



HAL
open science

Prédicteurs de l'utilité clinique du diagnostic moléculaire par séquençage haut débit de l'exome réalisé en première ligne chez les patients présentant un phénotype mendélien : étude observationnelle sur 603 patients consécutifs

Tom Alix

► **To cite this version:**

Tom Alix. Prédicteurs de l'utilité clinique du diagnostic moléculaire par séquençage haut débit de l'exome réalisé en première ligne chez les patients présentant un phénotype mendélien : étude observationnelle sur 603 patients consécutifs. Sciences pharmaceutiques. 2022. hal-04042809

HAL Id: hal-04042809

<https://hal.univ-lorraine.fr/hal-04042809>

Submitted on 23 Mar 2023

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

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



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

AVERTISSEMENT

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

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

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

Contact bibliothèque : ddoc-thesesexercice-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

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

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



UNIVERSITÉ
DE LORRAINE



FACULTÉ
DE PHARMACIE

UNIVERSITE DE LORRAINE
2022

FACULTE DE PHARMACIE

THÈSE

Présentée et soutenue publiquement
Le 03/06/2022, sur un sujet dédié à :

Prédicteurs de l'utilité clinique du diagnostic moléculaire par
séquençage haut débit de l'exome réalisé en première ligne chez des
patients présentant un phénotype mendélien : Étude observationnelle
sur 603 patients consécutifs

pour obtenir

le Diplôme d'Etat de Docteur en Pharmacie

par Tom ALIX
né(e) le 28/07/1994

Membres du Jury

Président :	Mme. Brigitte LEININGER-MULLER	PU : Faculté de Pharmacie
Juges :	M. Abderrahim OUSSALAH	PU-PH : Faculté de médecine, CHRU de Nancy
	M. Jean-Louis GUEANT	PU-PH : Faculté de médecine, CHRU de Nancy
	M. Jean-Yves JOUZEAU	PU-PH : Faculté de médecine, CHRU de Nancy
	Mme. Elise JEANNESSON	PH : CHRU de Nancy

UNIVERSITÉ DE LORRAINE
FACULTÉ DE PHARMACIE
Année universitaire 2021-2022

DOYEN

Raphaël DUVAL

Vice-Doyen

Julien PERRIN

Directrice des études

Marie SOCHA

Conseil de la Pédagogie

Président, Luc FERRARI

Vice-Présidente - vacant

Collège d'Enseignement Pharmaceutique Hospitalier

Présidente, Béatrice DEMORE

Commission Prospective Facultaire

Président, Igor CLAROT

Vice-Président, Raphaël DUVAL

Commission de la Recherche

Présidente, Caroline GAUCHER

Chargés de Mission

Communication

Aline BONTEMPS

Innovation pédagogique

Alexandrine LAMBERT

Référente ADE

Virginie PICHON

Référente dotation sur projet (DSP)

Marie-Paule SAUDER

Référent vie associative

Arnaud PALLOTTA

Responsabilités

Filière Officine

Caroline PERRIN-SARRADO

Filière Industrie

Julien GRAVOULET

Isabelle LARTAUD,

Jean-Bernard REGNOUF de VAINS

Filière Hôpital

Béatrice DEMORE

Marie SOCHA

Pharma Plus ENSIC

Jean-Bernard REGNOUF de VAINS

Pharma Plus ENSAIA

Xavier BELLANGER

Pharma Plus ENSGSI

Igor CLAROT

Cellule de Formation Continue et Individuelle

Luc FERRARI

Commission d'agrément des maîtres de stage

François DUPUIS

ERASMUS

Mihayl VARBANOV

DOYENS HONORAIRES

Chantal FINANCE

Francine PAULUS

Claude VIGNERON

PROFESSEURS EMERITES

Pierre LEROY

Philippe MAINCENT

Claude VIGNERON

PROFESSEURS HONORAIRES

Jean-Claude BLOCK

Pierre DIXNEUF

Chantal FINANCE

MAITRES DE CONFERENCES HONORAIRES

Monique ALBERT

Mariette BEAUD

François BONNEAUX

Marie-Madeleine GALTEAU
Thérèse GIRARD
Pierre LABRUDE
Vincent LOPPINET
Patrick MENU
Alain NICOLAS
Janine SCHWARTZBROD
Louis SCHWARTZBROD

