gr.070409.107.
[13] A.-S. Lebre et A. Brice, « Maladies par expansion de polyglutamine : données moléculaires et physiopathologiques », *médecine/sciences*, vol. 17, nº 11, p. 1149‑1157, nov. 2001, doi: 10.1051/medsci/200117111149.
[14] M. Koegl, T. Hoppe, S. Schlenker, H. D. Ulrich, T. U. Mayer, et S. Jentsch, « A Novel Ubiquitination Factor, E4, Is Involved in Multiubiquitin Chain Assembly », *Cell*, vol. 96, nº 5, p. 635‑644, mars 1999, doi: 10.1016/S0092-8674(00)80574-7.
[15] P. Sweeney *et al.*, « Protein misfolding in neurodegenerative diseases: implications and strategies », *Transl. Neurodegener.*, vol. 6, nº 1, p. 6, déc. 2017, doi: 10.1186/s40035-017-0077-5.
[16] M. S. Forman, J. Q. Trojanowski, et V. M.-Y. Lee, « Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs », *Nat. Med.*, vol. 10, nº 10, p. 1055‑1063, oct. 2004, doi: 10.1038/nm1113.
[17] V. Kumar, N. Sami, T. Kashav, A. Islam, F. Ahmad, et Md. I. Hassan, « Protein aggregation and neurodegenerative diseases: From theory to therapy », *Eur. J. Med. Chem.*, vol. 124, p. 1105‑1120, nov. 2016, doi: 10.1016/j.ejmech.2016.07.054.
[18] S. Gilad *et al.*, « Serum MicroRNAs Are Promising Novel Biomarkers », *PLoS ONE*, vol. 3, nº 9, p. e3148, sept. 2008, doi: 10.1371/journal.pone.0003148.
[19] V. S. Patil, R. Zhou, et T. M. Rana, « Gene regulation by non-coding RNAs », *Crit. Rev. Biochem. Mol. Biol.*, vol. 49, nº 1, p. 16‑32, janv. 2014, doi: 10.3109/10409238.2013.844092.
[20] J. K. Andersen, « Oxidative stress in neurodegeneration: cause or consequence? », *Nat. Med.*, vol. 10, nº S7, p. S18‑S25, juill. 2004, doi: 10.1038/nrn1434.

[21] C. Moreau ∗ , L. Defebvre, « Troubles de la marche », *AKOS Traité Médecine*, [En ligne]. Disponible sur: https://www.em-consulte.com/article/1068240/troubles-de-la-marche.
[22] « Territorialisation des vertèbres ». [En ligne]. Disponible sur: https://tel.archives-ouvertes.fr/tel-00855512).
[23] *Aids to the examination of the peripheral nervous system*. .
[24] E. H. Franssen, L. E. M. Somen, C. L. Torossian, et B. Reisberg, « Equilibrium and Limb Coordination in Mild Cognitive Impairment and Mild Alzheimer's Disease », *J. Am. Geriatr. Soc.*, vol. 47, nº 4, p. 463‑469, avr. 1999, doi: 10.1111/j.1532-5415.1999.tb07240.x.
[25] F. J. Huff, F. Boller, F. Lucchelli, R. Querriera, J. Beyer, et S. Belle, « The Neurologic Examination in Patients With Probable Alzheimer's Disease », *Arch. Neurol.*, vol. 44, nº 9, p. 929‑932, sept. 1987, doi: 10.1001/archneur.1987.00520210031015.
[26] A. Kluger *et al.*, « Patterns of Motor Impairment in Normal Aging, Mild Cognitive Decline, and Early Alzheimer' Disease », *J. Gerontol. B. Psychol. Sci. Soc. Sci.*, vol. 52B, nº 1, p. P28‑P39, janv. 1997, doi: 10.1093/geronb/52B.1.P28.
[27] R. J. Miller, J. Snowdon, et R. Vaughan, « The Use of the Cohen-Mansfield Agitation Inventory in the Assessment of Behavioral Disorders in Nursing Homes », *J. Am. Geriatr. Soc.*, vol. 43, nº 5, p. 546‑549, mai 1995, doi: 10.1111/j.1532-5415.1995.tb06104.x.
[28] « Orientation diagbnostic devant un trouble de la marche ». [En ligne]. Disponible sur: http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/neuro/troublspe/340/lecon340.htm.
[29] T. Schmitz-Hubsch *et al.*, « Scale for the assessment and rating of ataxia: Development of a new clinical scale », *Neurology*, vol. 66, nº 11, p. 1717‑1720, juin 2006, doi: 10.1212/01.wnl.0000219042.60538.92.
[30] M. Assadi, P. Leone, J. J. Veloski, R. J. Schwartzman, C. G. Janson, et J. V. Campellone, « Validating an Ataxia Functional Composite Scale in spinocerebellar ataxia », *J. Neurol. Sci.*, vol. 268, nº 1‑2, p. 136‑139, mai 2008, doi: 10.1016/j.jns.2007.11.016.
[31] A. Durr, « Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond », *Lancet Neurol.*, vol. 9, nº 9, p. 885‑894, sept. 2010, doi: 10.1016/S1474-4422(10)70183-6.
[32] B. L. Fogel et S. Perlman, « Clinical features and molecular genetics of autosomal recessive cerebellar ataxias », *Lancet Neurol.*, vol. 6, nº 3, p. 245‑257, mars 2007, doi: 10.1016/S1474-4422(07)70054-6.
[33] J. Finsterer et S. Zarrouk-Mahjoub, « Mitochondrial ataxia is genetically and phenotypically heterogeneous », *CNS Neurosci. Ther.*, vol. 24, nº 12, p. 1301‑1302, déc. 2018, doi: 10.1111/cns.13031.
[34] A. Dürr *et al.*, « Autosomal dominant cerebellar ataxia type I in Martinique (French West Indies): Clinical and neuropathological analysis of 53 patients from three unrealted SCA2 families », *Brain*, vol. 118, nº 6, p. 1573‑1581, 1995, doi: 10.1093/brain/118.6.1573.
[35] A. Lezin *et al.*, « Autosomal dominant cerebellar ataxia type I in Martinique (French West Indies): Genetic analysis of three unrelated SCA2 families », *Hum. Genet.*, vol. 97, nº 5, p. 671‑676, mai 1996, doi: 10.1007/BF02281881.
[36] F. O. Walker, « Huntington's disease », *The Lancet*, vol. 369, nº 9557, p. 218‑228, janv. 2007, doi: 10.1016/S0140-6736(07)60111-1.
[37] S. E. Holmes *et al.*, « A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2 », *Nat. Genet.*, vol. 29, nº 4, p. 377‑378, déc. 2001, doi: 10.1038/ng760.
[38] S. A. Schneider, K. E. Marshall, J. Xiao, et M. S. LeDoux, « JPH3 repeat expansions cause a progressive akinetic-rigid syndrome with severe dementia and putaminal rim in a five-generation African-American family », *neurogenetics*, vol. 13, nº 2, p. 133‑140, mai 2012, doi: 10.1007/s10048-012-0318-9.
[39] A. Krause *et al.*, « Junctophilin 3 (*JPH3*) expansion mutations causing Huntington disease like 2 (HDL2) are common in South African patients with African ancestry and a Huntington disease phenotype: JPH3 Mutations Causing HD Phenotype in Africa », *Am.*

J. Med. Genet. B Neuropsychiatr. Genet., vol. 168, nº 7, p. 573‑585, oct. 2015, doi: 10.1002/ajmg.b.32332.
[40] F. K. Baine, A. Krause, et L. J. Greenberg, « The Frequency of Huntington Disease and Huntington Disease-Like 2 in the South African Population », *Neuroepidemiology*, vol. 46, nº 3, p. 198‑202, 2016, doi: 10.1159/000444020.
[41] A. Y. Al-Tahan *et al.*, « A novel autosomal recessive "Huntington's disease-like" neurodegenerative disorder in a Saudi family », *Saudi Med. J.*, vol. 20, nº 1, p. 85‑89, janv. 1999.
[42] A. Ludolph *et al.*, « A revision of the El Escorial criteria - 2015 », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 16, nº 5‑6, p. 291‑292, août 2015, doi: 10.3109/21678421.2015.1049183.
[43] P. V. S. de Souza, W. B. V. de R. Pinto, M. A. T. Chieia, et A. S. B. Oliveira, « Clinical and genetic basis of familial amyotrophic lateral sclerosis », *Arq. Neuropsiquiatr.*, vol. 73, nº 12, p. 1026‑1037, déc. 2015, doi: 10.1590/0004-282X20150161.
[44] J. Costa, M. Swash, et M. de Carvalho, « Awaji Criteria for the Diagnosis of Amyotrophic Lateral Sclerosis: A Systematic Review », *Arch. Neurol.*, vol. 69, nº 11, p. 1410, nov. 2012, doi: 10.1001/archneurol.2012.254.
[45] B. R. Brooks, « El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors », *J. Neurol. Sci.*, vol. 124 Suppl, p. 96‑107, juill. 1994.
[46] F. Agosta *et al.*, « The El Escorial criteria: Strengths and weaknesses », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 16, nº 1‑2, p. 1‑7, mars 2015, doi: 10.3109/21678421.2014.964258.
[47] A. Al-Chalabi et O. Hardiman, « The epidemiology of ALS: a conspiracy of genes, environment and time », *Nat. Rev. Neurol.*, vol. 9, nº 11, p. 617‑628, oct. 2013, doi: 10.1038/nrneurol.2013.203.
[48] B. Marin *et al.*, « Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis », *Int. J. Epidemiol.*, p. dyw061, mai 2016, doi: 10.1093/ije/dyw061.
[49] G. Logroscino *et al.*, « Incidence of amyotrophic lateral sclerosis in Europe », *J. Neurol. Neurosurg. Psychiatry*, vol. 81, nº 4, p. 385‑390, avr. 2010, doi: 10.1136/jnnp.2009.183525.
[50] L. T. Kurland et D. W. Mulder, « Epidemiologic investigations of amyotrophic lateral sclerosis. 2. Familial aggregations indicative of dominant inheritance. I », *Neurology*, vol. 5, nº 3, p. 182‑196, mars 1955.
[51] M. Vidailhet, « Hétérogénéité de la maladie de Parkinson », *Bull. Académie Natl. Médecine*, vol. 187, nº 2, p. 259‑276, févr. 2003, doi: 10.1016/S0001-4079(19)34048-8.
[52] T. Shimohata, T. Ozawa, H. Nakayama, M. Tomita, H. Shinoda, et M. Nishizawa, « Frequency of nocturnal sudden death in patients with multiple system atrophy », *J. Neurol.*, vol. 255, nº 10, p. 1483‑1485, oct. 2008, doi: 10.1007/s00415-008-0941-4.
[53] S. Gilman *et al.*, « Second consensus statement on the diagnosis of multiple system atrophy », *Neurology*, vol. 71, nº 9, p. 670‑676, août 2008, doi: 10.1212/01.wnl.0000324625.00404.15.
[54] M. I. Papp, J. E. Kahn, et P. L. Lantos, « Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome) », *J. Neurol. Sci.*, vol. 94, nº 1‑3, p. 79‑100, déc. 1989, doi: 10.1016/0022-510X(89)90219-0.
[55] T. Ozawa, « The spectrum of pathological involvement of the striatonigral and olivopontocerebellar systems in multiple system atrophy: clinicopathological correlations », *Brain*, vol. 127, nº 12, p. 2657‑2671, nov. 2004, doi: 10.1093/brain/awh303.
[56] I. Litvan *et al.*, « Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop », *Neurology*, vol. 47, nº 1, p. 1‑9, juill. 1996, doi: 10.1212/WNL.47.1.1.
[57] G. Respondek *et al.*, « Accuracy of the national institute for neurological disorders and stroke/society for progressive supranuclear palsy and neuroprotection and natural

history in Parkinson plus syndromes criteria for the diagnosis of progressive supranuclear palsy: PSP Diagnostic Criteria », *Mov. Disord.*, vol. 28, nº 4, p. 504‑509, avr. 2013, doi: 10.1002/mds.25327.

[58] S. Lehericy *et al.*, « Magnetic resonance imaging lesion pattern in Guadeloupean parkinsonism is distinct from progressive supranuclear palsy », *Brain*, vol. 133, nº 8, p. 2410‑2425, août 2010, doi: 10.1093/brain/awq162.

[59] A. Lannuzel *et al.*, « Further evidence for a distinctive atypical degenerative parkinsonism in the Caribbean: A new cluster in the French West Indian Island of Martinique », *J. Neurol. Sci.*, vol. 388, p. 214‑219, mai 2018, doi: 10.1016/j.jns.2018.02.019.

[60] D. Caparros Lefebvre, « Food toxins and the Caribbean Parkinson plus types », *Rev. Neurol. (Paris)*, vol. 175, nº 10, p. 641‑643, déc. 2019, doi: 10.1016/j.neurol.2019.07.016.

[61] A. Bjornsdottir, G. Gudmundsson, H. Blondal, et E. Olafsson, « Incidence and prevalence of multiple system atrophy: a nationwide study in Iceland », *J. Neurol. Neurosurg. Psychiatry*, vol. 84, nº 2, p. 136‑140, févr. 2013, doi: 10.1136/jnnp-2012-302500.

[62] G. K. Wenning *et al.*, « The natural history of multiple system atrophy: a prospective European cohort study », *Lancet Neurol.*, vol. 12, nº 3, p. 264‑274, mars 2013, doi: 10.1016/S1474-4422(12)70327-7.

[63] Y. Ben-Shlomo, G. K. Wenning, F. Tison, et N. P. Quinn, « Survival of patients with pathologically proven multiple system atrophy: A meta-analysis », *Neurology*, vol. 48, nº 2, p. 384‑393, févr. 1997, doi: 10.1212/WNL.48.2.384.

[64] M. M. Mouradian, « Recent advances in the genetics and pathogenesis of Parkinson disease », *Neurology*, vol. 58, nº 2, p. 179‑185, janv. 2002, doi: 10.1212/WNL.58.2.179.

[65] S. Kato, M. Oda, H. Hayashi, A. Kawata, et T. Shimizu, « Participation of the limbic system and its associated areas in the dementia of amyotrophic lateral sclerosis », *J. Neurol. Sci.*, vol. 126, nº 1, p. 62‑69, oct. 1994, doi: 10.1016/0022-510X(94)90095-7.

[66] A. Chio *et al.*, « Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72 », *Brain*, vol. 135, nº 3, p. 784‑793, mars 2012, doi: 10.1093/brain/awr366.

[67] A. S. L. Ng, R. Rademakers, et B. L. Miller, « Frontotemporal dementia: a bridge between dementia and neuromuscular disease: Links between FTD and neuromuscular disease », *Ann. N. Y. Acad. Sci.*, vol. 1338, nº 1, p. 71‑93, mars 2015, doi: 10.1111/nyas.12638.

[68] Y. E. Kim *et al.*, « SCA2 family presenting as typical Parkinson's disease: 34 year follow up », *Parkinsonism Relat. Disord.*, vol. 40, p. 69‑72, juill. 2017, doi: 10.1016/j.parkreldis.2017.04.003.

[69] M. Zhang *et al.*, « *C9orf72* and *ATXN2* repeat expansions coexist in a family with ataxia, dementia, and parkinsonism: *C9orf72* AND *ATXN2* In Ataxia, Dementia, and Parkinsonism », *Mov. Disord.*, vol. 32, nº 1, p. 158‑162, janv. 2017, doi: 10.1002/mds.26841.

[70] Neuro-CEB Neuropathology Network *et al.*, « Interrupted CAG expansions in ATXN2 gene expand the genetic spectrum of frontotemporal dementias », *Acta Neuropathol. Commun.*, vol. 6, nº 1, déc. 2018, doi: 10.1186/s40478-018-0547-8.

[71] A. Chio *et al.*, « ATXN2 polyQ intermediate repeats are a modifier of ALS survival », *Neurology*, vol. 84, nº 3, p. 251‑258, janv. 2015, doi: 10.1212/WNL.0000000000001159.

[72] J. M. Shoffner *et al.*, « Mitochondrial DNA Variants Observed in Alzheimer Disease and Parkinson Disease Patients », *Genomics*, vol. 17, nº 1, p. 171‑184, juill. 1993, doi: 10.1006/geno.1993.1299.

[73] L. R. McQuade, A. Balachandran, H. A. Scott, S. Khaira, M. S. Baker, et U. Schmidt, « Proteomics of Huntington's Disease-Affected Human Embryonic Stem Cells Reveals an Evolving Pathology Involving Mitochondrial Dysfunction and Metabolic Disturbances », *J. Proteome Res.*, vol. 13, nº 12, p. 5648‑5659, déc. 2014, doi: 10.1021/pr500649m.

[74] L. Zeng *et al.*, « Differential recruitment of UBQLN2 to nuclear inclusions in the polyglutamine diseases HD and SCA3 », *Neurobiol. Dis.*, vol. 82, p. 281‑288, oct. 2015, doi: 10.1016/j.nbd.2015.06.017.

[75] M. R. Turner, J. Barnwell, A. Al-Chalabi, et A. Eisen, « Young-onset amyotrophic lateral sclerosis: historical and other observations », *Brain*, vol. 135, nº 9, p. 2883‑2891, sept. 2012, doi: 10.1093/brain/aws144.
[76] K. Iwai, « Anticipation in familial amyotrophic lateral sclerosis with SOD1-G93S mutation », *J. Neurol. Neurosurg. Psychiatry*, vol. 72, nº 6, p. 819‑820, juin 2002, doi: 10.1136/jnnp.72.6.819.
[77] J. K. Teer et J. C. Mullikin, « Exome sequencing: the sweet spot before whole genomes », *Hum. Mol. Genet.*, vol. 19, nº R2, p. R145‑R151, oct. 2010, doi: 10.1093/hmg/ddq333.
[78] I. Leong, J. Skinner, et D. Love, « Application of Massively Parallel Sequencing in the Clinical Diagnostic Testing of Inherited Cardiac Conditions », *Med. Sci.*, vol. 2, nº 2, p. 98‑126, juin 2014, doi: 10.3390/medsci2020098.
[79] F. Mertes *et al.*, « Targeted enrichment of genomic DNA regions for next-generation sequencing », *Brief. Funct. Genomics*, vol. 10, nº 6, p. 374‑386, nov. 2011, doi: 10.1093/bfgp/elr033.
[80] H. Li et R. Durbin, « Fast and accurate short read alignment with Burrows-Wheeler transform », *Bioinforma. Oxf. Engl.*, vol. 25, nº 14, p. 1754‑1760, juill. 2009, doi: 10.1093/bioinformatics/btp324.
[81] H. Li *et al.*, « The Sequence Alignment/Map format and SAMtools », *Bioinformatics*, vol. 25, nº 16, p. 2078‑2079, août 2009, doi: 10.1093/bioinformatics/btp352.
[82] H. V. Firth *et al.*, « DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources », *Am. J. Hum. Genet.*, vol. 84, nº 4, p. 524‑533, avr. 2009, doi: 10.1016/j.ajhg.2009.03.010.
[83] N.-L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, et P. C. Ng, « SIFT web server: predicting effects of amino acid substitutions on proteins », *Nucleic Acids Res.*, vol. 40, nº W1, p. W452‑W457, juill. 2012, doi: 10.1093/nar/gks539.
[84] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, et A. P. Chan, « Predicting the Functional Effect of Amino Acid Substitutions and Indels », *PLoS ONE*, vol. 7, nº 10, p. e46688, oct. 2012, doi: 10.1371/journal.pone.0046688.
[85] E. V. Davydov, D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, et S. Batzoglou, « Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++ », *PLoS Comput. Biol.*, vol. 6, nº 12, p. e1001025, déc. 2010, doi: 10.1371/journal.pcbi.1001025.
[86] H. A. Shihab *et al.*, « Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models », *Hum. Mutat.*, vol. 34, nº 1, p. 57‑65, janv. 2013, doi: 10.1002/humu.22225.
[87] I. Adzhubei, D. M. Jordan, et S. R. Sunyaev, « Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2 », *Curr. Protoc. Hum. Genet.*, vol. 76, nº 1, p. 7.20.1-7.20.41, janv. 2013, doi: 10.1002/0471142905.hg0720s76.
[88] J. M. Schwarz, D. N. Cooper, M. Schuelke, et D. Seelow, « MutationTaster2: mutation prediction for the deep-sequencing age », *Nat. Methods*, vol. 11, nº 4, p. 361‑362, avr. 2014, doi: 10.1038/nmeth.2890.
[89] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, et M. Kircher, « CADD: predicting the deleteriousness of variants throughout the human genome », *Nucleic Acids Res.*, vol. 47, nº D1, p. D886‑D894, janv. 2019, doi: 10.1093/nar/gky1016.
[90]; on behalf of the ACMG Laboratory Quality Assurance Committee *et al.*, « Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology », *Genet. Med.*, vol. 17, nº 5, p. 405‑423, mai 2015, doi: 10.1038/gim.2015.30.
[91] S. Miyatake *et al.*, « Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov Model: an 'exome-first' approach », *J. Hum. Genet.*, vol. 60, nº 4, p. 175‑182, avr. 2015, doi: 10.1038/jhg.2014.124.
[92] N. Krumm *et al.*, « Copy number variation detection and genotyping from exome sequence data », *Genome Res.*, vol. 22, nº 8, p. 1525‑1532, août 2012, doi: 10.1101/gr.138115.112.
[93] P. S. Samarakoon, H. S. Sorte, A. Stray-Pedersen, O. K. Rødningen, T. Rognes, et R. Lyle, « cnvScan: a CNV screening and annotation tool to improve the clinical utility of