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

ASSISTANTS HONORAIRES

Marie-Catherine BERTHE
Annie PAVIS

ENSEIGNANTS

Section CNU

*

Discipline d'enseignement

**PROFESSEURS DES UNIVERSITES -
PRATICIENS HOSPITALIERS**

Danièle BENSOUSSAN-LEJZEROWICZ

82

Béatrice DEMORE

81

Pharmacie clinique

Alexandre HARLE

82

Biologie cellulaire oncologique

Jean-Louis MERLIN

82

Biologie cellulaire

Jean-Michel SIMON

81

Economie de la santé, Législation pharmaceutique

Nathalie THILLY

81

Santé publique et Epidémiologie

PROFESSEURS DES UNIVERSITES

Ariane BOUDIER

85

Chimie Physique

Christine CAPDEVILLE-ATKINSON

86

Pharmacologie

Igor CLAROT

85

Chimie analytique

Joël DUCOURNEAU

85

Biophysique, Acoustique, Audioprothèse

Raphaël DUVAL

87

Microbiologie clinique

Béatrice FAIVRE

87

Hématologie, Biologie cellulaire

Luc FERRARI

86

Toxicologie

Pascale FRIANT-MICHEL

85

Mathématiques, Physique

Christophe GANTZER

87

Microbiologie

Frédéric JORAND

87

Eau, Santé, Environnement

Isabelle LARTAUD

86

Pharmacologie

Dominique LAURAIN-MATTAR

86

Pharmacognosie

Brigitte LEININGER-MULLER

87

Biochimie

Jean-Bernard REGNOUF de VAINS

86

Chimie thérapeutique

Anne SAPIN-MINET

85

Pharmacie galénique

Bertrand RIHN

87

Biochimie, Biologie moléculaire

**MAITRES DE CONFÉRENCES DES UNIVERSITÉS
- PRATICIENS HOSPITALIERS**

82

Caroline LAROYE		
Julien PERRIN	82	<i>Hématologie biologique</i>
Loïc REPEL	82	<i>Biothérapie</i>
Marie SOCHA	81	<i>Pharmacie clinique, thérapeutique et biotechnique</i>

MAITRES DE CONFÉRENCES

Xavier BELLANGER ^H	87	<i>Parasitologie, Mycologie médicale</i>
Emmanuelle BENOIT ^H	86	<i>Communication et Santé</i>
Isabelle BERTRAND ^H	87	<i>Microbiologie</i>
Michel BOISBRUN ^H	86	<i>Chimie thérapeutique</i>
Cédric BOURA ^H	86	<i>Physiologie</i>
Sandrine CAPIZZI	87	<i>Parasitologie</i>
Antoine CAROF	85	<i>Informatique</i>
Frédérique CHANGEY	87	<i>Microbiologie</i>
Sébastien DADE	85	<i>Bio-informatique</i>
Natacha DREUMONT ^H	87	<i>Biochimie générale, Biochimie clinique</i>
Florence DUMARCAY ^H	86	<i>Chimie thérapeutique</i>
François DUPUIS ^H	86	<i>Pharmacologie</i>
Reine EL OMAR	86	<i>Physiologie</i>
Adil FAIZ	85	<i>Biophysique, Acoustique</i>
Anthony GANDIN	87	<i>Mycologie, Botanique</i>
Caroline GAUCHER ^H	86	<i>Chimie physique, Pharmacologie</i>
Stéphane GIBAUD ^H	86	<i>Pharmacie clinique</i>
Jérémy GOUYON	85	<i>Chimie analytique</i>
Thierry HUMBERT	86	<i>Chimie organique</i>
Olivier JOUBERT ^H	86	<i>Toxicologie, Sécurité sanitaire</i>

ENSEIGNANTS (suite)

Section CNU

*

Discipline d'enseignement

Alexandrine LAMBERT	85	<i>Informatique, Biostatistiques</i>
Julie LEONHARD ^H	86/01	<i>Droit en Santé</i>
Balbine MAILLOU	85	<i>Biophysique, Acoustique, Audioprothèse</i>
Christophe MERLIN ^H	87	<i>Microbiologie environnementale</i>
Maxime MOURER ^H	86	<i>Chimie organique</i>
Coumba NDIAYE	86	<i>Epidémiologie et Santé publique</i>
Arnaud PALLOTTA	85	<i>Bioanalyse du médicament</i>
Marianne PARENT	85	<i>Pharmacie galénique</i>
Caroline PERRIN-SARRADO	86	<i>Pharmacologie</i>
Virginie PICHON	85	<i>Biophysique</i>
Sophie PINEL ^H	85	<i>Informatique en Santé (e-santé)</i>
Marie-Paule SAUDER	87	<i>Mycologie, Botanique</i>
Guillaume SAUTREY	85	<i>Chimie analytique</i>
Rosella SPINA	86	<i>Pharmacognosie</i>
Sabrina TOUCHET	86	<i>Pharmacochimie</i>
Mihayl VARBANOV	87	<i>Immuno-Virologie</i>
Marie-Noëlle VAULTIER ^H	87	<i>Mycologie, Botanique</i>
Emilie VELOT ^H	86	<i>Physiologie-Physiopathologie humaines</i>
Mohamed ZAIYOU ^H	87	<i>Biochimie et Biologie moléculaire</i>

PROFESSEUR ASSOCIE

Julien GRAVOULET 86 Pharmacie clinique

MAITRE DE CONFÉRENCE ASSOCIE

Pauline GILSON 82 Biologie cellulaire oncologique

PROFESSEUR AGREGE

Christophe COCHAUD 11 Anglais

^H Maître de conférences titulaire HDR

*** Disciplines du Conseil National des Universités :**

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

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

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

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

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

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

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

SERMENT DE GALIEN

En présence des Maitres de la Faculté, je fais le serment :

D'honorer ceux qui m'ont instruit(e) dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle aux principes qui m'ont été enseignés et d'actualiser mes connaissances

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

De ne jamais oublier ma responsabilité et mes devoirs envers la personne humaine et sa dignité

En aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser des actes criminels.

De ne dévoiler à personne les secrets qui m'auraient été confiés ou dont j'aurais eu connaissance dans l'exercice de ma profession

De faire preuve de loyauté et de solidarité envers mes collègues pharmaciens

De coopérer avec les autres professionnels de santé

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

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

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

Remerciements

A Madame le Professeur Brigitte Leininger

Je vous remercie d'avoir accepté la présidence de cette thèse. Veuillez recevoir mes sincères remerciements pour votre disponibilité, et pour les enseignements que vous m'avez prodigués durant cette année de master 2.

A Monsieur le Professeur Abderrahim Oussalah

Je vous remercie d'avoir accepté de diriger cette thèse, et de m'avoir aussi bien encadré durant tout ce travail. Je vous remercie également pour mes 2 passages en stage dans votre service où j'ai pu approfondir mes connaissances en biologie moléculaire.

A Monsieur le Professeur Jean-Louis Guéant

Je vous remercie d'avoir accepté de suivre l'avancée de ce travail, mais également pour mes passages en stage au sein de votre service. Je vous remercie pour vos conseils. Veuillez recevoir l'expression de mon respect et de ma profonde reconnaissance.

A Monsieur le Professeur Jean-Yves Jouzeau

Je souhaiterais vous remercier d'avoir accepté de participer à ce jury de thèse. Je vous remercie également, ainsi que toute l'équipe de Pharmaco-toxicologie pour mon court passage au sein de l'équipe.