computational CNV prediction from exome sequencing data », *BMC Genomics*, vol. 17, nº 1, p. 51, déc. 2016, doi: 10.1186/s12864-016-2374-2.
[94] V. Geoffroy *et al.*, « AnnotSV: an integrated tool for structural variations annotation », *Bioinformatics*, vol. 34, nº 20, p. 3572‑3574, oct. 2018, doi: 10.1093/bioinformatics/bty304.
[95] A. Halman et A. Oshlack, « Accuracy of short tandem repeats genotyping tools in whole exome sequencing data », *F1000Research*, vol. 9, p. 200, mars 2020, doi: 10.12688/f1000research.22639.1.
[96] N. Mousavi, S. Shleizer-Burko, R. Yanicky, et M. Gymrek, « Profiling the genome-wide landscape of tandem repeat expansions », *Nucleic Acids Res.*, vol. 47, nº 15, p. e90, 05 2019, doi: 10.1093/nar/gkz501.
[97] V. C. R. Gentleman, *genefilter*. Bioconductor, 2017.
[98] S. Millecamps *et al.*, « SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations », *J. Med. Genet.*, vol. 47, nº 8, p. 554‑560, août 2010, doi: 10.1136/jmg.2010.077180.
[99] S. Morgan et R. W. Orrell, « Pathogenesis of amyotrophic lateral sclerosis », *Br. Med. Bull.*, vol. 119, nº 1, p. 87‑98, sept. 2016, doi: 10.1093/bmb/ldw026.
[100] M. Aoki *et al.*, « Variance of age at onset in a Japanese family with amyotrophic lateral sclerosis associated with a novel cu/zn superoxide dismutase mutation », *Ann. Neurol.*, vol. 37, nº 5, p. 676‑679, mai 1995, doi: 10.1002/ana.410370518.
[101] A. Chio *et al.*, « Prevalence of SOD1 mutations in the Italian ALS population », *Neurology*, vol. 70, nº 7, p. 533‑537, févr. 2008, doi: 10.1212/01.wnl.0000299187.90432.3f.
[102] O. Abel, J. F. Powell, P. M. Andersen, et A. Al-Chalabi, « ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics », *Hum. Mutat.*, vol. 33, nº 9, p. 1345‑1351, sept. 2012, doi: 10.1002/humu.22157.
[103] D. R. Rosen *et al.*, « Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis », *Nature*, vol. 362, nº 6415, p. 59‑62, mars 1993, doi: 10.1038/362059a0.
[104] R. W. Orrell *et al.*, « Clinical characteristics of SOD1 gene mutations in UK families with ALS », *J. Neurol. Sci.*, vol. 169, nº 1‑2, p. 56‑60, oct. 1999, doi: 10.1016/S0022-510X(99)00216-6.
[105] W. Kim, J.-S. Kim, K.-S. Lee, Y.-J. Gwoun, J.-M. Kim, et K.-H. Lee, « Anticipation and Phenotypic Heterogeneity in Korean Familial Amyotrophic Lateral Sclerosis with Superoxide Dismutase 1 Gene Mutation », *J. Clin. Neurol. Seoul Korea*, vol. 3, nº 1, p. 38‑44, mars 2007, doi: 10.3988/jcn.2007.3.1.38.
[106] H. Zhang *et al.*, « A rare Cu/Zn superoxide dismutase mutation causing familial amyotrophic lateral sclerosis with variable age of onset and incomplete penetrance in China », *Amyotroph. Lateral Scler.*, vol. 6, nº 4, p. 234‑238, janv. 2005, doi: 10.1080/14660820510044478.
[107] S. Vucic, G. A. Nicholson, A. Chio, et M. C. Kiernan, « Apparent anticipation in SOD1 familial amyotrophic lateral sclerosis », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 14, nº 5‑6, p. 452‑456, sept. 2013, doi: 10.3109/21678421.2013.764569.
[108] C. Marelli *et al.*, « Mini-Exome Coupled to Read-Depth Based Copy Number Variation Analysis in Patients with Inherited Ataxias: HUMAN MUTATION », *Hum. Mutat.*, vol. 37, nº 12, p. 1340‑1353, déc. 2016, doi: 10.1002/humu.23063.
[109] L. Montagne *et al.*, « CoDE-seq, an augmented whole-exome sequencing, enables the accurate detection of CNVs and mutations in Mendelian obesity and intellectual disability », *Mol. Metab.*, vol. 13, p. 1‑9, juill. 2018, doi: 10.1016/j.molmet.2018.05.005.
[110] M. Fromer et S. M. Purcell, « Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data », *Curr. Protoc. Hum. Genet.*, vol. 81, nº 1, avr. 2014, doi: 10.1002/0471142905.hg0723s81.
[111] V. Plagnol *et al.*, « A robust model for read count data in exome sequencing experiments and implications for copy number variant calling », *Bioinformatics*, vol. 28, nº 21, p. 2747‑2754, nov. 2012, doi: 10.1093/bioinformatics/bts526.
[112] V. A. McKusick, « Mendelian Inheritance in Man and Its Online Version, OMIM », *Am. J. Hum. Genet.*, vol. 80, nº 4, p. 588‑604, avr. 2007, doi: 10.1086/514346.

[113] C. Munch *et al.*, « Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS », *Neurology*, vol. 63, nº 4, p. 724‑726, août 2004, doi: 10.1212/01.WNL.0000134608.83927.B1.
[114] P. Caroppo *et al.*, « *DCTN1* Mutation Analysis in Families With Progressive Supranuclear Palsy–Like Phenotypes », *JAMA Neurol.*, vol. 71, nº 2, p. 208, févr. 2014, doi: 10.1001/jamaneurol.2013.5100.
[115] D. A. Figlewicz *et al.*, « Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis », *Hum. Mol. Genet.*, vol. 3, nº 10, p. 1757‑1761, 1994, doi: 10.1093/hmg/3.10.1757.
[116] H.-X. Deng *et al.*, « Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia », *Nature*, vol. 477, nº 7363, p. 211‑215, sept. 2011, doi: 10.1038/nature10353.
[117] R. Chia, A. Chiò, et B. J. Traynor, « Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications », *Lancet Neurol.*, vol. 17, nº 1, p. 94‑102, janv. 2018, doi: 10.1016/S1474-4422(17)30401-5.
[118] H. Zhang *et al.*, « Screening for possible oligogenic pathogenesis in Chinese sporadic ALS patients », *Amyotroph. Lateral Scler. Front. Degener.*, vol. 19, nº 5‑6, p. 419‑425, juill. 2018, doi: 10.1080/21678421.2018.1432659.
[119] H. M. Blauw *et al.*, « Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen », *Lancet Neurol.*, vol. 7, nº 4, p. 319‑326, avr. 2008, doi: 10.1016/S1474-4422(08)70048-6.
[120] M. Vidal, M. E. Cusick, et A.-L. Barabási, « Interactome Networks and Human Disease », *Cell*, vol. 144, nº 6, p. 986‑998, mars 2011, doi: 10.1016/j.cell.2011.02.016.
[121] X. Peng, J. Wang, W. Peng, F.-X. Wu, et Y. Pan, « Protein?protein interactions: detection, reliability assessment and applications », *Brief. Bioinform.*, p. bbw066, juill. 2016, doi: 10.1093/bib/bbw066.
[122] J. Zahiri, J. Bozorgmehr, et A. Masoudi-Nejad, « Computational Prediction of Protein–Protein Interaction Networks: Algorithms and Resources », *Curr. Genomics*, vol. 14, nº 6, p. 397‑414, sept. 2013, doi: 10.2174/1389202911314060004.
[123] G. Alanis-Lobato, « Mining protein interactomes to improve their reliability and support the advancement of network medicine », *Front. Genet.*, vol. 6, sept. 2015, doi: 10.3389/fgene.2015.00296.
[124] J.-W. Chang, Y.-Q. Zhou, M. Ul Qamar, L.-L. Chen, et Y.-D. Ding, « Prediction of Protein?Protein Interactions by Evidence Combining Methods », *Int. J. Mol. Sci.*, vol. 17, nº 11, p. 1946, nov. 2016, doi: 10.3390/ijms17111946.
[125] Y. Qi, Z. Bar-Joseph, et J. Klein-Seetharaman, « Evaluation of different biological data and computational classification methods for use in protein interaction prediction », *Proteins Struct. Funct. Bioinforma.*, vol. 63, nº 3, p. 490‑500, janv. 2006, doi: 10.1002/prot.20865.
[126] J. De Las Rivas et C. Fontanillo, « Protein?Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks », *PLoS Comput. Biol.*, vol. 6, nº 6, p. e1000807, juin 2010, doi: 10.1371/journal.pcbi.1000807.
[127] D. Alonso-López *et al.*, « APID database: redefining protein–protein interaction experimental evidences and binary interactomes », *Database*, vol. 2019, janv. 2019, doi: 10.1093/database/baz005.
[128] D. Szklarczyk *et al.*, « STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets », *Nucleic Acids Res.*, vol. 47, nº D1, p. D607‑D613, janv. 2019, doi: 10.1093/nar/gky1131.
[129] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, et Q. Morris, « GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function », *Genome Biol.*, vol. 9, nº Suppl 1, p. S4, 2008, doi: 10.1186/gb-2008-9-s1-s4.
[130] P. Shannon, « Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks », *Genome Res.*, vol. 13, nº 11, p. 2498‑2504, nov. 2003, doi: 10.1101/gr.1239303.
[131] P. Smialowski *et al.*, « The Negatome database: a reference set of non-interacting protein pairs », *Nucleic Acids Res.*, vol. 38, nº Database issue, p. D540‑D544, janv. 2010, doi: 10.1093/nar/gkp1026.

[132] M. A. Mahdavi et Y.-H. Lin, « False positive reduction in protein-protein interaction predictions using gene ontology annotations », *BMC Bioinformatics*, vol. 8, p. 262, juill. 2007, doi: 10.1186/1471-2105-8-262.
[133] S. H. Jung, W.-H. Jang, H.-Y. Hur, B. Hyun, et D.-S. Han, « PROTEIN COMPLEX PREDICTION BASED ON MUTUALLY EXCLUSIVE INTERACTIONS IN PROTEIN INTERACTION NETWORK », nov. 2008, p. 77‑88, doi: 10.1142/9781848163324_0007.
[134] B. Aranda *et al.*, « PSICQUIC and PSISCORE: accessing and scoring molecular interactions », *Nat. Methods*, vol. 8, nº 7, p. 528‑529, juill. 2011, doi: 10.1038/nmeth.1637.
[135] G. Su, J. H. Morris, B. Demchak, et G. D. Bader, « BIOLOGICAL NETWORK EXPLORATION WITH CYTOSCAPE 3 », *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al*, vol. 47, p. 8.13.1-8.13.24, 2014, doi: 10.1002/0471250953.bi0813s47.
[136] Gene Ontology Consortium, « The Gene Ontology (GO) database and informatics resource », *Nucleic Acids Res.*, vol. 32, nº 90001, p. 258D ‑ 261, janv. 2004, doi: 10.1093/nar/gkh036.
[137] G. D. Bader et C. W. Hogue, « [No title found] », *BMC Bioinformatics*, vol. 4, nº 1, p. 2, 2003, doi: 10.1186/1471-2105-4-2.
[138] S. Montaut *et al.*, « Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders », *JAMA Neurol.*, vol. 75, nº 10, p. 1234, oct. 2018, doi: 10.1001/jamaneurol.2018.1478.
[139] M. Sekine et T. Makino, « Inference of Causative Genes for Alzheimer's Disease Due to Dosage Imbalance », *Mol. Biol. Evol.*, vol. 34, nº 9, p. 2396‑2407, sept. 2017, doi: 10.1093/molbev/msx183.
[140] V. La Cognata *et al.*, « A customized high-resolution array-comparative genomic hybridization to explore copy number variations in Parkinson's disease », *neurogenetics*, vol. 17, nº 4, p. 233‑244, oct. 2016, doi: 10.1007/s10048-016-0494-0.
[141] A. Vittori *et al.*, « Copy-number variation of the neuronal glucose transporter gene SLC2A3 and age of onset in Huntington's disease », *Hum. Mol. Genet.*, vol. 23, nº 12, p. 3129‑3137, juin 2014, doi: 10.1093/hmg/ddu022.
[142] C. Kun-Rodrigues *et al.*, « A comprehensive screening of copy number variability in dementia with Lewy bodies », *Neurobiol. Aging*, vol. 75, p. 223.e1-223.e10, mars 2019, doi: 10.1016/j.neurobiolaging.2018.10.019.
[143] A. C.-S. Yu *et al.*, « A Targeted Gene Panel That Covers Coding, Non-coding and Short Tandem Repeat Regions Improves the Diagnosis of Patients With Neurodegenerative Diseases », *Front. Neurosci.*, vol. 13, p. 1324, déc. 2019, doi: 10.3389/fnins.2019.01324.
[144] A. C. Ceylan, H. B. Erdem, İ. Şahin, et M. Agarwal, « SMN1 gene copy number analysis for spinal muscular atrophy (SMA) in a Turkish cohort by CODE-SEQ technology, an integrated solution for detection of SMN1 and SMN2 copy numbers and the "2+0" genotype », *Neurol. Sci.*, avr. 2020, doi: 10.1007/s10072-020-04365-x.
[145] M. F. Ullah, A. Ahmad, S. H. Bhat, F. M. Abu-Duhier, G. E. Barreto, et G. M. Ashraf, « Impact of sex differences and gender specificity on behavioral characteristics and pathophysiology of neurodegenerative disorders », *Neurosci. Biobehav. Rev.*, vol. 102, p. 95‑105, juill. 2019, doi: 10.1016/j.neubiorev.2019.04.003.
[146] R. H. Walker *et al.*, « Huntington's disease-like disorders in Latin America and the Caribbean », *Parkinsonism Relat. Disord.*, vol. 53, p. 10‑20, août 2018, doi: 10.1016/j.parkreldis.2018.05.021.
[147] M. H. Beaubrun, « Huntington's chorea in Trinidad », *West Indian Med. J.*, vol. 12, p. 39‑46, mars 1963.
[148] R. M. Castilhos *et al.*, « Huntington disease and Huntington disease-like in a case series from Brazil: Huntington disease-like phenotype in a case series », *Clin. Genet.*, vol. 86, nº 4, p. 373‑377, oct. 2014, doi: 10.1111/cge.12283.
[149] M. Leńska-Mieciek *et al.*, « Familial ataxia, tremor, and dementia in a polish family with a novel mutation in the *CCDC88C* gene », *Mov. Disord.*, vol. 34, nº 1, p. 142‑144, janv. 2019, doi: 10.1002/mds.27536.
[150] A. Rovelet-Lecrux *et al.*, « Partial deletion of the *MAPT* gene: A novel mechanism of FTDP-17 », *Hum. Mutat.*, vol. 30, nº 4, p. E591‑E602, avr. 2009, doi: 10.1002/humu.20979.

[151] G. R. Wilson *et al.*, « Mutations in RAB39B Cause X-Linked Intellectual Disability and Early-Onset Parkinson Disease with α-Synuclein Pathology », *Am. J. Hum. Genet.*, vol. 95, nº 6, p. 729‑735, déc. 2014, doi: 10.1016/j.ajhg.2014.10.015.

[152] M. K. Kotni, M. Zhao, et D.-Q. Wei, « Gene expression profiles and protein-protein interaction networks in amyotrophic lateral sclerosis patients with C9orf72 mutation », *Orphanet J. Rare Dis.*, vol. 11, nov. 2016, doi: 10.1186/s13023-016-0531-y.

[153] M. H. Schaefer, E. E. Wanker, et M. A. Andrade-Navarro, « Evolution and function of CAG/polyglutamine repeats in protein–protein interaction networks », *Nucleic Acids Res.*, vol. 40, nº 10, p. 4273‑4287, mai 2012, doi: 10.1093/nar/gks011.

[154] Anna-Gaelle Giguet-Valard1,2*, Kevin Raguette1, Stephanie Morin1, Remi et Bellance2 and Juliette Smith Ravin, « Gossypetin Derivatives are also Putative Inhibitors of SARS-COV 2: Results of a Computational Study », *J. Biomed. Res. Environ. Sci.*, vol. 1, nº 6, p. 201‑212, oct. 2020, doi: 10.37871/jbres1144.

ANNEXES

Annexes du

CHAPITRE II

Annexe 1.

Consentement mis en œuvre dans le cadre du diagnostic et de la recherche médicale sur les MNDREM au centre

CONSENTEMENT POUR UNE ETUDE DE GENETIQUE MOLECULAIRE

Décret n°2008-321 du 04 avril 2008 relatif à l'examen des caractéristiques génétiques d'une personne ou à son identification par empreintes génétiques à des fins médicales.
Un exemplaire doit être remis à la personne qui a accepté le prélèvement.

Renseignements pour une étude chez une personne majeure

Je soussigné (e) ....., né(e) le ....., reconnais avoir été informé par le Docteur .....sur les examens des caractéristiques génétiques qui seront réalisés à partir d'un (de) prélèvement(s) pour aider au diagnostic de la maladie dont je souffre ou que présente un membre de ma famille.

Renseignements pour une étude chez une personne mineure ou majeure sous tutelle

Nous soussignés ..... demeurant à ..... père et/ou mère et/ou tuteur légal de ..... , né(e) le ....., demandons au Docteur ..... que soit effectué chez notre enfant une recherche par génétique moléculaire. Celle-ci peut aider au diagnostic de la maladie dont il souffre ou que présente(nt) un (ou plusieurs) membre(s) de ma famille.

Je donne mon consentement pour ce prélèvement et reconnais avoir reçu l'ensemble des informations permettant la compréhension de cet acte biologique et sa finalité conformément aux articles R.1131-4 et R.1131-5 du décret n°2008-312 du 04 avril 2008 du Code de la Santé Publique. Je reconnais avoir été informé (e) des résultats susceptibles d'être obtenus et de leurs multiples conséquences, que certaines analyses génétiques sont susceptibles de mettre en évidence une anomalie génétique de découverte fortuite, que ces examens peuvent s'étendre sur plusieurs années et que ce prélèvement sera conservé le temps nécessaire.

Je reconnais avoir été informé (e) qu'il est indispensable que j'avertisse les membres de ma famille potentiellement concernés par l'anomalie génétique, et que mon silence leur ferait courir des risques ainsi qu'à leur descendance, dès lors que des mesures de prévention ou de soins peuvent être proposés à ceux-ci.

J'autorise :
- L'éventuelle utilisation des échantillons d'ADN afin de progresser dans la connaissance de la maladie concernée, l'envoie de mon ADN à un autre laboratoire pour complément d'étude, ainsi que la transmission des données dans le respect du secret professionnel.
- La conservation et l'utilisation de mes échantillons biologiques à des fins de recherche médicale, dans le cadre de la maladie dont je souffre ou que présente un membre de ma famille, sans limite de temps. Dans le cadre particulier d'analyse multi loci du génome ou d'analyse globale du génome (séquençage haut débit +/- capture de séquence, ...), l'analyse peut révéler d'autres affections que celles recherchées. Le patient, le titulaire légal, le tuteur ou le curateur pourra en être informé seulement en cas de bénéfice direct pour le patient ou sa famille en l'état actuel des connaissances, c'est-à-dire en cas de possibilité de prévention et/ou de traitement.
- Le recueil, la saisie, le traitement et la consultation des données contenues dans mon dossier au seul personnel médical, en toute confidentialité.
- La transmission de mon dossier médical auprès des médecins concernés par l'examen de mes caractéristiques génétiques et l'échange d'informations nécessaires à mon sujet.

Fait à ..... Le .....

Signature du patient / parents de l'enfant / ou tuteur légal

Signature du patient mineur ou majeur sous tutelle

ATTESTATION DE CONSULTATION DU MEDECIN PRESCRIPTEUR OU DU CONSEILLER EN GENETIQUE

Je soussigné (e) Docteur / Conseiller en génétique .....
Conformément aux articles R.1131-4 et R.1131-5 du décret n°2008-312 du 04 avril 2008, certifie avoir reçu ce jour en consultation le (la) patient(e) sous-nommé(e) afin de l'informer des caractéristiques de la maladie recherchée, des moyens de la détecter, du degré de fiabilité des analyses ainsi que des possibilités de prévention et de traitement. En outre, il (elle) est informé (e) des modalités de transmission génétique de la maladie recherchée et de leurs possibles conséquences chez d'autres membres de sa famille.

Fait à ..... Le .....

Signature et nom du médecin ou du conseiller en génétique :

Annexe 2.

Avis favorable n° N°2020/076 Institutional Review Board (IRB)
Centre Hospitalier Universitaire de Martinique (CHUM)

CHU de Martinique

CENTRE HOSPITALIER UNIVERSITAIRE DE MARTINIQUE

Institutional Review Board (IRB)

Docteur Rémi Bellance

Concerne : Avis de l'IRB du Centre Hospitalier et Universitaire Martinique (20 octobre 2020)

Monsieur et Cher confrère,

Vous avez sollicité l'avis de l'IRB du centre hospitalier et universitaire de Martinique concernant votre projet intitulé « Etude clinique et moléculaire des maladies neurodégénératives rares à expression motrice sans étiologie génétique ».

L'IRB émet un avis favorable. La référence de cet avis est 2020/076.
Nous vous remercions de nous avoir sollicités et vous souhaitons du succès dans les suites de votre projet.

Avec nos meilleures salutations.

Professeur Moustapha DRAME
Responsable de l'unité de soutien méthodologique à la recherche (USMR)
CHU Martinique Hôpital Pierre Zobda-Quitman CS 90632
97261 Fort-de-France
Secrétariat : +596 596 59 28 96
Poste interne : 9769
Courriel : moustapha.drame@chu-martinique.fr

Professeur Moustapha Dramé
Président de l'IRB

Annexe 3.

Fiche information relative à l'étude des MNDREM

NOTE D'INFORMATION

CHU de Martinique

BIOSPHERES Université des Antilles

CeRCA - CEntre de Reference CAribeen
DES MALADIES NEUROMUSCULAIRES ET NEUROLOGIQUES RARES

SITE CONSTITUTIF CENTRE AOC
CENTRE DE COMPETENCE DES ANOMALIES EMBRYONNAIRES DU DEVELOPPEMENT
CENTRE DE COMPETENCE DE LA MALADIE DE HUNTINGTON

maladies rares Filnemus

HOPITAL PIERRE ZOBDA-QUITMAN - NIVEAU –1
BP 632-97200 FORT DE France
N° FINESS 970 211 207

☎ : 0596 55 22 64 - Fax : 0596 75 52 66
cerca.secretariat@chu-martinique.fr

Madame, monsieur,

« La caractérisation clinique et moléculaire des Maladies Neurodégénératives Rares à Expression Motrices (MNDREM) sans étiologie génétique »,

Est une étude menée au CERCA, sous la direction du Docteur Rémi BELLANCE, coordonnateur et neurologue du Centre (05 96 55 22 64) ; En partenariat avec l'équipe de recherche BIOSPHERES/AREBIO de l'Université des Antilles basé au campus de Schœlcher en Martinique, dirigé par le Pr Juliette Smith-Ravin ; Et sous la supervision du chef de service de neurogénétique au CHU de Bordeaux, le Pr Cyril GOIZET.
Le délégué à la protection des données est Mr Gérald GALIM (service informatique CHUM).

L'objectif de cette étude est de mettre en évidence les spécificités cliniques et génétiques des patients de notre bassin de population. A ce titre, nous vous informons de l'utilisation des données de votre dossier médical conformément à l'article 6 du RGPD. Les informations recueillies ont caractère essentiellement identitaire, médical et familial. Ces données seront exploitées au sein même du service. Elles seront conservées au moins 2ans après la dernière publication des résultats.

Dans le respect de votre consentement donné pour l'analyse de vos caractéristiques génétiques à des fins de recherche médicales, votre échantillon d'ADN conservé au CERCA pourra faire l'objet d'analyses complémentaires.

Votre participation à cette étude n'est pas obligatoire. Vous pouvez, à tout moment accéder aux données médicales collectées, rectifier les données dont nous disposons ou faire opposition à cette participation par simple courrier à l'attention du responsable, le Dr BELLANCE.

Vous pourriez être sollicités par téléphone pour vérifier certaines données médicales de l'histoire de la maladie et de la famille.

L'absence de votre manifestation, vaut pour accord à la participation à cette étude.

Annexe 4.

Déclaration Comité National Informatique et Libertés (CNIL),
conforme au MR-001 pour l'analyse génomique des patients de la cohorte

CNIL.
COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

Référence CNIL :
2219621 v 0

Déclaration de conformité

au référentiel de méthodologie de référence MR-001

reçue le 12 octobre 2020

Madame Anna-Gaelle GIGUET-VALARD

2, AV EMILE MAURICE
97233 SCHOELCHER

ORGANISME DÉCLARANT

Nom : Madame GIGUET-VALARD Anna-Gaelle
Service :
Adresse : 2, AV EMILE MAURICE
CP : 97233
Ville : SCHOELCHER

N° SIREN/SIRET :
Code NAF ou APE :
Tél. : +596696224488
Fax. :

Par la présente déclaration, le déclarant atteste de la conformité de son/ses traitement(s) de données à caractère personnel au référentiel mentionné ci-dessus.

La CNIL peut à tout moment vérifier, par courrier ou par la voie d'un contrôle sur place ou en ligne, la conformité de ce(s) traitement(s).

Fait à Paris, le 12 octobre 2020

Les données personnelles nécessaires à l'accomplissement des missions de la CNIL sont conservées et traitées dans des fichiers destinés à son usage exclusif. Les personnes concernées peuvent exercer leurs droits Informatique et Libertés en s'adressant au délégué à la protection des données de la CNIL via un formulaire en ligne ou par courrier postal.

Pour en savoir plus : https://www.cnil.fr/donnees-personnelles

Annexe 5.

Commandes pipelines informatique mis en œuvre

Commande 1

```
#!/bin/bash
#Initialisation des variables
name_sample=$1 # sample name
gen_ref=$2 # fasta reference genome file
target_bed=$3 # fichier annotation exome hg38
CPU=$4 #nombre de thread ( Core logique processeur)
$dbsnp=$5
MillsG_indel=$6
date_get_reads=$7 #date obtention des reads pour le traçage (ex: D2018-12-20) output_dir=$8
BBMAP="/path/to_bbmap"
gatk="/path/to_gatk"
picard="/path/to_picard"
tmp_file="/path/tmp_dir" #Dossier temporaire utilisé par l'outil gatk freebayes="/path/to_freebayes"
snpEff="/path/to_snpEff_tools"
DB="path/to_all_Database_for_snpEff"
vt="/path/to_vt_tools"
tabix="/path/to_htslib/tabix"
bgzip="/path/to_htslib/bgzip"
vcftools="/path/to_vcftools"
scripts="/path/to_scripts_maison"
#On se place dans le dossier où seront stocker les sorties
cd $output_dir
```

Commande 2

```
fastqc fichier.fastq -o repertoire.sortie -j java
```

Commande 3

```
java -jar trimmomatic-0.35.jar PE -phred33 {name_sample}_R1.fastq.gz {name_sample}_R2.fastq.gz
{name_sample}_clean_R1.fastq.gz {name_sample}_clean_R2.fastq.gz \
output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz \
ILLUMINACLIP:TruSeq3 -PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

Commande 4

```
bwa mem -t $CPU -M -R "@RG\tID:${name_sample}_D${date_get_reads}\tLB:MedExome\
tPL:ILLUMINA\tSM:${name_sample}\tPU:NextSeq2" \
$gen_ref \
${name_sample}_clean_R1.fastq.gz \
${name_sample}_clean_R2.fastq.gz \
| samtools view -bS -@ $CPU - \
| samtools sort -o ${name_sample}.sort.bam
```

Commande 5

```
#Convertir en .bam
samtools view –b –h –o ${name_sample}.sort.bam ${name_sample}.sort.sam
#Obtenir les fichiers de mesures
samtools flagstat ${name_sample}.sort.bam
#Indexation du fichier et renommer le fichier
samtools index ${name_sample}.sort.bam
mv ${name_sample}.marked.bai ${name_sample}.marked.bam.bai
#Marquer les Duplicats
java -Djava.io.tmpdir=$tmp_file \
-jar $picard \
MarkDuplicates I=${name_sample}.sort.bam \ O=${name_sample}.marked.bam \
METRICS_FILE=${name_sample}.marked.sort.dedupMetrics.txt \ ASSUME_SORTED=true
CREATE_INDEX=true
```

```
#BaseRecalibrator, 1er tour
java -Djava.io.tmpdir=$tmp_file -jar $gatk \
-T BaseRecalibrator \
-R $gen_ref \
-L $target_bed \
-I ${name_sample}.marked.bam \
-o ${name_sample}.recal.table \
-knownSites $dbsnp \
-knownSites $MillsG_indel \ 2>${name_sample}.recal.BaseRecalibrator.err
#BaseRecalibrator, 2ème tour
java -Djava.io.tmpdir=$tmp_file \
-jar $gatk \
-T BaseRecalibrator \
-R $gen_ref \
-L $target_bed \
-I ${name_sample}.marked.bam \
-BQSR ${name_sample}.recal.table \
-o post_${name_sample}.recal.table \
-knownSites $dbsnp \
-knownSites $MillsG_indel \ 2>post_${name_sample}.recal.BaseRecalibrator.err
# Appliquer la “Recalibration” (PrintReads)
java -Djava.io.tmpdir=$tmp_file -jar $gatk -T PrintReads -R $gen_ref -I $ {name_sample}.marked.bam -L
$target_bed -BQSR ${name_sample}.recal.table -o $ {name_sample}.recal.bam
```

Commande 6

```
HaplotypeCaller
java -Djava.io.tmpdir=$tmp_file \ -jar $gatk \
-T HaplotypeCaller \
-R $gen_ref \
-I $name_sample.recal.bam \
-L $target_bed \
-ip 150 \
-drf DuplicateRead \
--variant_index_type LINEAR \ --variant_index_parameter 128000 \
-dt none \
-mbq 10 -mmq 20 -ERC GVCF \
-o ${output_dir}/${name_sample}_raw.variants.vcf
#Création d’un fichier liste par fusion des raw.variants.vcf obtenus pour chaque échantillon
echo "${output_dir}/${name_sample}_raw.variants.vcf" >> "${output_dir}/all_raw_variants.list"
#CombineVCF
java -Djava.io.tmpdir=$tmp_file \
-jar $gatk \
-T CombineGVCFs \
-R $gen_ref \
-V ${output_dir}/all_raw_variants.list \ --out ${output_dir}/all_raw_variants.vcf
#Genotype CombineVCF
java -Djava.io.tmpdir=$tmp_file \
-jar $gatk \
-T GenotypeGVCFs \
-R $gen_ref \
-ip 150 \
-V ${output_dir}/all_raw_variants.vcf \ --out ${output_dir}/raw.variants.vcf \ -dt NONE -stand_call_conf 0
#GATK SelectVariants
java -Djava.io.tmpdir=$tmp_file \
-jar $gatk \
-T SelectVariants \
-R $gen_ref \
-dt NONE \
```

```
-L $target_bed \
-V ${output_dir}/raw.variants.vcf \
-o ${output_dir}/${name_sample}.gatk.vcf \
-sn $name_sample \
--excludeNonVariants --keepOriginalDP --keepOriginalAC
#Filtration des vcf pour ne garder que les sites non multi-alléliques avec script python maison
${scripts}/oneAllelePerLine.py ${output_dir}/${name_sample}.gatk.vcf > sample_allele_filtrer_gatk.vcf
#Recherche toutes les variantes qui se produisent dans le premier échantillon avec le module SnpSift de l'outil snpEff
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "isVariant( GEN[0] )" sample_allele_filtrer_gatk.vcf >
${output_dir}/${name_sample}.gatk.monoAll.vcf
#Normalisation
$vt normalize ${output_dir}/${name_sample}.gatk.monoAll.vcf -r $gen_ref -o ${output_dir}/$
{name_sample}.gatk.norm.vcf
#Vartype
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar varType -v ${output_dir}/$ {name_sample}.gatk.norm.vcf >
${output_dir}/${name_sample}.gatk.VT.vcf
#Filter 1
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "( ( na QD ) | ( QD[*] >= 2.0 ) ) & ( ( na QUAL ) | (
QUAL[*] >= 20 ) ) & ( ( (exists DEL) & ( ( ( na InbreedingCoeff ) |
( InbreedingCoeff[*] >= -0.8 ) ) ) ) | ( (exists MNP) & ( ( ( na MQ ) | ( MQ[*] >= 40.0 ) ) ) ) |
( (exists INS) & ( ( ( na InbreedingCoeff ) | ( InbreedingCoeff[*] >= -0.8 ) ) ) ) | ( (exists SNP) & ( ( ( na MQ ) | (
MQ[*] >= 40.0 ) ) ) ) | (exists MIXED) )" -f ${output_dir}/$ {name_sample}.gatk.VT.vcf >
${output_dir}/${name_sample}.gatk.FQ2.vcf
#Filter 2
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "( ( na GEN[0].DP ) | ( GEN[0].DP >= 5 ) )" -f
${output_dir}/${name_sample}.gatk.FQ2.vcf > ${output_dir}/$ {name_sample}.gatk.GQ.vcf
# Ajout du mot « gatk» à la fin du fichier vcf obtenu pour le différencier de celui obtenu avec FREBAYES grâce à un
script python maison
python ${scripts}/addToolInfo.py ${output_dir}/${name_sample}.gatk.GQ.vcf gatk
```

Commande 7

```
BAMLIST="/path/to_recal.bam.list/" #Fichier au format txt avec le chemin ou se trouve chaque alignement
#run freebayes
$freebayes -L $BAMLIST -f $gen_ref --target $target_bed -q 10 -m 20 --use-duplicate-reads --vcf $
{output_dir}/all_vcf_freeb.vcf
#Tri du fichier vcf obtenu précédemment
grep '^#' ${output_dir}/all_vcf_freeb.vcf > ${output_dir}/all_vcf_freeb.sort.vcf && grep -v '^#' $
{output_dir}/all_vcf_freeb.vcf | LC_ALL=C sort -t $'\t' -V -k1,1 -k2,2n >> ${output_dir}/all_vcf_freeb.sort.vcf
#Si plusieurs échantillons ont été traités par Freebayes (dans notre cas on a 15 échantillons), on les sépare avec GATK
SelectVariants de la façon suivante :
java -Djava.io.tmpdir=$tmp_file -jar $gatk \
-T SelectVariants \
-R $gen_ref \
-L $target_bed \
-V all_vcf_freeb.sort.vcf \
-o ${output_dir}/${name_sample}.vcf \
-sn ${name_sample} \
---excludeNonVariants --keepOriginalDP --keepOriginalAC
#Filtration vcf pour ne garder que les sites qui ne sont pas multi-alléliques avec script maison python
${scripts}/oneAllelePerLine.py ${output_dir}/${name_sample}.freeb.vcf > sample_allele_filtrer_freeb.vcf
#Recherche de tous les variants qui se produisent dans le premier échantillon avec le module SnpSift de l'outil snpEff
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "isVariant( GEN[0] )" sample_allele_filtrer_freeb.vcf >
${output_dir}/${name_sample}.freeb.monoAll.vcf
#Normalisation
$vt normalize ${output_dir}/${name_sample}.freeb.monoAll.vcf -r $gen_ref -o ${output_dir}/$
{name_sample}.freeb.norm.vcf
#Vartype
```

```
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar varType -v ${output_dir}/$ {name_sample}.freeb.norm.vcf > ${output_dir}/${name_sample}.freeb.VT.vcf
#Filter 1
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "( ( na QUAL ) | ( QUAL[*] >= 20 ) ) & ( ( na SAR ) | ( SAR[*] >= 3 ) ) & ( ( na SAF ) | ( SAF[*] >= 3 ) )" -f ${output_dir}/$ {name_sample}.freeb.VT.vcf > ${output_dir}/${name_sample}.freeb.FQ2.vcf
#Filter 2
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar filter "( ( na GEN[0].DP ) | ( GEN[0].DP >= 5 ) )" -f ${output_dir}/${name_sample}.freeb.FQ2.vcf > ${output_dir}/$ {name_sample}.freeb.GQ.vcf
# Ajout du mot « freeb » à la fin du fichier vcf obtenu pour le différencier de celui obtenu avec GATK grâce à un script python maison
python ${scripts}/addToolInfo.py ${output_dir}/${name_sample}.freeb.GQ.vcf freeb
```

Commande 8

```
#Compression fichiers freeb pour utilisation vcftools
$bgzip -c ${output_dir}/${name_sample}.freeb.toolInfo.vcf > ${output_dir}/$ {name_sample}.freeb.toolInfo.vcf.gz
$tabix -p vcf ${output_dir}/${name_sample}.freeb.toolInfo.vcf.gz
#Compression fichiers gatk pour utilisation vcftools
$bgzip -c ${output_dir}/${name_sample}.gatk.toolInfo.vcf > ${output_dir}/$ {name_sample}.gatk.toolInfo.vcf.gz
$tabix -p vcf ${output_dir}/${name_sample}.gatk.toolInfo.vcf.gz
#Merge avec vcftool
$vcftools/vcf-merge ${output_dir}/${name_sample}.gatk.toolInfo.vcf.gz ${output_dir}/$ {name_sample}.freeb.toolInfo.vcf.gz > ${output_dir}/${name_sample}.mergeTool.vcf
```

Commande 9

```
#Extraction valeur AD, GT et DP avec script python maison
python get_genotype.py ${output_dir}/${name_sample}.mergeTool.vcf
#Annotation DB snpeff
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/snpEff.jar -noStats -no-downstream -no-upstream -no- intergenic -noMotif -noNextProt -noInteraction -v GRCh38.92 ${output_dir}/$ {name_sample}.mergeTool.vcf > ${output_dir}/snpeff_${name_sample}.mergeTool.ann.vcf
#Annotation DB Clinvar
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name ClinVar_ $DB/clinvar_20191013.vcf.gz ${output_dir}/snpeff_${name_sample}.mergeTool.ann.vcf > $ {output_dir}/clinvar_${name_sample}.mergeTool.ann.vcf
#Annotation DB dbsnp
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name dbSnp_ $DB/dbsnp_b150_GRCh38p7.vcf.gz ${output_dir}/clinvar_${name_sample}.mergeTool.ann.vcf > ${output_dir}/dbsnp_${name_sample}.mergeTool.ann.vcf
#Annotation DB ESP6500
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name ESP6500_ $DB/fixed_ESP6500SI-V2-SSA137.GRCh38-liftover.vcf.gz ${output_dir}/dbsnp_$ {name_sample}.mergeTool.ann.vcf > ${output_dir}/ESP6500_${name_sample}.mergeTool.ann.vcf
#Annotation DB gnomAD genome
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name GnomAD_ge_ $DB/gnomad.genomes.r2.1.1.sites.vcf.gz ${output_dir}/ESP6500_$ {name_sample}.mergeTool.ann.vcf > ${output_dir}/gnomAD_ge_$ {name_sample}.mergeTool.ann.vcf
#Annotation DB gnomAD exome
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name GnomAD_ex_ $DB/gnomad.exomes.r2.1.1.sites.vcf.gz ${output_dir}/gnomAD_ge_$ {name_sample}.mergeTool.ann.vcf > ${output_dir}/gnomAD_ex_$ {name_sample}.mergeTool.ann.vcf
#Annotation DB dbnsfp
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar dbnsfp -g GRCh38.92 -f SIFT4G_score,SIFT4G_pred,Polyphen2_HDIV_score,Polyphen2_HDIV_pred,Polyphen2_HVAR_score,Polyphen2_HVAR_pred,LRT_score,LRT_pred,MutationTaster_score,MutationTaster_pred,\
MutationAssessor_score,MutationAssessor_pred,FATHMM_score,FATHMM_pred,PROVEAN_score,PROVEAN_pred,VEST4_score,MetaSVM_score,MetaSVM_pred,MetaLR_score,MetaLR_pre d,M-CAP_score,\
M-CAP_pred,REVEL_score,MutPred_score,CADD_phred,DANN_score,fathmm- MKL_coding_score,fathmm-
```

```
MKL_coding_pred,Eigen-pred_coding,Eigen-PC-
phred_coding,GERP++_NR,GERP++_RS,phyloP100way_vertebrate,\
phyloP30way_mammalian,phastCons100way_vertebrate,phastCons30way_mammalian,SiPhy_29w
ay_logOdds,GTEx_V7_gene,GTEx_V7_tissue \
-db $DB/dbNSFP4.0a.txt.gz ${output_dir}/gnomAD_ex_${name_sample}.mergeTool.ann.vcf > $
{output_dir}/dbnsfp_${name_sample}.mergeTool.ann.vcf
#Annotation DB 1000G Phase 3
java -Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar annotate -name KGp3_
$DB/fixed_1KG.phase3.GRCh38.vcf.gz ${output_dir}/dbnsfp_$ {name_sample}.mergeTool.ann.vcf >
${output_dir}/${name_sample}.mergeTool.ann.vcf
#Filtration maf-eff à partir valeurs dbsnp_CAF ESP6500_ALL_AF GnomAD_ge_AF GnomAD_ex_AF KGp3_MAF
et du type de variation (ex : )
python ${scripts}/filtered_with_maf_eff.py ${output_dir}/${name_sample}.mergeTool.ann.vcf $DB/effect.list
#Conversion fichier vcf en fichier tsv
cat ${output_dir}/${name_sample}.mergeTool.FE.vcf | $snpEff/scripts/vcfEffOnePerLine.pl | java -
Djava.io.tmpdir=$tmp_file -jar $snpEff/SnpSift.jar extractFields - CHROM POS ID REF ALT "ANN[*].EFFECT"
"ANN[*].IMPACT" "ANN[*].GENE" "ANN[*].GENEID" "ANN[*].FEATURE" "ANN[*].FEATUREID"
"ANN[*].BIOTYPE" "ANN[*].RANK" "ANN[*].HGVS_C" "ANN[*].HGVS_P" "ANN[*].CDNA_POS"
"ANN[*].CDNA_LEN" "ANN[*].CDS_POS" "ANN[*].CDS_LEN" "ANN[*].AA_POS" "ANN[*].AA_LEN"
"ANN[*].DISTANCE" GnomAD_ex_AF GnomAD_ex_AC GnomAD_ex_AN GnomAD_ex_AF_popmax
GnomAD_ge_AF GnomAD_ge_AC GnomAD_ge_AN GnomAD_ge_AF_popmax ClinVar_CLNHGVS
ClinVar_CLNSIG ESP6500_MAF ESP6500_GS KGp3_AF KGp3_AC KGp3_EUR_AF KGp3_AMR_AF
KGp3_EAS_AF KGp3_SAS_AF KGp3_AFR_AF dbSnp_dbSNPBuildID dbNSFP_SIFT4G_score
dbNSFP_SIFT4G_pred dbNSFP_Polyphen2_HDIV_score dbNSFP_Polyphen2_HDIV_pred
dbNSFP_Polyphen2_HVAR_score dbNSFP_Polyphen2_HVAR_pred dbNSFP_LRT_score dbNSFP_LRT_pred
dbNSFP_MutationTaster_score dbNSFP_MutationTaster_pred dbNSFP_MutationAssessor_score
dbNSFP_MutationAssessor_pred dbNSFP_FATHMM_score dbNSFP_FATHMM_pred dbNSFP_PROVEAN_score
dbNSFP_PROVEAN_pred dbNSFP_VEST4_score dbNSFP_MetaSVM_score dbNSFP_MetaSVM_pred
dbNSFP_MetaLR_score dbNSFP_MetaLR_pred dbNSFP_M_CAP_score dbNSFP_M_CAP_pred
dbNSFP_REVEL_score dbNSFP_MutPred_score dbNSFP_CADD_phred dbNSFP_DANN_score
dbNSFP_fathmm_MKL_coding_score dbNSFP_fathmm_MKL_coding_pred dbNSFP_Eigen_pred_coding
dbNSFP_Eigen_PC_phred_coding dbNSFP_GERP___NR dbNSFP_GERP___RS
dbNSFP_phyloP100way_vertebrate dbNSFP_phastCons30way_mammalian dbNSFP_SiPhy_29way_logOdds
dbNSFP_GTEx_V7_gene dbNSFP_GTEx_V7_tissue "LOF[*].GENE" "LOF[*].GENEID" "NMD[*].GENE"
"NMD[*].GENEID" > ${output_dir}/tmp1_${name_sample}.mergeTool.ann.txt
#Annotation score pLi avec DB Exac, phenotypes et mode de transmission avec la DB d'OMIM, et tissus avec DB
Tiger + Ajout colonne génotype
python ${scripts}/printable_annotation.py ${output_dir}/tmp2_${name_sample}.mergeTool.ann.txt
$DB/fordist_cleaned_exac_r03_march16_z_pli_rec_null_data.txt $DB/genemap2.txt $DB/tiger_tissue.txt
${output_dir}/genotype_${name_sample}.tsv > annotation_$ {name_sample}.mergeTool.ann.txt
```

Commande 10

```
#Annotation avec liste gènes impliqués dans ND et SLA
python ${scripts}/annotation_with_disease.py ${output_dir}/tmp1_$ {name_sample}.mergeTool.ann.txt
$DB/gene_result_ALS_01072019.txt $DB/gene_result_ND_01072019.txt >
tmp2_${name_sample}.mergeTool.ann.txt
```

Commande 11

```
#!/bin/bash
#Initialisation des variables
seqdb="/path/to_seqdb_hg38"
locdb="/path/to_locdb_hg38"
bam_list="/path/to_recal_bam.list"
gen_ref="/path/to_genome_ref" #resources_broad_hg38_v0_Homo_sapiens_assembly38.fasta
tmp_dir="/path/to_tmp_dir"
gatk="/path/to_gatk"
xhmm="/path/to_xhmm"
plinkseq="/path/to_plinkseq-0.10"
```

```
output="/path/to_output_dir"
scripts="/path/to_homemade_scripts" target_bed="/path/to_MedExome_hg38_capture_targets.bed"
list_target_regions="/path/to_MedExome_hg38_capture_targets_interval.list"
```

Commande 12

```
# GATK « depth of coverage » avec Medexome hg38
java -Xmx7G -Djava.io.tmpdir=$tmp_dir \
-jar $gatk \
-T DepthOfCoverage \
-R $gen_ref \
-L $target_bed \
-I $bam_list \
-dt BY_SAMPLE -dcov 5000 -l INFO --omitDepthOutputAtEachBase --omitLocusTable \ --minBaseQuality 0 --
minMappingQuality 20 --start 1 --stop 5000 --nBins 200 \ --includeRefNSites \
--countType COUNT_FRAGMENTS \
-o $output/group.recal.DATA \
2>$output/call_xhmm.log
#Combinaison résultats GATK « Depth-of-Coverage » pour plusieurs échantillons au même locus
$xhmm/statgen-xhmm-cc14e528d909/xhmm --mergeGATKdepths \
-o $output/DATA.RD.txt \
--GATKdepths $output/group1.recal.DATA.sample_interval_summary \ --GATKdepths
$output/group2.recal.DATA.sample_interval_summary \ 2>>$output/call_xhmm.log
# Calculer le contenu GC and créer une liste de targets avec contenus extrême en GC
java -Djava.io.tmpdir=$tmp_dir \
-jar $gatk \
-T GCContentByInterval \
-L $target_bed \
-R $gen_ref \
-o $output/DATA.locus_GC.txt \ 2>>$output/call_xhmm.log
cat $output/DATA.locus_GC.txt \
| awk '{if ($2 < 0.1 || $2 > 0.9) print $1}' \ > $output/extreme_gc_targets.txt \ 2>>$output/call_xhmm.log
#Utiliser PLINK/Seq pour calculer la fraction de « repeat-masked bases » dans chaque cible et créer une liste de cible
à partir de laquelle on effectuera la filtration
$xhmm/statgen-xhmm-cc14e528d909/sources/scripts/interval_list_to_pseq_reg \ $list_target_regions \
> $output/DATA.EXOME.targets.reg \ 2>>$output/call_xhmm.log
$plinkseq/pseq . loc-load --locdb $output/DATA.EXOME.targets.LOCDB --file $output/DATA.EXOME.targets.reg -
-group targets --out $output/DATA.EXOME.targets.LOCDB.loc-load \
2>>$output/call_xhmm.log
$plinkseq/pseq . loc-stats --locdb $output/DATA.EXOME.targets.LOCDB --group targets --seqdb $seqdb | awk '{if
(NR > 1) print $_}' | sort -k1 -g | awk '{print $10}' \
| paste $list_target_regions - | awk '{print $1"\t"$2}' > $output/DATA.locus_complexity.txt \
2>>$output/call_xhmm.log
cat $output/DATA.locus_complexity.txt \ | awk '{if ($2 > 0.25) print $1}' \
> $output/low_complexity_targets.txt \ 2>>$output/call_xhmm.log
#Filtrer échantillons et cibles, leurs « mean-centers » les cibles
$xhmm/statgen-xhmm-cc14e528d909/xhmm --matrix -r \ $output/DATA.RD.txt \
--centerData --centerType target \
-o $output/DATA.filtered_centered.RD.txt \
--outputExcludedTargets $output/DATA.filtered_centered.RD.txt.filtered_targets.txt \ --outputExcludedSamples
$output/DATA.filtered_centered.RD.txt.filtered_samples.txt \ --minTargetSize 10 --maxTargetSize 10000 \
--minMeanTargetRD 10 --maxMeanTargetRD 500 \
--minMeanSampleRD 25 --maxMeanSampleRD 200 \
--maxSdSampleRD 150 \
2>>$output/call_xhmm.log
#Exclure les cibles
xhmm_output_hg19/all.process.DATA.extreme_gc_targets.txt --excludeTargets
xhmm_output_hg19/all.process.low_complexity_targets.txt
#Utiliser les PCA sur les données « mean-centered »
```

```
$xhmm/statgen-xhmm-cc14e528d909/xhmm --PCA \
-r $output/DATA.filtered_centered.RD.txt \ --PCAfiles $output/DATA.RD_PCA \ 2>>$output/call_xhmm.log
#Normaliser les données « mean-centered » data en utilisant les informations de PCA (PS: changement 0.5 par 0.7)
$xhmm/statgen-xhmm-cc14e528d909/xhmm \
--normalize -r $output/DATA.filtered_centered.RD.txt \
--PCAfiles $output/DATA.RD_PCA \
--normalizeOutput $output/DATA.PCA_normalized.txt \ --PCnormalizeMethod PVE_mean --PVE_mean_factor 0.7 \
2>>$output/call_xhmm.log
#Filtres et z-score centres (par échantillons), PCA-normalisé
$xhmm/statgen-xhmm-cc14e528d909/xhmm \
--matrix -r $output/DATA.PCA_normalized.txt \
--centerData --centerType sample --zScoreData \
-o $output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt \ --outputExcludedTargets
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt.filtered_targets.txt \ --outputExcludedSamples
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt.filtered_samples.txt \ --maxSdTargetRD 30 \
2>>$output/call_xhmm.log
#Filtrer les donnée de « read-depth » pour faire correspondre aux données filtrées et normalisées
$xhmm/statgen-xhmm-cc14e528d909/xhmm \
--matrix -r $output/DATA.RD.txt \
--excludeTargets $output/DATA.filtered_centered.RD.txt.filtered_targets.txt \ --excludeTargets
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt.filtered_targets.txt \ --excludeSamples
$output/DATA.filtered_centered.RD.txt.filtered_samples.txt \ --excludeSamples
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt.filtered_samples.txt \ -o
$output/DATA.same_filtered.RD.txt \
2>>$output/call_xhmm.log
#Détecter les CNVs dans les données normaisées
$xhmm/statgen-xhmm-cc14e528d909/xhmm --discover -p $xhmm/statgen-xhmm-cc14e528d909/params.txt -r
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt -R $output/DATA.same_filtered.RD.txt -c
$output/DATA.xcnv -a $output/DATA.aux_xcnv -s $output/ DATA 2>>$output/call_xhmm.log
#Génotyper les CNVs détectés
$xhmm/statgen-xhmm-cc14e528d909/xhmm --genotype -p $xhmm/statgen-xhmm-cc14e528d909/params.txt -r
$output/DATA.PCA_normalized.filtered.sample_zscores.RD.txt -R $output/DATA.same_filtered.RD.txt -g
$output/DATA.xcnv -F $gen_ref -v $output/DATA.vcf 2>>$output/call_xhmm.log
# Annoter les cibles exome avec les gènes correspondants dans hg38
$plinkseq/pseq . loc-intersect --group refseq --locdb $locdb --file $target_bed --out
$output/DATA.annotated_targets.refseq \
2>>$output/call_xhmm.log
```

Commande 13

```
#source("http://bioconductor.org/biocLite.R")
#biocLite("Rsamtools")
#biocLite("GenomicAlignments")
#install.packages("ExomeDepth")
library(GenomicAlignments)
library(Rsamtools)
library(ExomeDepth)

#Chargement données
bedFile <- file.path("/path/MedExome_design_files/hg38", "exons_MedExome_hg38_capture_targets.bed")
targets <- read.table(bedFile,head=F)
colnames(targets) <- c("chromosome","start","end","name") hg38<-
file.path("/path/tools_bioinfo_NGS/GATK_TOOLS/bwa-
mem_hg38","resources_broad_hg38_v0_Homo_sapiens_assembly38.fasta") bamName <-
list.files("/path/recal.bam_dir", pattern = '*.bam$')
bamFile <- file.path("/path/recal.bam_dir",bamName)
baiName <- list.files("/path/recal.bam_dir", pattern = '*.bam.bai$')
baiFile <- file.path("/path/recal.bam_dir",baiName)
```

```
samples_data <- scan("/path/samples_data", quiet=T,what="character") samples_control <-
scan("/path/samples_control", quiet=T,what="character")
#count bam
TargetCount <- getBamCounts(bed.frame = targets,bam.files = bamFile,index.files = baiFile, include.chr =
FALSE,referenceFasta = hg38)
#renommer les échantillons dont getBamCounts a changé le nom
name_all_samples <- scan("/path/name_file_all_samples", quiet=T,what="character")
for (names_samples in name_all_samples) {
names(TargetCount)[gsub("[X]|\\.recal.bam", "",names(TargetCount)) == names_samples] <-
paste(names_samples,".recal.bam",sep="")
}
#matrice all_data
TargetCount.df <- as(TargetCount[, colnames(TargetCount)], 'data.frame') TargetCount.df$chromosome <-
gsub(as.character(TargetCount.df$chromosome),pattern = 'chr',replacement = '')
#matrice échantillons contrôles
my.ref.samples <- paste0(samples_control,".recal.bam") my.reference.set <- as.matrix(TargetCount.df[,
my.ref.samples])
#CNV sur chaque échantillon for (Sample in samples_data){
my.test <- TargetCount.df[,paste0(Sample,".recal.bam")] my.choice <- select.reference.set (test.counts = my.test,
reference.counts = my.reference.set,
bin.length = (TargetCount.df$end - TargetCount.df$start)/1000, n.bins.reduced = 10000)

my.matrix <- as.matrix( TargetCount.df[, my.choice$reference.choice, drop = FALSE]) my.reference.selected <-
apply(X = my.matrix, MAR = 1, FUN = sum)

all.exons <- new('ExomeDepth',
test = my.test,
reference = my.reference.selected, formula = 'cbind(test, reference) ~ 1')

all.exons <- CallCNVs(x = all.exons,
transition.probability = 10^-4,
chromosome = TargetCount.df$chromosome, start = TargetCount.df$start,
end = TargetCount.df$end,
name = TargetCount.df$exon)

if(nrow(all.exons@CNV.calls)>0){
output.file <- paste0(Sample,".cnv.txt")
write.table(file = file.path("/path/output_exomedepth",output.file),x =
cbind(Sample,all.exons@CNV.calls),row.names = FALSE,quote=F,sep="\t") }
}
```

Commande 14

```
#Créer les fichiers RPKM
python conifer.py rpkm \
--probes probes.txt \
#Conifer#
--input $name_sample_recal.bam \
--output RPKM/$name_sample.rpkm.txt
#Analyse
python $conifer analyze --probes $target_bed --rpkm_dir ${output_dir}/RPKM/ \
--output ${output_dir}/analysis_datas_exome.hdf5 \
--svd 2 \
--write_svals ${output_dir}/singular_values_datas_exome.txt \
--plot_scree ${output_dir}/screenplot_datas_exome.png \
--write_sd ${output_dir}/sd_values_datas_exome.txt
#Appel de variants
python $conifer call \
```

```
--input ${output_dir}/analysis_datas_exome.hdf5 \
--output ${output_dir}/call_datas_exome.txt
#Conversion call au format bed avec script maison pour analyse ultérieure
python ${scripts}/parse_conifer_output.py ${output_dir}/call_datas_exome.txt
$sample_name
#Plots
python $conifer plotcalls \
--input ${output_dir}/analysis_datas_exome.hdf5 \
--calls ${output_dir}/call_datas_exome.txt \
--outputdir ${output_dir}/call_imgs
```

Commande 15

```
#Assemblage des résultats conifer, exomedepth, et xhmm avec script maison >
merge_out_callers.bed
```

Commande 16

```
#création des différents dossier où seront stocker les sorties
mkdir results
mkdir -p results/filtre
mkdir DB/
#Mise en forme data exomedepth
for f in `ls *.txt`; do name=`echo $(basename $f) | sed 's/.cnv.txt//g'`; grep -v "Sample" $f | awk
'BEGIN{OFS="\t";cnv=""}{ if($4=="deletion"){cnv=1}else if($4=="duplication"){cnv=3} split($9,a,":");
split(a[2],b,"-"); print substr(a[1],4),b[1],b[2],cnv,$10}' > ${name}'_cnvScan.ready.bed'; done
for f in `ls *_cnvScan.ready.bed`; do name=`echo $(basename $f) | sed 's/_cnvScan.ready.bed//g'`;awk
'BEGIN{OFS="\t"}{print $0, "'$name'"}' $f >> DB/DB.txt; done
cd DB/
sort -k1,1 -k2n,2 -k3n,3 DB.txt > DB.txt.sorted.bed bgzip DB.txt.sorted.bed
tabix -p bed DB.txt.sorted.bed.gz
#run cnvscan
for f in `ls *_cnvScan.ready.bed`; do name=`echo $(basename $f) | sed 's/_cnvScan.ready.bed//g'`;python
/path/cnvScan_run.py -i $f -o results/${name}'_cnvScan.out.tab' - db DB/DB.txt.sorted.bed.gz -s /path/resources/;
done
#filtration des fichiers annotés
for f in results/*_cnvScan.out.tab; do name=`echo $(basename $f) | sed 's/_cnvScan.out.tab//g'`;python
/path/cnvScan_VarFilt.py -s 10 -i $f -o results/filtre/${name}'_filter_cnvScan.out.tab' ;done

#Étape supplémentaire par rapport à Exomedepth pour avoir un échantillon par fichier
for ident in $(grep -v "SAMPLE" cnv_xhmm.txt | cut -f1 | sort -u) ; do grep $ident cnv_xhmm.txt > all/${ident}.txt
;done
#Mise en forme data xhmm
for f in `ls all/*.txt`; do name=`echo $(basename $f) | sed 's/.txt//g'`; awk 'BEGIN{OFS="\t";cnv=2} {
if($2=="DEL"){cnv=1}else if($2=="DUP"){cnv=3} split($3,a,":"); split(a[2],b,"-"); print a[1], b[1], b[2], cnv,$10}' $f
| sed 's/^chr//g' > ${name}'_cnvScan.ready.bed'; done
for f in `ls *_cnvScan.ready.bed`; do name=`echo $(basename $f) | sed 's/_cnvScan.ready.bed//g'`;awk
'BEGIN{OFS="\t"}{print $0, "'$name'"}' $f >> DB/DB.txt; done
sort -k1,1 -k2n,2 -k3n,3 DB.txt > DB.txt.sorted.bed bgzip DB.txt.sorted.bed
tabix -p bed DB.txt.sorted.bed.gz
#run cnvscan
for f in `ls *_cnvScan.ready.bed`; do name=`echo $(basename $f) | sed 's/_cnvScan.ready.bed//g'`;python
/path/cnvScan_run.py -i $f -o results/${name}'_cnvScan.out.tab' - db DB/DB.txt.sorted.bed.gz -s /path/resources/;
done
#filtration des fichiers annotés
for f in results/*_cnvScan.out.tab; do name=`echo $(basename $f) | sed 's/_cnvScan.out.tab//g'`;python
/path/cnvScan_VarFilt.py -s 10 -i $f -o results/filtre/${name}'_filter_cnvScan.out.tab' ;done
```

Commande 17

```
#merge tous les échantillons d'exomedepth annotés par cnvscan dans un seul fichier
```

```
for f in `ls exomedepth/results/filtre/*_filter_cnvScan.out.tab`; do name=`echo $(basename $f) | sed 's/_filter_cnvScan.out.tab//g'`; grep -v "#chr" $f | sed 's/^/chr/g'| awk -F "\t" '{OFS=FS; if($4==1) {$4="DEL"}else if($4==3){$4="DUP"} print}'| awk '{print $0"\t""exomedepth_cnvscan"}' | awk 'BEGIN{OFS="\t"}{print $0, "'$name'"}'| awk -F'\t' -v OFS="\t" '{$5=$NF OFS $5;$NF=""}1' >> exomedepth/results/filtre/cnv_sample_exomedepth_cnvscan.tsv; done
#merge tous les échantillons d'xhmm annoté par cnvscan dans un seul fichier
for f in `ls results/filtre/*_filter_cnvScan.out.tab`; do name=`echo $(basename $f) | sed 's/_filter_cnvScan.out.tab//g'`; grep -v "#chr" $f | sed 's/^/chr/g'| awk -F "\t" '{OFS=FS; if($4==1) {$4="DEL"}else if($4==3){$4="DUP"} print}'| awk '{print $0"\t""xhmm_cnvscan"}' | awk 'BEGIN{OFS="\t"}{print $0, "'$name'"}'| awk -F'\t' -v OFS="\t" '{$5=$NF OFS $5;$NF=""}1' >> xhmm/results/filtre/cnv_sample_xhmm_cnvscan.tsv; done
#merge annotation xhmm + exomedepth par cnvscan
cat cnv_sample_exomedepth_cnvscan.tsv cnv_sample_xhmm_cnvscan.tsv | sort -k1,1V -k2,2n - k3,3n > cnv_sample_cnvscan.tsv
#Ajout colonne au fichier final annoté
sed -i '1s/^/#chr\tstart\tend\tcnv_state\tsample\tdefault_score\tlen\tinDB_count\tinDBScore_MinMaxMedian\tgene_name\tgene_type\tgene_id\texon_count\tUTR\ttranscript\tphastConElement_count\tphastConElement_minMax\thaplo_insufIdx_count\thaplo_insufIdx_score\tGene_intolarance_score\tsanger_cnv\tdgv_cnv\tdgv_varType\tdgv_varSubType\tdgv_pubmedId\tDGV_Stringency2_count\tDGV_Stringency2_PopFreq\tDGV_Stringency12_count\tDGV_Stringency12_popFreq\t1000g_del\t1000g_ins\tomim_morbidMap\tddd_mutConsequence\tddd_diseaseName\tddd_pubmedId\tclinVar_disease\thgvs_varName\tcaller_and_filter\n/' cnv_sample_cnvscan.tsv
```

Commande 18

```
#Annotation avec AnnotSV
export ANNOTSV=/path/AnnotSV_2.2
./AnnotSV.tcl -SVinputFile /path/merge_out_callers.bed -SVinputInfo 1 -outputFile /path/Annotation_merge_out_callers.tsv -svtBEDcol 4 -genomeBuild GRCh38
#Annotation avec database decipher avec script maison écrit en python
>Annotation_decipher_merge_out_callers.tsv
```

Annexe 6.

Liste des gènes indexés à la SLA dans la base de données NBCI Gene (juin 2019)

Symbol	description	other_designations	map_location
SETX	senataxin	probable helicase senataxin|SEN1 homolog|amyotrophic lateral sclerosis 4 protein	9q34.13
ALS2	ALS2, alsin Rho guanine nucleotide exchange factor	alsin|amyotrophic lateral sclerosis 2 (juvenile)|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 6 protein|amyotrophic lateral sclerosis 2 protein	2q33.1
C9orf72	chromosome 9 open reading frame 72	guanine nucleotide exchange C9orf72|protein C9orf72	9p21.2
SOD1	superoxide dismutase 1	superoxide dismutase [Cu-Zn]|Cu/Zn superoxide dismutase|SOD, soluble|epididymis secretory protein Li 44|indophenoloxidase A|superoxide dismutase 1, soluble|superoxide dismutase, cystolic	21q22.11
UNC13A	unc-13 homolog A	protein unc-13 homolog A	19p13.11
C2CD6	C2 calcium dependent domain containing 6	C2 calcium-dependent domain-containing protein 6|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 11|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 11 protein|amyotrophic lateral sclerosis 2 chromosome region candidate 11|testicular tissue protein Li 16	2q33.1
VCP	valosin containing protein	transitional endoplasmic reticulum ATPase|15S Mg(2+)-ATPase p97 subunit|TER ATPase	9p13.3
RAPH1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1	ras-associated and pleckstrin homology domains-containing protein 1|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 18|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 9|lamellipodin|proline-rich EVH1 ligand 2	2q33.2
OPTN	optineurin	optineurin|E3-14.7K-interacting protein|FIP-2|HIP-7|Huntingtin interacting protein L|huntingtin yeast partner L|huntingtin-interacting protein 7|huntingtin-interacting protein L|nemo-related protein|optic neuropathy-inducing protein|transcription factor IIIA-interacting protein|transcrption factor IIIA-interacting protein|tumor necrosis factor alpha-inducible cellular protein containing leucine zipper domains	10p13
ERBB4	erb-b2 receptor tyrosine kinase 4	receptor tyrosine-protein kinase erbB-4|ERBB4 transcript variant I12DEL|ERBB4 transcript variant I20DEL|avian erythroblastic leukemia viral (v-erb-b2) oncogene homolog 4|human epidermal growth factor receptor 4|proto-oncogene-like protein c-ErbB-4|tyrosine kinase-type cell surface receptor HER4|v-erb-a erythroblastic leukemia viral oncogene homolog 4|v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4	2q34
STRADB	STE20 related adaptor beta	STE20-related kinase adapter protein beta|ILP-interacting protein ILPIPA|STE20-related kinase adaptor beta|STRAD beta|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 2 protein|pseudokinase ALS2CR2	2q33.1
TMEM237	transmembrane protein 237	transmembrane protein 237|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 4|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 4 protein	2q33.1
UBQLN2	ubiquilin 2	ubiquilin-2|Nedd4 binding protein 4|protein linking IAP with cytoskeleton 2|ubiquitin-like product Chap1/Dsk2	Xp11.21
MPP4	membrane palmitoylated	MAGUK p55 subfamily member 4|amyotrophic lateral sclerosis 2 chromosomal	2q33.1

	protein 4	region candidate gene 5 protein|discs large homolog 6|membrane protein, palmitoylated 4 (MAGUK p55 subfamily member 4)	
CFLAR-AS1	CFLAR antisense RNA 1	CFLAR antisense RNA 1 (non-protein coding)|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 10	2q33.1
TARDBP	TAR DNA binding protein	TAR DNA-binding protein 43|TAR DNA-binding protein-43	1p36.22
FUS	FUS RNA binding protein	RNA-binding protein FUS|75 kDa DNA-pairing protein|FUS/ERG chimeric protein|FUS/ERG fusion protein|fus-like protein|fused in sarcoma|fusion gene in myxoid liposarcoma|heterogeneous nuclear ribonucleoprotein P2|oncogene FUS|oncogene TLS|translocated in liposarcoma protein	16p11.2
ANG	angiogenin	angiogenin|RNase 5|angiogenin, ribonuclease, RNase A family, 5|epididymis luminal protein 168|ribonuclease 5|ribonuclease A A1|ribonuclease A family member 5	14q11.2
HNRNPA1	heterogeneous nuclear ribonucleoprotein A1	heterogeneous nuclear ribonucleoprotein A1|epididymis secretory sperm binding protein|helix-destabilizing protein|heterogeneous nuclear ribonucleoprotein A1B protein|heterogeneous nuclear ribonucleoprotein B2 protein|heterogeneous nuclear ribonucleoprotein core protein A1|hnRNP core protein A1-like 3|nuclear ribonucleoprotein particle A1 protein|putative heterogeneous nuclear ribonucleoprotein A1-like 3|single-strand DNA-binding protein UP1|single-strand RNA-binding protein	12q13.13
TBK1	TANK binding kinase 1	serine/threonine-protein kinase TBK1|NF-kB-activating kinase|NF-kappa-B-activating kinase	12q14.2
PFN1	profilin 1	profilin-1|epididymis tissue protein Li 184a|profilin I	17p13.2
VAPB	VAMP associated protein B and C	vesicle-associated membrane protein-associated protein B/C|VAMP (vesicle-associated membrane protein)-associated protein B and C|VAMP-associated 33 kDa protein	20q13.32
NEFH	neurofilament heavy	neurofilament heavy polypeptide|200 kDa neurofilament protein|NF-H|neurofilament triplet H protein|neurofilament, heavy polypeptide 200kDa	22q12.2
SQSTM1	sequestosome 1	sequestosome-1|EBI3-associated protein of 60 kDa|EBI3-associated protein p60|EBIAP|oxidative stress induced like|phosphotyrosine independent ligand for the Lck SH2 domain p62|phosphotyrosine-independent ligand for the Lck SH2 domain of 62 kDa|ubiquitin-binding protein p62	5q35.3
CHCHD10	coiled-coil-helix-coiled-coil-helix domain containing 10	coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial|MIX17 homolog A	22q11.23
DCTN1	dynactin subunit 1	dynactin subunit 1|150 kDa dynein-associated polypeptide|dynactin 1 (p150, glued homolog, Drosophila)	2p13.1
SIGMAR1	sigma non-opioid intracellular receptor 1	sigma non-opioid intracellular receptor 1|SR31747 binding protein 1|aging-associated gene 8 protein|sigma 1-type opioid receptor	9p13.3
SPG11	SPG11, spatacsin vesicle trafficking associated	spatacsin|colorectal carcinoma-associated protein|spastic paraplegia 11 (autosomal recessive)|spastic paraplegia 11 protein	15q21.1
PRPH	peripherin	peripherin|neurofilament 4 (57kD)	12q13.12
CHMP2B	charged multivesicular body protein 2B	charged multivesicular body protein 2b|VPS2 homolog B|chromatin modifying protein 2B|vacuolar protein-sorting-associated protein 2-2	3p11.2
DPP6	dipeptidyl peptidase like 6	dipeptidyl aminopeptidase-like protein 6|DPP VI|dipeptidyl aminopeptidase IV-	7q36.2

		related protein|dipeptidyl peptidase IV-related protein|dipeptidyl peptidase VI|dipeptidyl-peptidase 6	
MATR3	matrin 3	matrin-3|vocal cord and pharyngeal weakness with distal myopathy	5q31.2
DISC1	DISC1 scaffold protein	disrupted in schizophrenia 1 protein	1q42.2
FIG4	FIG4 phosphoinositide 5-phosphatase	polyphosphoinositide phosphatase|FIG4 homolog, SAC domain containing lipid phosphatase|FIG4 homolog, SAC1 domain containing lipid phosphatase|FIG4 homolog, SAC1 lipid phosphatase domain containing|SAC domain-containing protein 3|Sac domain-containing inositol phosphatase 3|phosphatidylinositol 3,5-bisphosphate 5-phosphatase	6q21
TRPM7	transient receptor potential cation channel subfamily M member 7	transient receptor potential cation channel subfamily M member 7|LTRPC ion channel family member 7|channel-kinase 1|long transient receptor potential channel 7|transient receptor potential-phospholipase C-interacting kinase	15q21.2
SDC1	syndecan 1	syndecan-1|CD138 antigen|heparan sulfate proteoglycan fibroblast growth factor receptor|syndecan proteoglycan 1	2p24.1
LIPC	lipase C, hepatic type	hepatic triacylglycerol lipase|Triacylglycerol lipase|hepatic lipase|lipase member C|lipase, hepatic	15q21.3
KIFAP3	kinesin associated protein 3	kinesin-associated protein 3|small G protein GDP dissociation stimulator|smg GDS-associated protein	1q24.2
MAP3K7	mitogen-activated protein kinase kinase kinase 7	mitogen-activated protein kinase kinase kinase 7|TGF-beta activated kinase 1|transforming growth factor-beta-activated kinase 1	6q15
PROCR	protein C receptor	endothelial protein C receptor|APC receptor|CD201 antigen|activated protein C receptor|cell cycle, centrosome-associated protein|centrocyclin|protein C receptor, endothelial	20q11.22
PAX3	paired box 3	paired box protein Pax-3|paired box homeotic gene 3|paired domain gene 3|paired domain gene HuP2|transcriptional factor PAX3	2q36.1
SELL	selectin L	L-selectin|CD62 antigen-like family member L|gp90-MEL|leukocyte surface antigen Leu-8|leukocyte-endothelial cell adhesion molecule 1|lymph node homing receptor|lymphocyte adhesion molecule 1|pln homing receptor	1q24.2
TFAP2A	transcription factor AP-2 alpha	transcription factor AP-2-alpha|AP-2 transcription factor|AP2-alpha|activating enhancer-binding protein 2-alpha|activator protein 2|transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	6p24.3
NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase	E3 ubiquitin-protein ligase NEDD4-like|HECT-type E3 ubiquitin transferase NED4L|ubiquitin-protein ligase Rsp5	18q21.31
ALCAM	activated leukocyte cell adhesion molecule	CD166 antigen	3q13.11
ITPR2	inositol 1,4,5-trisphosphate receptor type 2	inositol 1,4,5-trisphosphate receptor type 2|IP3 receptor|IP3R 2|cilia and flagella associated protein 48|type 2 InsP3 receptor	12p11.23
TUBA4A	tubulin alpha 4a	tubulin alpha-4A chain|tubulin H2-alpha|tubulin alpha-1 chain|tubulin, alpha 1 (testis specific)	2q35
SMARCA2	SWI/SNF related, matrix associated, actin dependent	probable global transcription activator SNF2L2|ATP-dependent helicase SMARCA2|BAF190B|BRG1-associated factor 190B|SNF2-alpha|SNF2/SWI2-like	9p24.3

	regulator of chromatin, subfamily a, member 2	protein 2|SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a2|brahma homolog|global transcription activator homologous sequence|protein brahma homolog|sucrose nonfermenting 2-like protein 2	
EFEMP1	EGF containing fibulin extracellular matrix protein 1	EGF-containing fibulin-like extracellular matrix protein 1|EGF containing fibulin like extracellular matrix protein 1|extracellular protein S1-5|fibulin-3	2p16.1
TMPRSS2	transmembrane serine protease 2	transmembrane protease serine 2|epitheliasin|serine protease 10|transmembrane protease, serine 2	21q22.3
NEK1	NIMA related kinase 1	serine/threonine-protein kinase Nek1|NIMA (never in mitosis gene a)-related kinase 1|never in mitosis A-related kinase 1|nimA-related protein kinase 1|protein-serine/threonine kinase|renal carcinoma antigen NY-REN-55	4q33
EPB41	erythrocyte membrane protein band 4.1	protein 4.1|EPB4.1|P4.1|band 4.1|elliptocytosis 1, RH-linked|erythrocyte surface protein band 4.1	1p35.3
ATXN1	ataxin 1	ataxin-1|alternative ataxin1|spinocerebellar ataxia type 1 protein	6p22.3
ANK3	ankyrin 3	ankyrin-3|ankyrin 3, node of Ranvier (ankyrin G)	10q21.2
LAMA3	laminin subunit alpha 3	laminin subunit alpha-3|BM600 150kD subunit|LAM3, alpha-3 subunit|epiligrin 170 kda subunit|epiligrin alpha 3 subunit|kalinin 165kD subunit|kalinin subunit alpha|laminin 5, alpha-3 subunit|laminin A3|laminin, alpha 3 (nicein (150kD), kalinin (165kD), BM600 (150kD), epilegrin)|laminin-5 alpha 3 chain|laminin-5 subunit alpha|laminin-6 subunit alpha|laminin-7 subunit alpha|nicein 150kD subunit|nicein subunit alpha	18q11.2
NOG	noggin	noggin|symphalangism 1 (proximal)	17q22
MASP1	mannan binding lectin serine peptidase 1	mannan-binding lectin serine protease 1|C4/C2 activating component of Ra-reactive factor|Ra-reactive factor serine protease p100|complement factor MASP-3|complement-activating component of Ra-reactive factor|mannose-binding lectin-associated serine protease 1|mannose-binding protein-associated serine protease|serine protease 5	3q27.3
LAMA2	laminin subunit alpha 2	laminin subunit alpha-2|laminin M chain|laminin, alpha 2|laminin-12 subunit alpha|laminin-2 subunit alpha|laminin-4 subunit alpha|merosin heavy chain|mutant laminin subunit alpha 2	6q22.33
BMPR1B	bone morphogenetic protein receptor type 1B	bone morphogenetic protein receptor type-1B|BMP type-1B receptor|BMPR-1B|activin receptor-like kinase 6|bone morphogenetic protein receptor, type IB|serine/threonine receptor kinase	4q22.3
TYRP1	tyrosinase related protein 1	5,6-dihydroxyindole-2-carboxylic acid oxidase|DHICA oxidase|catalase B|glycoprotein 75|melanoma antigen gp75	9p23
TNFRSF1B	TNF receptor superfamily member 1B	tumor necrosis factor receptor superfamily member 1B|TNF-R2|TNF-RII|p75 TNF receptor|p80 TNF-alpha receptor|soluble TNFR1B variant 1|tumor necrosis factor beta receptor|tumor necrosis factor binding protein 2|tumor necrosis factor receptor 2|tumor necrosis factor receptor type II	1p36.22
CCR2	C-C motif chemokine receptor 2	C-C chemokine receptor type 2|MCP-1 receptor|chemokine (C-C motif) receptor 2|monocyte chemoattractant protein 1 receptor|monocyte chemotactic protein 1 receptor	3p21.31
NOTCH3	notch 3	neurogenic locus notch homolog protein 3|Notch homolog 3	19p13.12

PARK7	Parkinsonism associated deglycase	protein/nucleic acid deglycase DJ-1|Parkinson disease (autosomal recessive, early onset) 7|epididymis secretory sperm binding protein Li 67p|maillard deglycase|oncogene DJ1|parkinson protein 7|protein DJ-1|protein deglycase DJ-1	1p36.23
PIN1	peptidylprolyl cis/trans isomerase, NIMA-interacting 1	peptidyl-prolyl cis-trans isomerase NIMA-interacting 1|PPIase Pin1|protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1|rotamase Pin1	19p13.2
C9orf72	chromosome 9 open reading frame 72	guanine nucleotide exchange C9orf72|protein C9orf72	9p21.2
VCP	valosin containing protein	transitional endoplasmic reticulum ATPase|15S Mg(2+)-ATPase p97 subunit|TER ATPase	9p13.3
FUS	FUS RNA binding protein	RNA-binding protein FUS|75 kDa DNA-pairing protein|FUS/ERG chimeric protein|FUS/ERG fusion protein|fus-like protein|fused in sarcoma|fusion gene in myxoid liposarcoma|heterogeneous nuclear ribonucleoprotein P2|oncogene FUS|oncogene TLS|translocated in liposarcoma protein	16p11.2
PINK1	PTEN induced kinase 1	serine/threonine-protein kinase PINK1, mitochondrial|PTEN induced putative kinase 1|PTEN-induced putative kinase protein 1|protein kinase BRPK	1p36.12
NOS1	nitric oxide synthase 1	nitric oxide synthase, brain|NOS type I|constitutive NOS|neuronal NOS|nitric oxide synthase 1 (neuronal)|peptidyl-cysteine S-nitrosylase NOS1	12q24.22
SDC1	syndecan 1	syndecan-1|CD138 antigen|heparan sulfate proteoglycan fibroblast growth factor receptor|syndecan proteoglycan 1	2p24.1
CYBB	cytochrome b-245 beta chain	cytochrome b-245 heavy chain|CGD91-phox|NADPH oxidase 2|cytochrome b(558) subunit beta|cytochrome b-245 beta polypeptide|cytochrome b558 subunit beta|heme-binding membrane glycoprotein gp91phox|neutrophil cytochrome b 91 kDa polypeptide|p22 phagocyte B-cytochrome|superoxide-generating NADPH oxidase heavy chain subunit	Xp21.1-p11.4
HSPA8	heat shock protein family A (Hsp70) member 8	heat shock cognate 71 kDa protein|LPS-associated protein 1|N-myristoyltransferase inhibitor protein 71|constitutive heat shock protein 70|epididymis luminal protein 33|epididymis secretory sperm binding protein Li 72p|heat shock 70kDa protein 8|heat shock 70kd protein 10|heat shock cognate protein 54|lipopolysaccharide-associated protein 1	11q24.1
SMN1	survival of motor neuron 1, telomeric	survival motor neuron protein|component of gems 1|gemin-1|survival motor neuron 1 protein|survival of motor neuron 1 isoform D2A2B345|survival of motor neuron 1 isoform D2A2B3457|survival of motor neuron 1 isoform D2A3457|survival of motor neuron 1 isoform D2B3457|survival of motor neuron 1 isoform D345|survival of motor neuron 1 isoform D3457|survival of motor neuron 1 isoform D347|tudor domain containing 16A	5q13.2
ERBB4	erb-b2 receptor tyrosine kinase 4	receptor tyrosine-protein kinase erbB-4|ERBB4 transcript variant I12DEL|ERBB4 transcript variant I20DEL|avian erythroblastic leukemia viral (v-erb-b2) oncogene homolog 4|human epidermal growth factor receptor 4|proto-oncogene-like protein c-ErbB-4|tyrosine kinase-type cell surface receptor HER4|v-erb-a erythroblastic leukemia viral oncogene homolog 4|v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4	2q34
GZMB	granzyme B	granzyme B|T-cell serine protease 1-3E|cathepsin G-like 1|cytotoxic T-lymphocyte	14q12

		proteinase 2|cytotoxic serine protease B|fragmentin 2|granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)|human lymphocyte protein	
CTSD	cathepsin D	cathepsin D|ceroid-lipofuscinosis, neuronal 10|epididymis secretory sperm binding protein Li 130P|lysosomal aspartyl peptidase|lysosomal aspartyl protease	11p15.5
LIPC	lipase C, hepatic type	hepatic triacylglycerol lipase|Triacylglycerol lipase|hepatic lipase|lipase member C|lipase, hepatic	15q21.3
TIMP3	TIMP metallopeptidase inhibitor 3	metalloproteinase inhibitor 3|MIG-5 protein|TIMP-3|protein MIG-5|tissue inhibitor of metalloproteinases 3	22q12.3
OPA1	OPA1, mitochondrial dynamin like GTPase	dynamin-like 120 kDa protein, mitochondrial|dynamin-like guanosine triphosphatase|mitochondrial dynamin-like GTPase|optic atrophy 1 (autosomal dominant)|optic atrophy protein 1	3q29
DISC1	DISC1 scaffold protein	disrupted in schizophrenia 1 protein	1q42.2
HDAC6	histone deacetylase 6	histone deacetylase 6|protein phosphatase 1, regulatory subunit 90	Xp11.23
KRT18	keratin 18	keratin, type I cytoskeletal 18|cell proliferation-inducing gene 46 protein|cytokeratin 18|keratin 18, type I	12q13.13
EWSR1	EWS RNA binding protein 1	RNA-binding protein EWS|EWS RNA-binding protein variant 6|Ewing sarcoma breakpoint region 1|Ewings sarcoma EWS-Fli1 (type 1) oncogene	22q12.2
AIFM1	apoptosis inducing factor mitochondria associated 1	apoptosis-inducing factor 1, mitochondrial|apoptosis-inducing factor, mitochondrion-associated, 1|programmed cell death 8 (apoptosis-inducing factor)|striatal apoptosis-inducing factor|testicular secretory protein Li 4	Xq26.1
MAP3K7	mitogen-activated protein kinase kinase kinase 7	mitogen-activated protein kinase kinase kinase 7|TGF-beta activated kinase 1|transforming growth factor-beta-activated kinase 1	6q15
DNM1L	dynamin 1 like	dynamin-1-like protein|Dnm1p/Vps1p-like protein|dynamin family member proline-rich carboxyl-terminal domain less|dynamin-like protein 4|dynamin-like protein IV|dynamin-related protein 1	12p11.21
FXN	frataxin	frataxin, mitochondrial|Friedreich ataxia protein	9q21.11
PROCR	protein C receptor	endothelial protein C receptor|APC receptor|CD201 antigen|activated protein C receptor|cell cycle, centrosome-associated protein|centrocyclin|protein C receptor, endothelial	20q11.22
CP	ceruloplasmin	ceruloplasmin|ceruloplasmin (ferroxidase)	3q24-q25.1
LOX	lysyl oxidase	protein-lysine 6-oxidase	5q23.1
PAX3	paired box 3	paired box protein Pax-3|paired box homeotic gene 3|paired domain gene 3|paired domain gene HuP2|transcriptional factor PAX3	2q36.1
ANG	angiogenin	angiogenin|RNase 5|angiogenin, ribonuclease, RNase A family, 5|epididymis luminal protein 168|ribonuclease 5|ribonuclease A A1|ribonuclease A family member 5	14q11.2
DDIT3	DNA damage inducible transcript 3	DNA damage-inducible transcript 3 protein|C/EBP zeta|CCAAT/enhancer-binding protein homologous protein|alternative DDIT3 protein|c/EBP-homologous protein 10|growth arrest and DNA damage-inducible protein GADD153	12q13.3
HNRNPA1	heterogeneous nuclear ribonucleoprotein A1	heterogeneous nuclear ribonucleoprotein A1|epididymis secretory sperm binding protein|helix-destabilizing protein|heterogeneous nuclear ribonucleoprotein A1B protein|heterogeneous nuclear ribonucleoprotein B2 protein|heterogeneous nuclear ribonucleoprotein core protein A1|hnRNP core protein A1-like 3|nuclear	12q13.13

		ribonucleoprotein particle A1 protein|putative heterogeneous nuclear ribonucleoprotein A1-like 3|single-strand DNA-binding protein UP1|single-strand RNA-binding protein	
SELL	selectin L	L-selectin|CD62 antigen-like family member L|gp90-MEL|leukocyte surface antigen Leu-8|leukocyte-endothelial cell adhesion molecule 1|lymph node homing receptor|lymphocyte adhesion molecule 1|pln homing receptor	1q24.2
TBK1	TANK binding kinase 1	serine/threonine-protein kinase TBK1|NF-kB-activating kinase|NF-kappa-B-activating kinase	12q14.2
CHRNA4	cholinergic receptor nicotinic alpha 4 subunit	neuronal acetylcholine receptor subunit alpha-4|cholinergic receptor, nicotinic alpha 4|cholinergic receptor, nicotinic, alpha 4 (neuronal)|cholinergic receptor, nicotinic, alpha polypeptide 4|neuronal nicotinic acetylcholine receptor alpha-4 subunit	20q13.33
PLA2G4A	phospholipase A2 group IVA	cytosolic phospholipase A2|calcium-dependent phospholipid-binding protein|lysophospholipase|phosphatidylcholine 2-acylhydrolase|phospholipase A2, group IVA (cytosolic, calcium-dependent)	1q31.1
SMN2	survival of motor neuron 2, centromeric	survival motor neuron protein|component of gems 1|gemin-1|survival of motor neuron 2 isoform D2A2B345|survival of motor neuron 2 isoform D2A2B3457|survival of motor neuron 2 isoform D2A3457|survival of motor neuron 2 isoform D2B3457|survival of motor neuron 2 isoform D345|survival of motor neuron 2 isoform D3457|survival of motor neuron 2 isoform D347|tudor domain containing 16B	5q13.2
TREM2	triggering receptor expressed on myeloid cells 2	triggering receptor expressed on myeloid cells 2|triggering receptor expressed on monocytes 2|triggering receptor expressed on myeloid cells 2a	6p21.1
TRPM7	transient receptor potential cation channel subfamily M member 7	transient receptor potential cation channel subfamily M member 7|LTRPC ion channel family member 7|channel-kinase 1|long transient receptor potential channel 7|transient receptor potential-phospholipase C-interacting kinase	15q21.2
TFAP2A	transcription factor AP-2 alpha	transcription factor AP-2-alpha|AP-2 transcription factor|AP2-alpha|activating enhancer-binding protein 2-alpha|activator protein 2|transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	6p24.3
VDAC1	voltage dependent anion channel 1	voltage-dependent anion-selective channel protein 1|outer mitochondrial membrane protein porin 1|plasmalemmal porin|porin 31HL|porin 31HM|sperm binding protein 1a	5q31.1
TFAM	transcription factor A, mitochondrial	transcription factor A, mitochondrial|mitochondrial transcription factor 1|mitochondrial transcription factor A|transcription factor 6|transcription factor 6-like 1|transcription factor 6-like 2 (mitochondrial transcription factor)|transcription factor 6-like 3	10q21.1
CHRNA3	cholinergic receptor nicotinic alpha 3 subunit	neuronal acetylcholine receptor subunit alpha-3|cholinergic receptor, nicotinic alpha 3|cholinergic receptor, nicotinic, alpha 3 (neuronal)|cholinergic receptor, nicotinic, alpha polypeptide 3|neuronal nicotinic acetylcholine receptor, alpha3 subunit	15q25.1
HDAC4	histone deacetylase 4	histone deacetylase 4|histone deacetylase A	2q37.3
PRMT1	protein arginine methyltransferase 1	protein arginine N-methyltransferase 1|HMT1 (hnRNP methyltransferase, S. cerevisiae)-like 2|heterogeneous nuclear ribonucleoprotein methyltransferase 1-	19q13.33

		like 2|histone-arginine N-methyltransferase PRMT1|interferon receptor 1-bound protein 4	
OPTN	optineurin	optineurin|E3-14.7K-interacting protein|FIP-2|HIP-7|Huntingtin interacting protein L|huntingtin yeast partner L|huntingtin-interacting protein 7|huntingtin-interacting protein L|nemo-related protein|optic neuropathy-inducing protein|transcription factor IIIA-interacting protein|transcrption factor IIIA-interacting protein|tumor necrosis factor alpha-inducible cellular protein containing leucine zipper domains	10p13
PON2	paraoxonase 2	serum paraoxonase/arylesterase 2|A-esterase 2|PON 2|aromatic esterase 2|arylesterase 2|paraoxonase nirs|serum aryldialkylphosphatase 2	7q21.3
FTL	ferritin light chain	ferritin light chain|epididymis secretory sperm binding protein|ferritin L subunit|ferritin L-chain|ferritin light polypeptide-like 3|ferritin, light polypeptide	19q13.33
GLRA1	glycine receptor alpha 1	glycine receptor subunit alpha-1|glycine receptor 48 kDa subunit|glycine receptor strychnine-binding subunit	5q33.1
NEDD4L	neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin protein ligase	E3 ubiquitin-protein ligase NEDD4-like|HECT-type E3 ubiquitin transferase NED4L|ubiquitin-protein ligase Rsp5	18q21.31
ATXN2	ataxin 2	ataxin-2|spinocerebellar ataxia type 2 protein|trinucleotide repeat-containing gene 13 protein	12q24.12
ALCAM	activated leukocyte cell adhesion molecule	CD166 antigen	3q13.11
SLC1A2	solute carrier family 1 member 2	excitatory amino acid transporter 2|excitotoxic amino acid transporter 2|glutamate/aspartate transporter II|sodium-dependent glutamate/aspartate transporter 2|solute carrier family 1 (glial high affinity glutamate transporter), member 2	11p13
NEFL	neurofilament light	neurofilament light polypeptide|light molecular weight neurofilament protein|neurofilament protein, light chain|neurofilament subunit NF-L|neurofilament triplet L protein|neurofilament, light polypeptide 68kDa|protein phosphatase 1, regulatory subunit 110	8p21.2
PABPC1	poly(A) binding protein cytoplasmic 1	polyadenylate-binding protein 1|poly(A) binding protein, cytoplasmic 2	8q22.3
PFN1	profilin 1	profilin-1|epididymis tissue protein Li 184a|profilin I	17p13.2
RTN4	reticulon 4	reticulon-4|Human NogoA|My043 protein|foocen|neurite growth inhibitor 220|neurite outgrowth inhibitor|neuroendocrine-specific protein C homolog|reticulon 5	2p16.1
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	heterogeneous nuclear ribonucleoproteins A2/B1|HNRNPA2B1/MYC fusion|epididymis secretory sperm binding protein|hnRNP A2 / hnRNP B1|nuclear ribonucleoprotein particle A2 protein	7p15.2
TXNRD1	thioredoxin reductase 1	thioredoxin reductase 1, cytoplasmic|KM-102-derived reductase-like factor|gene associated with retinoic and IFN-induced mortality 12 protein|gene associated with retinoic and interferon-induced mortality 12 protein|oxidoreductase|testis tissue sperm-binding protein Li 46a|thioredoxin reductase GRIM-12|thioredoxin	12q23.3

		reductase TR1	
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	probable global transcription activator SNF2L2|ATP-dependent helicase SMARCA2|BAF190B|BRG1-associated factor 190B|SNF2-alpha|SNF2/SWI2-like protein 2|SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a2|brahma homolog|global transcription activator homologous sequence|protein brahma homolog|sucrose nonfermenting 2-like protein 2	9p24.3
AKT3	AKT serine/threonine kinase 3	RAC-gamma serine/threonine-protein kinase|PKB gamma|RAC-gamma serine/threonine protein kinase|v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)	1q43-q44
DCTN1	dynactin subunit 1	dynactin subunit 1|150 kDa dynein-associated polypeptide|dynactin 1 (p150, glued homolog, Drosophila)	2p13.1
EFEMP1	EGF containing fibulin extracellular matrix protein 1	EGF-containing fibulin-like extracellular matrix protein 1|EGF containing fibulin like extracellular matrix protein 1|extracellular protein S1-5|fibulin-3	2p16.1
GRM1	glutamate metabotropic receptor 1	metabotropic glutamate receptor 1|glutamate receptor, metabotropic 1|protein phosphatase 1, regulatory subunit 85	6q24.3
PSAP	prosaposin	prosaposin|proactivator polypeptide|sphingolipid activator protein-1	10q22.1
TRPM2	transient receptor potential cation channel subfamily M member 2	transient receptor potential cation channel subfamily M member 2|estrogen-responsive element-associated gene 1 protein|long transient receptor potential channel 2|transient receptor potential channel 7|transient receptor potential melastatin 2	21q22.3
TMPRSS2	transmembrane serine protease 2	transmembrane protease serine 2|epitheliasin|serine protease 10|transmembrane protease, serine 2	21q22.3
SIGMAR1	sigma non-opioid intracellular receptor 1	sigma non-opioid intracellular receptor 1|SR31747 binding protein 1|aging-associated gene 8 protein|sigma 1-type opioid receptor	9p13.3
EPB41	erythrocyte membrane protein band 4.1	protein 4.1|EPB4.1|P4.1|band 4.1|elliptocytosis 1, RH-linked|erythrocyte surface protein band 4.1	1p35.3
ATXN1	ataxin 1	ataxin-1|alternative ataxin1|spinocerebellar ataxia type 1 protein	6p22.3
CYP27A1	cytochrome P450 family 27 subfamily A member 1	sterol 26-hydroxylase, mitochondrial|5-beta-cholestane-3-alpha, 7-alpha, 12-alpha-triol 26-hydroxylase|5-beta-cholestane-3-alpha, 7-alpha, 12-alpha-triol 27-hydroxylase|cholestanetriol 26-monooxygenase|cytochrome P-450C27/25|cytochrome P450 27|cytochrome P450, family 27, subfamily A, polypeptide 1|cytochrome P450, subfamily XXVIIA (steroid 27-hydroxylase, cerebrotendinous xanthomatosis), polypeptide 1|sterol 27-hydroxylase|vitamin D(3) 25-hydroxylase	2q35
P4HB	prolyl 4-hydroxylase subunit beta	protein disulfide-isomerase|cellular thyroid hormone-binding protein|collagen prolyl 4-hydroxylase beta|glutathione-insulin transhydrogenase|p55|procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide|prolyl 4-hydroxylase, beta polypeptide|protein disulfide isomerase family A, member 1|protein disulfide isomerase-associated 1|protein disulfide isomerase/oxidoreductase|protocollagen hydroxylase|testicular secretory protein Li 32|thyroid hormone-binding protein p55	17q25.3
MYO6	myosin VI	unconventional myosin-VI|myosin VI transcript variant 007|myosin VI transcript	6q14.1

		variant 008|unconventional myosin-6	
BCL11A	BCL11A, BAF complex component	B-cell lymphoma/leukemia 11A|B cell CLL/lymphoma 11A|B-cell CLL/lymphoma 11A (zinc finger protein) isoform 2|BCL11A B-cell CLL/lymphoma 11A (zinc finger protein) isoform 1|C2H2-type zinc finger protein|COUP-TF-interacting protein 1|ecotropic viral integration site 9 homolog|ecotropic viral integration site 9 protein homolog|zinc finger protein 856	2p16.1
GRIA1	glutamate ionotropic receptor AMPA type subunit 1	glutamate receptor 1|AMPA 1|AMPA-selective glutamate receptor 1|gluR-1|gluR-A|gluR-K1|glutamate receptor, ionotropic, AMPA 1	5q33.2
SLC11A2	solute carrier family 11 member 2	natural resistance-associated macrophage protein 2|DMT-1|NRAMP 2|divalent cation transporter 1|solute carrier family 11 (proton-coupled divalent metal ion transporter), member 2|solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	12q13.12
ANK3	ankyrin 3	ankyrin-3|ankyrin 3, node of Ranvier (ankyrin G)	10q21.2
LAMA3	laminin subunit alpha 3	laminin subunit alpha-3|BM600 150kD subunit|LAM3, alpha-3 subunit|epiligrin 170 kda subunit|epiligrin alpha 3 subunit|kalinin 165kD subunit|kalinin subunit alpha|laminin 5, alpha-3 subunit|laminin A3|laminin, alpha 3 (nicein (150kD), kalinin (165kD), BM600 (150kD), epilegrin)|laminin-5 alpha 3 chain|laminin-5 subunit alpha|laminin-6 subunit alpha|laminin-7 subunit alpha|nicein 150kD subunit|nicein subunit alpha	18q11.2
HSPB8	heat shock protein family B (small) member 8	heat shock protein beta-8|E2-induced gene 1 protein|alpha-crystallin C chain|heat shock 22kDa protein 8|heat shock 27kDa protein 8|protein kinase H11|small stress protein-like protein HSP22	12q24.23
NOG	noggin	noggin|symphalangism 1 (proximal)	17q22
VAPB	VAMP associated protein B and C	vesicle-associated membrane protein-associated protein B/C|VAMP (vesicle-associated membrane protein)-associated protein B and C|VAMP-associated 33 kDa protein	20q13.32
CNTF	ciliary neurotrophic factor	ciliary neurotrophic factor	11q12.1
MASP1	mannan binding lectin serine peptidase 1	mannan-binding lectin serine protease 1|C4/C2 activating component of Ra-reactive factor|Ra-reactive factor serine protease p100|complement factor MASP-3|complement-activating component of Ra-reactive factor|mannose-binding lectin-associated serine protease 1|mannose-binding protein-associated serine protease|serine protease 5	3q27.3
UBQLN1	ubiquilin 1	ubiquilin-1|hPLIC-1|protein linking IAP with cytoskeleton 1|testicular tissue protein Li 219	9q21.2-q21.3
HK1	hexokinase 1	hexokinase-1|brain form hexokinase|glycolytic enzyme|hexokinase IR|hexokinase type I	10q22.1
ATL1	atlastin GTPase 1	atlastin-1|GBP-3|GTP-binding protein 3|brain-specific GTP-binding protein|guanine nucleotide-binding protein 3|guanylate-binding protein 3|hGBP3|spastic paraplegia 3 protein A	14q22.1
TIA1	TIA1 cytotoxic granule associated RNA binding protein	nucleolysin TIA-1 isoform p40|nucleolysin TIA-1|T-cell-restricted intracellular antigen-1|p40-TIA-1 (containing p15-TIA-1)	2p13.3

LAMA2	laminin subunit alpha 2	laminin subunit alpha-2|laminin M chain|laminin, alpha 2|laminin-12 subunit alpha|laminin-2 subunit alpha|laminin-4 subunit alpha|merosin heavy chain|mutant laminin subunit alpha 2	6q22.33
SPG11	SPG11, spatacsin vesicle trafficking associated	spatacsin|colorectal carcinoma-associated protein|spastic paraplegia 11 (autosomal recessive)|spastic paraplegia 11 protein	15q21.1
BMPR1B	bone morphogenetic protein receptor type 1B	bone morphogenetic protein receptor type-1B|BMP type-1B receptor|BMPR-1B|activin receptor-like kinase 6|bone morphogenetic protein receptor, type IB|serine/threonine receptor kinase	4q22.3
TYRP1	tyrosinase related protein 1	5,6-dihydroxyindole-2-carboxylic acid oxidase|DHICA oxidase|catalase B|glycoprotein 75|melanoma antigen gp75	9p23
TUBA4A	tubulin alpha 4a	tubulin alpha-4A chain|tubulin H2-alpha|tubulin alpha-1 chain|tubulin, alpha 1 (testis specific)	2q35
LAMC1	laminin subunit gamma 1	laminin subunit gamma-1|S-LAM gamma|S-laminin subunit gamma|laminin B2 chain|laminin, gamma 1 (formerly LAMB2)|laminin-10 subunit gamma|laminin-11 subunit gamma|laminin-2 subunit gamma|laminin-3 subunit gamma|laminin-4 subunit gamma|laminin-6 subunit gamma|laminin-7 subunit gamma|laminin-8 subunit gamma|laminin-9 subunit gamma	1q25.3
CACNA1S	calcium voltage-gated channel subunit alpha1 S	voltage-dependent L-type calcium channel subunit alpha-1S|calcium channel, L type, alpha 1 polypeptide, isoform 3 (skeletal muscle, hypokalemic periodic paralysis)|calcium channel, voltage-dependent, L type, alpha 1S subunit|dihydropyridine receptor alpha 1 subunit|dihydropyridine-sensitive L-type calcium channel alpha-1 subunit|voltage-gated calcium channel subunit alpha Cav1.1	1q32.1
DAO	D-amino acid oxidase	D-amino-acid oxidase	12q24.11
FGF9	fibroblast growth factor 9	fibroblast growth factor 9|fibroblast growth factor 9 (glia-activating factor)|heparin-binding growth factor 9	13q12.11
TBXAS1	thromboxane A synthase 1	thromboxane-A synthase|TXA synthase|cytochrome P450 5A1|cytochrome P450, family 5, subfamily A, polypeptide 1|platelet, cytochrome P450, subfamily V|thromboxane A synthase 1 (platelet)	7q34
MATR3	matrin 3	matrin-3|vocal cord and pharyngeal weakness with distal myopathy	5q31.2
NEFH	neurofilament heavy	neurofilament heavy polypeptide|200 kDa neurofilament protein|NF-H|neurofilament triplet H protein|neurofilament, heavy polypeptide 200kDa	22q12.2
ADARB1	adenosine deaminase, RNA specific B1	double-stranded RNA-specific editase 1|RNA editing deaminase 1|RNA-editing enzyme 1|adenosine deaminase, RNA-specific, B1 (homolog of rat RED1)|dsRNA adenosine deaminase DRADA2	21q22.3
CHGB	chromogranin B	secretogranin-1|cgB|secretogranin B|secretogranin I|sgI	20p12.3
SETX	senataxin	probable helicase senataxin|SEN1 homolog|amyotrophic lateral sclerosis 4 protein	9q34.13
KCNA2	potassium voltage-gated channel subfamily A member 2	potassium voltage-gated channel subfamily A member 2|potassium channel, voltage gated shaker related subfamily A, member 2|potassium voltage-gated channel, shaker-related subfamily, member 2|voltage-gated K(+) channel HuKIV|voltage-gated potassium channel HBK5|voltage-gated potassium channel protein Kv1.2|voltage-gated potassium channel subunit Kv1.2	1p13.3
UBQLN2	ubiquilin 2	ubiquilin-2|Nedd4 binding protein 4|protein linking IAP with cytoskeleton	Xp11.21

		2|ubiquitin-like product Chap1/Dsk2	
KIF1B	kinesin family member 1B	kinesin-like protein KIF1B|kinesin superfamily protein KIF1B	1p36.22
PON3	paraoxonase 3	serum paraoxonase/lactonase 3|arylesterase 3|paraoxanase-3	7q21.3
DACH1	dachshund family transcription factor 1	dachshund homolog 1|dac homolog	13q21.33
CRBN	cereblon	protein cereblon|protein x 0001	3p26.2
KIF5A	kinesin family member 5A	kinesin heavy chain isoform 5A|KIF5A variant protein|kinesin heavy chain neuron-specific 1|kinesin, heavy chain, neuron-specific|neuronal kinesin heavy chain	12q13.3
VEGFB	vascular endothelial growth factor B	vascular endothelial growth factor B|VEGF-related factor	11q13.1
MIR338	microRNA 338		17q25.3
CHMP2B	charged multivesicular body protein 2B	charged multivesicular body protein 2b|VPS2 homolog B|chromatin modifying protein 2B|vacuolar protein-sorting-associated protein 2-2	3p11.2
PTPRF	protein tyrosine phosphatase, receptor type F	receptor-type tyrosine-protein phosphatase F|LCA-homolog|leukocyte antigen-related (LAR) PTP receptor|leukocyte antigen-related tyrosine phosphatase|leukocyte common antigen related|protein tyrosine phosphatase, receptor type, F polypeptide|receptor-linked protein-tyrosine phosphatase LAR	1p34.2
GZMA	granzyme A	granzyme A|CTL tryptase|Cytotoxic T-lymphocyte-associated serine esterase-3|Granzyme A (Cytotoxic T-lymphocyte-associated serine esterase-3; Hanukah factor serine protease)|HF|Hanukah factor serine protease)|cytotoxic T-lymphocyte proteinase 1|fragmentin-1|granzyme 1|granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)|h factor	5q11.2
ALS2	ALS2, alsin Rho guanine nucleotide exchange factor	alsin|amyotrophic lateral sclerosis 2 (juvenile)|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 6 protein|amyotrophic lateral sclerosis 2 protein	2q33.1
CTNND2	catenin delta 2	catenin delta-2|T-cell delta-catenin|catenin (cadherin-associated protein), delta 2 (neural plakophilin-related arm-repeat protein)|neurojungin	5p15.2
GRIA3	glutamate ionotropic receptor AMPA type subunit 3	glutamate receptor 3|AMPA-selective glutamate receptor 3|dJ1171F9.1|gluR-3|glutamate receptor C|glutamate receptor subunit 3|glutamate receptor, ionotrophic, AMPA 3|glutamate receptor, ionotropic, AMPA 3	Xq25
PCSK5	proprotein convertase subtilisin/kexin type 5	proprotein convertase subtilisin/kexin type 5|prohormone convertase 5|proprotein convertase 6|protease PC6|subtilase|subtilisin/kexin-like protease PC5	9q21.13
MAP2K5	mitogen-activated protein kinase kinase 5	dual specificity mitogen-activated protein kinase kinase 5|MAP kinase kinase 5|MAP kinase kinase MEK5b|MAPK/ERK kinase 5|MAPKK 5|MEK 5	15q23
TAF15	TATA-box binding protein associated factor 15	TATA-binding protein-associated factor 2N|RBP56/CSMF fusion|TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa|TATA box binding protein (TBP)-associated factor, RNA polymerase II, N, 68kD (RNA-binding protein 56)|TATA box-binding protein-associated factor 2N (RNA-binding protein 56)|TBP-associated factor 15	17q12
PFKP	phosphofructokinase, platelet	ATP-dependent 6-phosphofructokinase, platelet type|6-phosphofructokinase type C|6-phosphofructokinase, platelet type|phosphofructo-1-kinase isozyme C|phosphofructokinase 1|phosphohexokinase	10p15.2

ITGA9	integrin subunit alpha 9	integrin alpha-9|integrin alpha-RLC	3p22.2
ATP2B2	ATPase plasma membrane Ca2+ transporting 2	plasma membrane calcium-transporting ATPase 2|ATPase, Ca++ transporting, plasma membrane 2|plasma membrane Ca(2+)-ATPase|plasma membrane Ca2+ pump 2|plasma membrane calcium ATPase|plasma membrane calcium pump	3p25.3
SOX5	SRY-box 5	transcription factor SOX-5|SRY (sex determining region Y)-box 5	12p12.1
HOXD10	homeobox D10	homeobox protein Hox-D10|homeo box 4D|homeo box D10|homeobox protein Hox-4D|homeobox protein Hox-4E	2q31.1
SLC25A12	solute carrier family 25 member 12	calcium-binding mitochondrial carrier protein Aralar1|araceli hiperlarga|calcium binding mitochondrial carrier superfamily member Aralar1|mitochondrial aspartate glutamate carrier 1|solute carrier family 25 (aspartate/glutamate carrier), member 12|solute carrier family 25 (mitochondrial carrier, Aralar), member 12	2q31.1
DPP6	dipeptidyl peptidase like 6	dipeptidyl aminopeptidase-like protein 6|DPP VI|dipeptidyl aminopeptidase IV-related protein|dipeptidyl peptidase IV-related protein|dipeptidyl peptidase VI|dipeptidyl-peptidase 6	7q36.2
CNTN4	contactin 4	contactin-4|axonal-associated cell adhesion molecule|brain-derived immunoglobulin superfamily protein 2|neural cell adhesion protein BIG-2	3p26.3-p26.2
ITPR2	inositol 1,4,5-trisphosphate receptor type 2	inositol 1,4,5-trisphosphate receptor type 2|IP3 receptor|IP3R 2|cilia and flagella associated protein 48|type 2 InsP3 receptor	12p11.23
HADH	hydroxyacyl-CoA dehydrogenase	hydroxyacyl-coenzyme A dehydrogenase, mitochondrial|L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain|medium and short-chain L-3-hydroxyacyl-coenzyme A dehydrogenase|short-chain 3-hydroxyacyl-CoA dehydrogenase|testis secretory sperm-binding protein Li 203a	4q25
PRPH	peripherin	peripherin|neurofilament 4 (57kD)	12q13.12
RGS6	regulator of G protein signaling 6	regulator of G-protein signaling 6|regulator of G-protein signalling 6	14q24.2
NFASC	neurofascin	neurofascin|neurofascin homolog|neurofascin isoform 140	1q32.1
SCG2	secretogranin II	secretogranin-2|chromogranin-C|secretoneurin	2q36.1
CCS	copper chaperone for superoxide dismutase	copper chaperone for superoxide dismutase|superoxide dismutase copper chaperone	11q13.2
NRXN3	neurexin 3	neurexin 3|neurexin III|neurexin-3-alpha	14q24.3-q31.1
ITIH4	inter-alpha-trypsin inhibitor heavy chain family member 4	inter-alpha-trypsin inhibitor heavy chain H4|ITI heavy chain H4|inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein)|inter-alpha-inhibitor heavy chain 4|inter-alpha-trypsin inhibitor family heavy chain-related protein|inter-alpha-trypsin inhibitor, heavy chain-like, 1|plasma kallikrein-sensitive glycoprotein 120	3p21.1
OPCML	opioid binding protein/cell adhesion molecule like	opioid-binding protein/cell adhesion molecule|IgLON family member 1|opioid binding protein/cell adhesion molecule-like preprotein	11q25
SLITRK1	SLIT and NTRK like family member 1	SLIT and NTRK-like protein 1|leucine-rich repeat-containing protein 12|slit and trk like gene 1	13q31.1
CHCHD10	coiled-coil-helix-coiled-coil-helix domain containing 10	coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial|MIX17 homolog A	22q11.23
LDHC	lactate dehydrogenase C	L-lactate dehydrogenase C chain|LDH testis subunit|LDH-C|LDH-X|cancer/testis	11p15.1

		antigen 32|lactate dehydrogenase C4|lactate dehydrogenase c variant 1|lactate dehydrogenase c variant 3|lactate dehydrogenase c variant 4|lactate dehydrogenase c variant 5|testis secretory sperm-binding protein Li 234P	
ANKRD1	ankyrin repeat domain 1	ankyrin repeat domain-containing protein 1|ankyrin repeat domain 1 (cardiac muscle)|cardiac ankyrin repeat protein|cytokine-inducible gene C-193 protein|cytokine-inducible nuclear protein|epididymis secretory sperm binding protein|liver ankyrin repeat domain 1	10q23.31
NEK1	NIMA related kinase 1	serine/threonine-protein kinase Nek1|NIMA (never in mitosis gene a)-related kinase 1|never in mitosis A-related kinase 1|nimA-related protein kinase 1|protein-serine/threonine kinase|renal carcinoma antigen NY-REN-55	4q33
ASIC2	acid sensing ion channel subunit 2	acid-sensing ion channel 2|acid sensing (proton gated) ion channel 2|amiloride-sensitive cation channel 1, neuronal|brain sodium channel 1|mammalian degenerin homolog|neuronal amiloride-sensitive cation channel 1	17q11.2-q12
C19orf12	chromosome 19 open reading frame 12	protein C19orf12|membrane protein-associated neurodegeneration|neurodegeneration with brain iron accumulation 3	19q12
IFRD1	interferon related developmental regulator 1	interferon-related developmental regulator 1|12-O-tetradecanoylphorbol-13-acetate-induced sequence 7|TPA induced sequence 7|nerve growth factor-inducible protein PC4|pheochromocytoma cell-4	7q31.1
NIPA1	NIPA magnesium transporter 1	magnesium transporter NIPA1|non imprinted in Prader-Willi/Angelman syndrome 1|non-imprinted in Prader-Willi/Angelman syndrome region protein 1|spastic paraplegia 6 protein	15q11.2
KCNIP1	potassium voltage-gated channel interacting protein 1	Kv channel-interacting protein 1|A-type potassium channel modulatory protein 1|Kv channel interacting protein 1|potassium channel interacting protein 1|vesicle APC-binding protein	5q35.1
FIG4	FIG4 phosphoinositide 5-phosphatase	polyphosphoinositide phosphatase|FIG4 homolog, SAC domain containing lipid phosphatase|FIG4 homolog, SAC1 domain containing lipid phosphatase|FIG4 homolog, SAC1 lipid phosphatase domain containing|SAC domain-containing protein 3|Sac domain-containing inositol phosphatase 3|phosphatidylinositol 3,5-bisphosphate 5-phosphatase	6q21
TTBK2	tau tubulin kinase 2	tau-tubulin kinase 2	15q15.2
GLE1	GLE1, RNA export mediator	nucleoporin GLE1|GLE1 RNA export mediator homolog|GLE1 RNA export mediator-like|GLE1-like protein|GLE1-like, RNA export mediator	9q34.11
CABIN1	calcineurin binding protein 1	calcineurin-binding protein cabin-1|calcineurin binding protein cabin 1|calcineurin inhibitor	22q11.23
ANKS1B	ankyrin repeat and sterile alpha motif domain containing 1B	ankyrin repeat and sterile alpha motif domain-containing protein 1B|E2a-Pbx1-associated protein|amyloid-beta precursor protein intracellular domain associated protein 1|cajalin 2	12q23.1
ST3GAL3	ST3 beta-galactoside alpha-2,3-sialyltransferase 3	CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase|Gal beta-1,3(4)GlcNAc alpha-2,3 sialyltransferase|alpha 2,3-ST 3|alpha 2,3-sialyltransferase III|alpha-2,3-sialyltransferase II|sialyltransferase 6 (N-acetyllacosaminide alpha 2,3-sialyltransferase)|sialyltransferase 6 (N-acetyllactosaminide alpha 2,3-sialyltransferase)	1p34.1
RAPH1	Ras association (RalGDS/AF-	ras-associated and pleckstrin homology domains-containing protein	2q33.2

	6) and pleckstrin homology domains 1	1|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 18|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 9|lamellipodin|proline-rich EVH1 ligand 2	
ANKRD2	ankyrin repeat domain 2	ankyrin repeat domain-containing protein 2|ankyrin repeat domain 2 (stretch responsive muscle)|skeletal muscle ankyrin repeat protein	10q24.2
KIFAP3	kinesin associated protein 3	kinesin-associated protein 3|small G protein GDP dissociation stimulator|smg GDS-associated protein	1q24.2
ZFYVE26	zinc finger FYVE-type containing 26	zinc finger FYVE domain-containing protein 26|FYVE domain-containing centrosomal protein|spastizin|zinc finger, FYVE domain containing 26	14q24.1
MACROD2	MACRO domain containing 2	O-acetyl-ADP-ribose deacetylase MACROD2|MACRO domain-containing protein 2|[Protein ADP-ribosylglutamate] hydrolase	20p12.1
UNC13A	unc-13 homolog A	protein unc-13 homolog A	19p13.11
PDGFRL	platelet derived growth factor receptor like	platelet-derived growth factor receptor-like protein|PDGF receptor beta-like tumor suppressor|PDGFR-like protein|platelet-derived growth factor-beta-like tumor suppressor	8p22
SS18L1	SS18L1, nBAF chromatin remodeling complex subunit	calcium-responsive transactivator|SS18-like protein 1|SYT homolog 1|synovial sarcoma translocation gene on chromosome 18-like 1	20q13.33
CFAP410	cilia and flagella associated protein 410	cilia- and flagella-associated protein 410|nuclear encoded mitochondrial protein C21orf2|C21orf-HUMF09G8.5|leucine rich repeat containing 76|leucine-rich repeat-containing protein 76|protein C21orf2	21q22.3
ARHGEF28	Rho guanine nucleotide exchange factor 28	rho guanine nucleotide exchange factor 28|190 kDa guanine nucleotide exchange factor|Rho guanine nucleotide exchange factor (GEF) 28|Rho interacting protein 2|p190-RhoGEF|rho-guanine nucleotide exchange factor	5q13.2
SCN7A	sodium voltage-gated channel alpha subunit 7	sodium channel protein type 7 subunit alpha|putative voltage-gated sodium channel subunit alpha Nax|sodium channel protein cardiac and skeletal muscle subunit alpha|sodium channel protein type VII subunit alpha|sodium channel, voltage-gated, type VI, alpha polypeptide|sodium channel, voltage-gated, type VII, alpha subunit|voltage-dependent sodium channel alpha subunit	2q24.3
FGGY	FGGY carbohydrate kinase domain containing	FGGY carbohydrate kinase domain-containing protein|D-ribulokinase	1p32.1
STK36	serine/threonine kinase 36	serine/threonine-protein kinase 36|fused homolog|testicular tissue protein Li 188	2q35
KIF5C	kinesin family member 5C	kinesin heavy chain isoform 5C|kinesin heavy chain|kinesin heavy chain neuron-specific 2	2q23.1-q23.2
TTBK1	tau tubulin kinase 1	tau-tubulin kinase 1|brain-derived tau kinase	6p21.1
IFT74	intraflagellar transport 74	intraflagellar transport protein 74 homolog|capillary morphogenesis gene 1 protein|capillary morphogenesis protein 1|coiled-coil domain containing 2|coiled-coil domain-containing protein 2|intraflagellar transport 74 homolog	9p21.2
GLRX5	glutaredoxin 5	glutaredoxin-related protein 5, mitochondrial|epididymis secretory sperm binding protein|glutaredoxin 5 homolog|monothiol glutaredoxin-5	14q32.13
VPS54	VPS54, GARP complex subunit	vacuolar protein sorting-associated protein 54|hepatocellular carcinoma protein 8|protein phosphatase 1, regulatory subunit 164|tumor antigen HOM-HCC-8|tumor antigen SLP-8p|vacuolar protein sorting 54 homolog	2p15-p14
PIGL	phosphatidylinositol glycan	N-acetylglucosaminyl-phosphatidylinositol de-N-acetylase|N-	17p11.2

	anchor biosynthesis class L	acetylglucosaminylphosphatidylinositol deacetylase|PIG-L|phosphatidylinositol glycan, class L|phosphatidylinositol-glycan biosynthesis class L protein	
GSE1	Gse1 coiled-coil protein	genetic suppressor element 1|CTC-786C10.1|Gse1 coiled-coil protein homolog	16q24.1
CHODL	chondrolectin	chondrolectin|transmembrane protein MT75	21q21.1
STRADB	STE20 related adaptor beta	STE20-related kinase adapter protein beta|ILP-interacting protein ILPIPA|STE20-related kinase adaptor beta|STRAD beta|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 2 protein|pseudokinase ALS2CR2	2q33.1
TMEM237	transmembrane protein 237	transmembrane protein 237|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 4|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 4 protein	2q33.1
TMEM132E	transmembrane protein 132E	transmembrane protein 132E	17q12
MPP4	membrane palmitoylated protein 4	MAGUK p55 subfamily member 4|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 5 protein|discs large homolog 6|membrane protein, palmitoylated 4 (MAGUK p55 subfamily member 4)	2q33.1
STON1	stonin 1	stonin-1|stoned B homolog 1|stoned B-like factor|stoned-b1	2p16.3
FAT3	FAT atypical cadherin 3	protocadherin Fat 3|FAT tumor suppressor homolog 3|cadherin family member 15|cadherin-related family member 10	11q14.3
ZNF142	zinc finger protein 142	zinc finger protein 142|zinc finger protein 142 (clone pHZ-49)	2q35
STON1-GTF2A1L	STON1-GTF2A1L readthrough	STON1-GTF2A1L protein|General transcription factor II A, 1-like factor|TFIIA-alpha and beta-like factor|stoned B/TFIIA-alpha/beta-like factor	2p16.3
C2CD6	C2 calcium dependent domain containing 6	C2 calcium-dependent domain-containing protein 6|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 11|amyotrophic lateral sclerosis 2 chromosomal region candidate gene 11 protein|amyotrophic lateral sclerosis 2 chromosome region candidate 11|testicular tissue protein Li 16	2q33.1
LIPC-AS1	LIPC antisense RNA 1		15q21.3
CFLAR-AS1	CFLAR antisense RNA 1	CFLAR antisense RNA 1 (non-protein coding)|amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 10	2q33.1
SNORD114-31	small nucleolar RNA, C/D box 114-31	14q32.31	

Annexes du

CHAPITRE III

Annexe 1.

Résultats CoDEseq SPARK

095079

SYNTHESE CLINIQUE

Maladie de Parkinson familiale, compatible avec un mode de transmission dominant : Le père atteint avait été diagnostiqué probablement vers 70ans. Après 10 ans d'évolution, la maladie s'était compliquée de dyskinésies, d'hallucinations, de pollakiurie, d'hypersalivation et une perte de la sensibilité des extrémités. Chez le patient 095079, la maladie a débutée vers l'âge de 57ans. Le patient présentait une akinésie et une rigidité dyssymétriques, sans tremblement. La réponse était faiblement dopasensible. A 59 ans l'IRM cérébrale sans modification significative. A 60 ans, le bilan neuropsychologique est toutefois en faveur d'un léger déficit fronto-sous-cortical. La patient présente par ailleurs, une hypertrophie ventriculaire gauche.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 4/5	CLASS 3
ABCA7 rs77403558, ch19, heterozygote **Stop gained** NM_019112.3, exon 9, c.853A>T p.Arg285* Susceptibility to Alzheimer disease 9 No prediction data, 6.14 .10^{-5} in GnomAD AF	*DNAJC13* rs79953286, ch3, Missense heterozygote NM_001329126.1, exon44, c.5033A>G, p.Tyr1678Cys Autosomal dominant PARKINSON disease 21 prédit DELETERE >3 logiciels, 0.04 in Gnomad AF
	MAPT rs63750072, ch17, Missense heterozygote NM_001123066.3, exon 6, c.689A>G, p.Gln230Arg Autosomal dominant Dementia, frontotemporal, with or without Parkinsonism OR Isolated cases, Autosomal dominant Pick disease OR **Autosomal dominant Supranuclear palsy, progressive** OR Autosomal recessive Supranuclear palsy, progressive atypical OR **Multifactorial, Isolated cases, susceptibility to Parkinson disease** prédit DELETERE >3 logiciels

Copy Number Variants, CNVs

(Annotation variants cnv all methods)

Ch17 pos 45989856-45990096 del 17q21.1 240pb exons 7 et 8 de MAPT

MAPT: Dementia, frontotemporal, with or without parkinsonism (#600274), Autosomal dominant; **susceptibility to Parkinson disease** (#168600), **Autosomal dominant, Multifactorial**; Pick disease, (#172700), Autosomal dominant; Supranuclear palsy, progressive, (#601104), Autosomal dominant; Supranuclear palsy, progressive atypical, (#260540), Autosomal recessive.

INTERPRETATION

Le phénotype observé chez le patient et l'histoire familiale, pourraient être en rapport avec la conjonction de 2 variants pathogènes :

- ✓ Le variant dans le gène *ABCA7* est très rare et provoque un arrêt prématuré de la transcription du gène. *ABCA7* (ATP Binding Cassette Subfamily A Member 7) code pour une protéine de la superfamille des ATP-binding cassette (ABC) transporteurs, impliquées dans le transport de diverses molécules à travers les membranes intra- et extracellulaires. Les mutations de ce gène sont associées à la maladie d'Alzheimer autosomique dominante et précoce. Mais Karen Nuytemans et al. ont montré que les variants fonctionnels dans *ABCA7* sont responsables de chevauchement entre la maladie de Parkinson et la maladie d'Alzheimer. [1]. La mutation stop retrouvée dans ce gène est très probablement de fort impact fonctionnel et possiblement en lien avec le phénotype observé chez le patient.
- ✓ La délétion partielle (exons 7 et 8) du gène *MAPT* est une délétion de 240pb très probablement impliquée dans le phénotype. En effet, des délétions plus larges, des exons 6 à 9, ont été décrites comme étant un facteur de susceptibilité à la maladie de Parkinson de transmission dominante ou multifactorielle[2]. En plus, elle est associée à un variant de signification incertaine de l'exon 6 de ce même gène, qui pourrait être un facteur de susceptibilité de la maladie de Parkinson, comme l'a montré, l'étude d'une grande famille du sud de la République tchèque [3].

L'enquête familiale permettrait de vérifier si la ségrégation des variants mis en évidence confirme le caractère délétère de leur occurrence conjointe.

REFERENCES

[1] K. Nuytemans *et al.*, « Overlap between Parkinson disease and Alzheimer disease in *ABCA7* functional variants », *Neurol. Genet.*, vol. 2, n° 1, p. e44, févr. 2016, doi: 10.1212/NXG.0000000000000044.

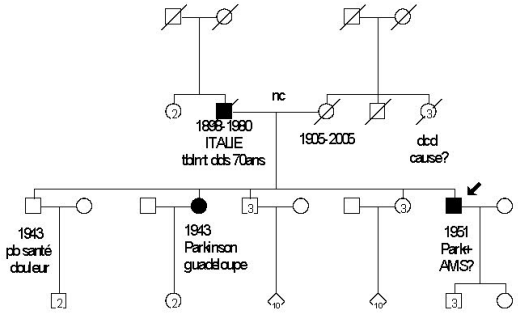
[2] A. Rovelet-Lecrux *et al.*, « Partial deletion of the *MAPT* gene: A novel mechanism of FTDP-17 », *Hum. Mutat.*, vol. 30, n° 4, p. E591‑E602, avr. 2009, doi: 10.1002/humu.20979.

[3] T. Bartoníková *et al.*, « New endemic familial parkinsonism in south Moravia, Czech Republic and its genetical background »:, *Medicine (Baltimore)*, vol. 97, n° 38, p. e12313, sept. 2018, doi: 10.1097/MD.0000000000012313.

15S310

SYNTHESE CLINIQUE

Trouble de la marche type Parkinson et modification du comportement. Age de début 62ans.
Hallucinations visuelles, trouble mnésique, troubles sommeil depuis 60-62ans.
Tremblement de repos côté droit. Hypertonie extrapyramidale droite.
Tableau anxio-dépressif bien marqué. IRM cérébrale normale.
Une sœur atteinte avec un âge de début 45ans. Transmission récessive/dominante avec anticipation.
Origine Italienne par le père décédé à 82ans avait des tremblements apparus à l'âge de 70ans.



RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
	GIGYF2 rs527464858, ch2 , hétérozygote disruptive inframe deletion NM_001103146.1, exon29, c.3626_3646delTGCCACAGCAGCAGCAGCAGC , p.Leu1209_Gln1215del, Parkinson 11 Mode de transmission indéfini absent dans gnomAD et pas de prédiction in sillico

CLASS 2
VPS13C
rs34060567, ch15, heterozygote missense variant
NM_020821.2, exon 61, c.8423A>G p.Lys2808Arg,
Parkinson disease 23, autosomal recessive, early onset
GnomAD AF 0,00979282

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

Pas de CNVs associés à la pathologie

INTERPRETATION

L'analyse n'a pas permis de mettre en évidence de variants clairement pathogène susceptibles d'être impliqués dans la pathologie.
Un variant de signification incertaine de nature à interrompre le cadre de lecture d'un motif répété dans l'exon 29 du gène GIGYF2 est cependant mis en évidence.
Précédemment, des études d'association avaient permis de montrer un lien entre les mutations de GIGYF2 et la maladie de Parkinson de type 11 dans les populations française et Italiennes [1]
Une étude menée par des chinois, montre que les indels de ce gène sont impliqués dans la maladie de Parkinson d'une population issue du nord de la Chine [2]

L'analyse des variants de type expansion de motifs répétés par l'outil GangSTR, confirme l'insertion de plusieurs motifs dont les répétitions de CAG. Une étude de la sœur serait également contributive.

REFERENCES

[1] C. Lautier *et al.*, « Mutations in the GIGYF2 (TNRC15) Gene at the PARK11 Locus in Familial Parkinson Disease », *Am. J. Hum. Genet.*, vol. 82, n° 4, p. 822‑833, avr. 2008, doi: 10.1016/j.ajhg.2008.01.015.

[2] Z. Huo *et al.*, « Genetic analysis of indel markers in three loci associated with Parkinson's disease », *PLOS ONE*, vol. 12, n° 9, p. e0184269, sept. 2017, doi: 10.1371/journal.pone.0184269.

17S218

SYNTHESE CLINIQUE

Symptôme inaugural, dépression à l'âge de 77ans. Puis troubles cognitifs (désorientation), troubles praxiques, hypersalivation, faciès figé, hypoTA orthostatique, dysarthrie, extinction de voix, dysphagie, trouble de la posture et de l'équilibre.
Limitation verticale du mouvement des yeux, saccades rapides altérées.
Important syndrome extrapyramidal : Akinésie faciale majeure. Bavage. Marche rapide et ample mais sans balancement des bras qui restent semi fléchis. Attitude voutée. Rigidité axiale et segmentaire majeure. Tremblement d'attitude.
Trouble du comportement avec agitation et akathisie.
Pas de dopa-sensibilité ni d'antécédents familiaux.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
LRRK2 rs35303786; 39200, Ch12, **hétérozygote missense** NM_198578.3, exon 34, c.4937T>C, p.Met1646Thr Autosomal **dominant** Parkinson disease 8 Conflicting_interpretations_of_pathogenicity, GnomAD 0,0092	
	SORL1 rs ?, Ch11, **hétérozygote missense** NM_003105.5, exon 34, c.4696G>C, p.Asp1566His Autosomal **dominant**, pathogenesis association with Alzheimer disease prédit DELETERE >3 logiciels, GnomAD AF 2,53E-05
	HLA-DRB1 rs1059582, ch6, **hétérozygote missense** NM_002124.3.2, exon 2, c.239C>G, p.Thr80Arg **Multifactorial susceptibility** to Multiple sclerosis prédit DELETERE >3 logiciels, GnomAD AF 0.00928409

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

Pas de CNVs en rapport avec la pathologie

INTERPRETATION

L'analyse révèle un variant de pathogénicité discutée dans le gène LRRK2 qui est impliqué dans la maladie de Parkinson 8 de transmission dominante.[1]
L'absence d'histoire familiale plaide en faveur de variant de novo qu'il conviendrait de pouvoir vérifier par une étude de ségrégation en testant les parents ou apparentés non atteints.

On trouve aussi 2 autres variants candidats dans les gènes :

- HLA-DRB1, facteur de susceptibilité multifactoriel à l'atrophie multisystématisée.
- SORL1 impliqué dans la maladie d'Alzheimer de transmission dominante

Une analyse en trio aurait été plus contributive.

REFERENCES

[1] O. A. Ross *et al.*, « Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case–control study », *Lancet Neurol.*, vol. 10, n° 10, p. 898‑908, oct. 2011, doi: 10.1016/S1474-4422(11)70175-2.

15S132

SYNTHESE CLINIQUE

A 79ans : Incontinence urinaire. Instabilité à la marche. Instabilité avec rétropulsion. Hypoacousie gauche. Probable épilepsie à crises partielles complexes. L'IRM cérébrale montre une atrophie diffuse à prédominance temporale.
Diabète non-insulino dépendant.
Bilan neuropsy à 80ans : Dysfonctionnement fronto-sous cortical, MMS 19/30
A 85ans : agitation motrice et mouvement involontaire.
La patiente est retraitée de la fonction d'ouvrière agricole. Les antécédents familiaux rapportés font état d'antécédents de probable déficience intellectuelle avec épilepsie.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
LRRK2 rs72546327, chr12, heterozygote missense NM_198578.3, exon 30, c.4229C>T, p.Thr1410Met Autosomal dominant, Parkinson disease 8 prédit délétère >3 softs, (GnomAD AF 0,0020)	TRPM7 chr15, ▪ rs148771191;rs750996657 , hétérozygote disruptive_inframe_deletion NM_017672.5, exon 29, c.4500_4502delTAG, p.Ser1500del (GnomAD AF 0,0071) ▪ rs79443717, heterozygote splice_region_variant&intron_variant NM_017672.5, exon9, c.1131+3A>G, p. ? (GnomAD 0,008) Autosomal dominant susceptibility to Amyotrophic lateral sclerosis-parkinsonism/dementia complex
ANKK1 rs113005509, ch11, heterozygote **stop_gained** NM_178510.1, exon 8, c.1759G>T, p.Glu587* Mode de transmission ?, reduced brain density of dopamine receptor D2 prédit délétère >3 softs , GnomAD AF 0.00197843	

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

Ch X 147912159_155279367_DUP Xq28 7Mb loss of function 5'UTR-3'UTR mental retardation X-linked type 72 (mrx72) +/- parkinsons
Contenant notamment le gène RAB39B dont les mutations sont associée à DI, épilepsie et parkinsonisme de révélation variable.

INTERPRETATION

L'analyse met en évidence un variant pathogène dans le gène ANKK1.
Ce gène code pour une Ser/Thr kinase étroitement liée au récepteur D2 de la dopamine (DRD2).
Les mutations DRD2 ont été beaucoup étudiées dans les comportements addictifs notamment à l'alcool mais aussi la schizophrénie.
Aujourd'hui, leur implication dans les phénotypes tels que dystonie myoclonique et parkinsonisme sont discutés. [1]
L'arrêt prématuré de la transcription, induite par l'apparition d'un codon STOP est très fortement en faveur de l'implication du variant dans la symptomatologie.

Par ailleurs, on trouve également un variant de class 4 dans le gène LRRK2 impliqué dans la maladie de Parkinson 8 dominante. [2]
Des variants de signification incertaines dans le gène TRPM7 sont susceptibles d'interagir dans la symptomatologie.
Une étude de ségrégation familiale permettrait de confirmer l'implication de ce variant dans le phénotype.

Un variant de nombre de copies du chromosome 7 associé à la déficience intellectuelle et des syndromes parkinsoniens tardifs a été mis en évidence. Cette duplication du chromosome X est également susceptible de participer au phénotype. [3]
L'enquête familiale à la recherche d'épilepsie et DI serait contributive à la filtration de ce variant.

REFERENCES

[1] R. L. Oliveri *et al.*, « Dopamine D2 receptor gene polymorphism and the risk of levodopa-induced dyskinesias in PD », *Neurology*, vol. 53, n° 7, p. 1425‑1425, oct. 1999, doi: 10.1212/WNL.53.7.1425.
[2] F. S. Refai, S. H. Ng, et E.-K. Tan, « Evaluating *LRRK2* Genetic Variants with Unclear Pathogenicity », *BioMed Res. Int.*, vol. 2015, p. 1‑6, 2015, doi: 10.1155/2015/678701.
[3] G. R. Wilson *et al.*, « Mutations in RAB39B Cause X-Linked Intellectual Disability and Early-Onset Parkinson Disease with α-Synuclein Pathology », *Am. J. Hum. Genet.*, vol. 95, n° 6, p. 729‑735, déc. 2014, doi: 10.1016/j.ajhg.2014.10.015.

17S116

SYNTHESE CLINIQUE

Ataxie et dysarthrie apparues à l'âge de 48ans.
Spasmes pharyngés et troubles de la déglutition.
A 58ans : IRM cérébrale montre une atrophie cérébelleuse et médullaire et atrophie corticale ; ENMG polyneuropathie axonale sensitivo motrice aux 4 membres.
Syndrome dépressif et trouble de l'humeur.
A 67ans : le tableau clinique associe : ataxie cérébelleuse, neuropathie axonale sensitivo-motrice chronique, fasciculations, amyotrophie des extrémités et syndrome frontal.
A 70ans : biopsie musculaire montre des ragged red fibers et des amas mitochondriaux sous sarcolémiques en faveur d'une cytopathie mitochondriale. Sans déficit de la chaîne respiratoire. Amplification SCA 1 de 35 CAG (méthode de référence, PCR)
Pas de délétion de l'ADNmt. Les mutations du gène FRDA ainsi que les délétions de l'ADNmt ont été exlues.
L'histoire familiale révèle que deux cousins germains sont atteints de pathologies neurodégénératives : l'un atteint de SLA et l'autre, de neuropathie amyloïde.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
RARS2 rs377591456, ch 6, heterozygote **missense variant** NM_020320.4, exon 6, c.422A>G, p.His141Arg Autosomal recessive, Pontocerebellar hypoplasia, type 6 prédit DELETERE >3 logiciels, GnomAD AF 1,81E-05	
NDUFAF1 rs ?, ch15, heterozygote **missense variant** NM_016013.3, exon 3, c.623T>C, p.Leu208Pro Autosomal recessive/Mitochondrial/X-linked dominant, Mitochondrial complex I deficiency prédit DELETERE >3 logiciels, GnomAD AF 1,44E-05	
ANO10 rs61732728;345187, ch3, heterozygote **missense variant** NM_018075.3, exon 6, c.1133G>A, p.Arg378Gln Autosomal recessive, Spinocerebellar ataxia 10 prédit délétère >3 softs, GnomAD AF 8,74E-04	ANO10 rs763039129, ch3, heterozygote **splice_region_variant&intron_variant** NM_018075.3, exon 7, c.1219-6delT Autosomal recessive, Spinocerebellar ataxia 10 no prediction data, GnomAD AF ?

Copy Number Variants, CNVs

(AnnotSV-merge_cnv_all_methods)

Ch19 pos 40875670-40875901 dup 19q13.2 231pb intron6-exon7 de PLD3
PLD3 : Spinocerebellar ataxia 46, **Autosomal dominant**

INTERPRETATION

Des variants délétères associées à l'ataxie spinocérébelleuse ont été mis en évidence. Mais aucun n'est suffisant à lui seul pour expliquer la maladie :

- La mutation dans RARS2, répond à un mode de transmission récessif et le phénotype qu'il produit est plus sévère.
- La mutation du gène NDUFAF1 serait plus en lien mais elle cause une atteinte du complexe 1 de la chaîne respiratoire ; or les analyses de celle-ci sont données pour normales chez ce patient.

Seuls les variants du gène ANO10, responsable des formes récessives d'ataxie avec déficit en co-enzyme Q10 et fasciculations pourraient être mis en cause dans le phénotype. A condition que le variant intronique (rs763039129, NM_018075.3, exon 7, c.1219-6delT) soit effectivement pathogène.[1]
Par ailleurs, on note également une petite duplication interstitielle du gène PLD3 associé à SCA 46. Il n'existe aucune description clinique en lien avec cette délétion interstitielle.
Pour aller plus loin dans l'interprétation des variants délétères chez ce patient, une analyse in vivo de l'effet du variant intronique de ANO10 serait intéressante.

Par ailleurs, conformément au résultat de la filière diagnostic, la méthode GangSTR confirme la présence d'expansion de motifs à la limite supérieure dans les gènes ATXN1 et 3, avec respectivement 35 et 32 répétitions CTG. Le screening des autres gènes de CA sujets aux expansions de motifs : ATXN2, ATXN7, CACNA1A, KLHL1AS, ATXN10, PPP2R2B, TBP, ATXN1L, ATXN2L, ATN1, NOP56 et BEAN, est non contributif. On relève seulement une amplification border-line de 25 CAG dans le gène TBP.

REFERENCES

[1] A. Balreira *et al.*, « ANO10 mutations cause ataxia and coenzyme Q10 deficiency », *J. Neurol.*, vol. 261, n° 11, p. 2192‑2198, nov. 2014, doi: 10.1007/s00415-014-7476-7.

Annexe 2.

Résultats CoDEseq CA

18S231 (mère) et 17S162 (fille)

SYNTHESE CLINIQUE

Les cas 18S231 et 17S162 sont mère et fille. Elles présentent toutes les 2, un syndrome cérébelleux statique apparu vers l'âge de 60 ans; prédominant aux membres inférieurs pour la mère. Il y aurait une petite fille atteinte, avec un âge de début 40ans.
Chez le cas 18S231, l'IRM cérébrale et la ponction lombaire avaient été données pour normales ; Les anticorps anti-GAD étaient négatifs. Sur le plan génétique, les SCA 1, 2, 3, 6, 7, et 17 avaient été exclus.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
	ATXN3 heterozygote conservative/disruptive_inframe_insertion GnomAD AF ?, **Machado-Joseph disease** Autosomal dominant ▪ rs12895357;rs763541221, NM_004993.5 , ch14 exon 10 ***18S231:*** c.915_916insCAGCAGCAGCAGCAGCAGCAGCAG p.Gln297_Gln305dup ***17S162:*** c.892_915dupCAGCAGCAGCAGCAGCAGCAG p.Gln298_Gln305dup ▪ rs ?, NM_001164774.1, ch14, exon 3 ***18S231:*** c.232_233insCAGCAGCAGCAGCAGCAGCAGCAG p.Ala77_Gly78insAlaAlaAlaAlaAlaAlaAlaAla ***17S162:*** c.209_232dupCAGCAGCAGCAGCAGCAGCAG p.Ala77_Gly78insAlaAlaAlaAlaAlaAlaAla ▪ rs ?, NM_001164778.1, ch14, Exon 6 ***18S231:*** c.430_431insCAGCAGCAGCAGCAGCAGCAGCAG p.Ala143_Gly144insAlaAlaAlaAlaAlaAlaAlaAla ***17S162:*** c.407_430dupCAGCAGCAGCAGCAGCAGCAG p.Ala143_Gly144insAlaAlaAlaAlaAlaAlaAla ***17S162*** SPTBN2 chr11, rs750934185, NM_006946.2 exon28, c.5939C>T, p.Ala1980Val Spinocerebellar ataxia 5 Autosomal dominant GnomAD AF 4,06E-06 prédit délétère >3 softs

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

18S231

ch19 pos 40844715-40875901	19q13.2duplication 31Kb contenant PLD3	spinocerebellar ataxia 46	autosomal dominant
chX pos147921899-155545312	Xq28 duplication 7,6Mb contenant ATP2B3	spinocerebellar ataxia liée à l'X	

INTERPRETATION

Aucun variant non ambigus n'a été mis en évidence par la méthode.
Cependant, on retrouve plusieurs insertions conservatives dans le gène ATXN3.
L'expansion de polyGln dans l'exon 10 ne semble pas pathogène mais les multiples insertions semblent concourir finalement à la génération de tracts polyAla. Or Gaspar C et al. ont prouvé que leur présence est toxique[1]
Par la suite, cette tocicité est confirmée sur des modèles animaux et neurones de mamifères.[2]
Ces tracts pourraient être à l'origine du phénotype partagé par mère et fille.
La méthode GangSTR utilisée pour préciser les tailles d'expansion, permet de rapporter une expansion dans ATXN3 de 25 CTG chez le patient 17S162 contre seulement 9 chez le patient 18S231. Aucune autre expansion suspecte n'a été mise en évidence dans les gènes ATXN1, ATXN2, ATXN7, ATXN10, CACNA1A, KLHL1AS, PPP2R2B, TBP, ATXN1L, ATXN2L, ATN1, NOP56 ni BEAN explorés.
En parallèle, on rapporte d'une part, un CNV probablement bénin chez la mère (18S231).
D'autre part, un variant de classe 3 dans le gène SPTBN2, chez la fille.

REFERENCES

[1] C. Gaspar, « CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation », *Hum. Mol. Genet.*, vol. 9, n° 13, p. 1957‑1966, août 2000, doi: 10.1093/hmg/9.13.1957.

[2] S. J. Stochmanski *et al.*, « Expanded ATXN3 frameshifting events are toxic in Drosophila and mammalian neuron models », *Hum. Mol. Genet.*, vol. 21, n° 10, p. 2211‑2218, mai 2012, doi: 10.1093/hmg/dds036.

10S103

SYNTHESE CLINIQUE

Syndrome cérébelleux statique et cinétique relativement symétrique. Ataxie cérébelleuse isolée et sporadique.
Nystagmus horizontal prédominant à gauche.
Syndrome pyramidal. Les réflexes rotuliens abolis et les réflexes achilléens peu marqués.
Age de début 28ans.
Il existe des dysesthésies des mains et des pieds. Une importante fatigabilité à l'effort et des troubles de la vision épisodiques
Une crise convulsive en 1992, suite à l'accouchement de son 3e et dernier enfant.
Ptosis unilatéral intermittent.
Suivi pour le cœur ? troubles dysimmunitaires

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 4/5	CLASS 3
SYNE1 ch 6, heterozygote missense_variant ▪ rs180727534, NM_182961.3 exon 71 c.11675T>C, p.Leu3892Ser GnomAD AF 1,59E-04, prédit délétère >3 softs ▪ rs757562714, NM_182961.3 exon 49 c.7345C>T, p.Leu2449Phe GnomAD AF 4,06E-05, prédit délétère >3 softs Spinocerebellar ataxia, autosomal recessive (conflicting interpretation in ClinVar)	*SYNE1* rs116600265, NM_182961.3 exon 42 c.6135T>G, p.Ile2045Met GnomAD 0,00112193, prédit délétère >3 softs

Copy Number Variants, CNVs

(Annotation variants cnv all methods)

ch16 pos 89647651-89648282 Del 16q24.3 631Kb contenant l'intron 3-4 du gène *CHMP1A*
Associé à hypoplasie Pontocerebelleuse de type 8 (OMIM 614961) de transmission autosomal récessive, ou à un syndrome neurodéveloppemental sévère.

INTERPRETATION

Aucun variant de classe 5 n'est trouvé mais plusieurs variants de class 4 dans le gène SYNE1 pourraient être à l'origine du phénotype.
Ce qui est également cohérent avec le mode de transmission.
Les mutations de ce gène ont été mises en évidence dans une cohorte franco-canadienne où il représentait le 1e gène responsable d'ataxie spinocérébelleuse pure héréditaire et récessive. [1]
L'étude, par la même méthode, du fils non atteint, montre qu'il présente des variants bénins de SYNE1 mais a hérité du variant rs757562714 de sa mère.

Le CNV n'est probablement pas impliqué en propre dans le phénotype.

REFERENCES

[1] F. Gros-Louis *et al.*, « Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia », *Nat. Genet.*, vol. 39, n° 1, p. 80‑85, janv. 2007, doi: 10.1038/ng1927.

145121

SYNTHESE CLINIQUE

Probable atrophie multisystématisée caractérisée par l'évolution suivante :
Syndrome cérébelleux débuté à l'âge de 39ans, associé à des troubles psychiatriques (hallucinations).
L'IRM cérébrale montre une atrophie vermienne et cérébelleuse.
A 48ans : Syndrome cérébelleux stable, sans déficit sensitivomoteur, pas de myoclonies, pas de troubles cognitifs. Apparition d'un tremblement au repos et à la marche, d'une rigidité axiale et segmentaire et d'une limitation des mouvements oculaires en particulier vers le haut.
Antécédents de psychose et troubles psychologiques majeurs suivi pendant plus de 10ans. Le discours fait apparaitre des thèmes de persécutions plutôt inquiétants.
A 52ans: Syndrome pyramidal sans BBK, nystagmus persistant, dysarthrie cérébelleuse, incontinence urinaire et constipation, dysmétrie axiale. Score SARA à 29,5.

Mutations : SCA2, FMR1, NPC1 et NPC2 négatifs.

Antécédents familiaux : on suspecte la grand-mère maternelle d'avoir été atteinte.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 4/5	CLASS 3
	GIGYF2 rs34845648, ch2, Missense heterozygote NM_001103147.1, exon 14, c.1331C>T p.Pro444Leu GnomAD AF 0,0044 prédit délétère par >3 logiciels **Parkinson disease 11 susceptibility** transmission non précisée, défaut de pénétrance
	GIGYF2 rs766013709 ch2 disruptive frameshift variant NM_001103147.1, exon 26, c.3126_3128dupGCA, p.Gln1043dup GnomAD AF ? , pas de prédiction **Parkinson disease 11 susceptibility** transmission non précisée, défaut de pénétrance
	CCDC88C rs776229233, ch14, Heterozygote splice region variant NM_001080414.3, exon10,c.1050+6dupG p. ? **Spinocerebellar ataxia 40** Autosomal dominant (GnomAD AF 7,66.10^{-5}) Pas de données de prédiction in sillico

Copy Number Variants, CNVs

(Annotation variants cnv all methods)

Aucun variant de signification non ambiguë

INTERPRETATION

Aucun variant de signification non ambigu n'est trouvé dans l'analyse.
Les variants rapportés dans GIGYF2, sont de signification incertaine et n'expliquent pas à eux seuls le phénotype.
Le VUS dans CCDC88C serait à explorer. Des méthodes in sillico indépendantes de l'analyse (type HSF_Pro analysis) pourraient être mise en œuvre dans le but de prédire l'impact de ce variant d'épissage.
Le mutations faux sens de ce gène ont été rapportées dans des familles chinoises et polonaise [1]

REFERENCES

[1] M. Leńska-Mieciek *et al.*, « Familial ataxia, tremor, and dementia in a polish family with a novel mutation in the *CCDC88C* gene », *Mov. Disord.*, vol. 34, n° 1, p. 142‑144, janv. 2019, doi: 10.1002/mds.27536.

Annexe 3.

Résultats coDEseq SLA

17S353

SYNTHESE CLINIQUE

Troubles dysarthriques apparus vers l'âge de 68ans.
IRM cérébrale (71ans) : Normale. Leucopathie microvasculaire. ENMG (72ans): activité spontanée tonique discontinue
Examen clinique : Dysphonie, Dysarthrie parétique. Très probable syndrome de Meige.
Amyotrophie des premiers espaces interosseux et fasciculations de la langue.

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5/4	CLASS 3
	TRPM7 NM_017672.5, rs148771191; rs750996657 ch15, heterozygote disruptive_inframe_deletion exon 29, c.4500_4502delTAG, p.Ser1500del GnomAD AF (0,0071) susceptibility to **Amyotrophic lateral sclerosis-parkinsonism/dementia complex**, Autosomal dominant NEFH NM_021076.3, rs149571560; rs373980795 chr22, exon 4, c.1973_1978delAAGAGG, p.Glu658_Glu659del GnomAD AF(0,005) heterozygote disruptive_inframe_deletion susceptibility to **Amyotrophic lateral sclerosis**, Autosomal recessive

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

chr2, 201701817-201768913, 67kb	2q33.1del	loss of function	ALS2, ALS2-RELATED DISORDERS
chr9, 35057092-35077410, 20kb	9p13.3del	loss of function	VCP/FANCG, VCP-related developmental disorder (monoallelic)

INTERPRETATION

Aucun variant clairement pathogène et pouvant expliquer le phénotype n'a été mis en évidence.

On relève cependant la présence d'un VUS dans TRPM7 dont les mutations ont été associées l'ALS de l'île de Guam.[1]
Leur implication est finalement controversée et sont finalement considéré comme des facteurs de susceptibilité, notamment dans les populations de l'Ouest Pacific. [2]
Une étude de l'impact de ce variant permettrait de conclure à son caractère délétère.

Par ailleurs, la présence de petites délétions emportant tout ou partie des gènes ALS2, VCP et FANCG est à noter. Les mutations de se gènes sont responsables de phénotype ALS +/- atteintes cognitives, syndromiques ou juvénile.
L'étude des CNVs dans l'ALS tend à montrer leur existence dans ce phénotype.

REFERENCES

[1] Y. Sun, P. Sukumaran, A. Schaar, et B. B. Singh, « TRPM7 and its role in neurodegenerative diseases », *Channels*, vol. 9, n° 5, p. 253‑261, sept. 2015, doi: 10.1080/19336950.2015.1075675.

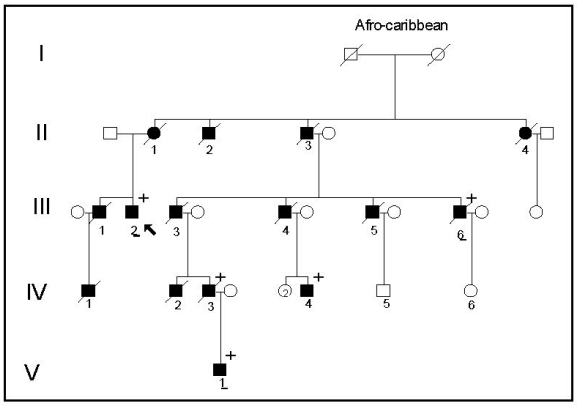
[2] M. C. Hermosura et R. M. Garruto, « TRPM7 and TRPM2-Candidate susceptibility genes for Western Pacific ALS and PD? », *Biochim. Biophys. Acta*, vol. 1772, n° 8, p. 822‑835, août 2007, doi: 10.1016/j.bbadis.2007.02.008.

13S001_16S073_10S030

SYNTHESE CLINIQUE

Les patients 13S001, 10S030 et 16S073 sont des individus d'une même famille. Ils sont atteints de SLA génétique en lien avec une mutation du gène *SOD1* (Gly94Val). Dans cette grande famille martiniquaise, on dénombre au moins 9 cas. Certains ont été suivis au centre. Nous avons appliqué un test de Mann-Whitney pour comparer l'âge d'apparition, le délai de survie et l'âge de décès parmi les générations III et IV les plus informatives. Le test a révélé des différences statistiques d'âge d'apparition (p = 0, 03) et d'âge de décès (p = 0, 01), mais pas de délai de survie (p = 0, 05). Il y aurait donc une anticipation statistiquement prouvée dans cette famille. Ce phénomène a déjà été rapporté dans la SLA mais jamais avec cette mutation.

Les variants d'expansion dans les gènes associées à la SLA ont été exclus. CoDESeq a été mis en œuvre pour trouver des facteurs synergiques susceptibles d'expliquer les âges de début précoces au fil des générations.



Correspondance des symboles : 13S001 ; III.2 / 10S030 ; III.6 / 16S073 ; V.1

RESULTATS

SINGLE NUCLEOTIDS VARIANTS, SNVs

CLASS 5	CLASS 4
13S001, 10S030, 16S073: *SOD1* NM_000454.4 exon4, c.281G>T, p.Gly94Val, heterozygote missense GnomAD AF ?, prédit pathogène >3 softs **Amyotrophic Lateral Sclerosis 1, autosomal dominant**	***13S001:*** *DCTN1* rs184147813 c.3517C>G; p.Pro1173Ala, exon 27 heterozygote missense GnomAD Afro 0,0001% susceptibility to autosomal recessive and dominant ALS and autosomal dominant distal hereditary motor neuropathy type VIIB
	10S030: *NEFH* - rs200634512 c.1941G>C; p.Lys647Asn exon 4 heterozygote missense GnomAD 0,001% - rs138156220 c.2601G>T; p.Lys867Asn exon 4 heterozygote missense GnomAD 0,0005%
	13S001, 10S030, 16S073: *TIAM1* - rs ? c.2464C>T p.Pro822Ser exon 13 missense hétérozygote GnomAD AF $4{,}06.10^{-6}$ - rs567504389 c.3882C>T c.3882C>T exon 24 missense hétérozygote GnomAD AF $2{,}43.10^{-5}$ facteur de susceptibilité de la SLA

Copy Number Variants, CNVs

(annotSv_merge_cnv_all_methods/annotation_decipher_cnv_allmethods)

16S073:

Xp11.21 Pos 43711851 to 77620567 hemizygote deletion 3.2 Mb *UBQLN2* ALS/ALS15 avec ou sans FTD [1]

21q22.3 Pos 44 318 436 to 44 333 294 duplication 14,858Kb *PFKL* and *C21orf2* ALS susceptibility

INTERPRETATION

L'analyse CoDESeq de ces ADN révèle que les patients de la 2e génération ont au moins un 2nd variant probablement pathogène dans un autre gène de susceptibilité à la SLA : *DCTN1* et *NEFH* [2]–[4]. Mais le plus jeune cas atteint de la famille ne présente a priori pas d'autres SNVs d'interprétation non ambigus susceptibles d'intervenir dans l'âge de début précoce de la maladie. Ce jeune patient, au contraire de ses apparentés, présente des CNVs contenant des gènes de susceptibilité à la SLA. Il s'agit d'une délétion hémizygote emportant le gène *UBQLN2* et une duplication comportant le gène *C21orf2*.[1], [5]

Ces variants supplémentaires au variant SOD1 plaideraient en faveur de mécanisme oligogénique pouvant expliquer des âges de début variables dans cette famille. D'ailleurs, le variant TIAM1 est également rapporté ségrége avec le variant pathogène *SOD1*, chez tous les apparentés. Ce gène est situé à proximité du gène SOD1 sur la bande 21q22.11 et est considéré comme un facteur de susceptibilité de la SLA par les étude d'association génome entier [6]

REFERENCES

[1] H.-X. Deng *et al.*, « Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia », *Nature*, vol. 477, n° 7363, p. 211‑215, sept. 2011, doi: 10.1038/nature10353.

[2] C. Munch *et al.*, « Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS », *Neurology*, vol. 63, n° 4, p. 724‑726, août 2004, doi: 10.1212/01.WNL.0000134608.83927.B1.

[3] P. Caroppo *et al.*, « *DCTN1* Mutation Analysis in Families With Progressive Supranuclear Palsy–Like Phenotypes », *JAMA Neurol.*, vol. 71, n° 2, p. 208, févr. 2014, doi: 10.1001/jamaneurol.2013.5100.

[4] K. Tripolszki *et al.*, « Comprehensive Genetic Analysis of a Hungarian Amyotrophic Lateral Sclerosis Cohort », *Front. Genet.*, vol. 10, p. 732, août 2019, doi: 10.3389/fgene.2019.00732.

[5] R. Chia, A. Chiò, et B. J. Traynor, « Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications », *Lancet Neurol.*, vol. 17, n° 1, p. 94‑102, janv. 2018, doi: 10.1016/S1474-4422(17)30401-5.

[6] H. Laaksovirta *et al.*, « Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study », *Lancet Neurol.*, vol. 9, n° 10, p. 978‑985, oct. 2010, doi: 10.1016/S1474-4422(10)70184-8.