A Madame le Docteur Elise Jeannesson

Je te remercie d'avoir accepté de juger ce travail. Je te remercie également pour le temps que tu as pris lors de mes passages en stage pour me former, et me faire découvrir les secteurs dont tu t'occupes.

A ma famille

Tout d'abord, un grand merci à toute ma famille sans qui je ne serais certainement pas là. Merci à **Papa** et **Maman** de s'être autant investis dans notre éducation et notre épanouissement : toutes les activités que vous nous avez fait faire, tous les trajets parcourus pour nous emmener ici et là, tous les voyages qu'on a faits ensemble... Merci à **Laure**, pour tout ce qu'on a pu partager, et même ce qui nous divise (la nourriture à base de graines pour oiseaux, les programmes TV). Merci à tous les 3.

Un grand merci à tous mes grands-parents : **papi Bob**, **mamie Françoise**, **mamie Hélène** et **papi Gilbert** d'avoir toujours été là pour nous. Merci également à **Gaël & André**, **Rémi & Mathias** pour tous les Noëls, toutes les croisières et toutes les parties de Mario qu'on a pu faire (c'est moins drôle depuis que je gagne moins).

Une grosse pensée pour **Martine** et **Michel** ! Merci d'avoir toujours été présents, même si les entrevues se font rares (surtout depuis ma montée dans le Nord...).

En bref : merci à tous d'avoir toujours été là et pour le soutien que vous apportez.

A toute l'équipe de Chassieu Gym (les vieux et les moins vieux)

Bertrand : Pour m'avoir supporté toutes ces années : finalement, je suis peut-être en partie responsable des cheveux blancs ? Merci pour les centaines de pépites que tu nous as sorti (dont les meilleurs : « Je suis pas maitre-chien moi ! » ; « C'est pas une réception c'est une déception ça ! »).

Jean Lo : Pour m'avoir fait commencer la gym, et pour ton « petit » côté râleur qui me rendait moins seul au gymnase.

La team des vieux (**Marine**, **Marwan**, **Didou**, **Victo**, **Flavie**, **Audran**, **Vincent**) et Chassieu Gym la résistance (**Rémi**, **Enzo/Enzo**, **Léo**, **Mathias**, **Martin**) : A tous les entrainements passés autour du bac à magnésie ou accoudé sur le cheval d'arçons, pour toutes les soirées et évènements et à tous les débats débiles des entrainements.

Aux amis de la Fac de Lyon

Vilain petit canard et les internistes : A tous ceux de la 1^{ère} heure de Repas fac/Tarot/TACOS et aux arrivées plus tardives. **Antoine & Rousseline** pour les soirées en trio à jouer au tarot banane, **Mélissa** pour les voyages au bout du monde et ton dynamisme naturel, **Elsa** la suisse pour tes cheveux cheveux, **Mélanie** pour ta bonne humeur au volant, **Chloé** pour la découverte de l'escalade, **Laura** la fameuse retardataire infiltrée parmi les internistes et aux perdues de vue **Coralie** et **Elise**...

Les amis de P1 : **Cyrielle Noémie Pauline** : Pour cette double année, pas si horrible que ça !

Pour tous les délires de ces années fac (pas encore terminées pour certains) et aux délires futurs !

A tous mes amis du grand Nord

« Vous savez, moi je ne crois pas qu'il y ait de bonne ou de mauvaise situation. Moi, si je devais résumer ma vie aujourd'hui avec vous, je dirais que c'est d'abord des rencontres ». Grâce à vous l'expatriation nancéenne est moins rude !

A mes colocataires des 4 Terres :

Stéphanie : Pour le Nord-Ouest ! Pour ton débit verbal largement supérieur à celui d'Eminem et ta capacité à raconter encore et toujours la même histoire avec les mêmes mots, le même entrain et le même ton. Attention aux parasites ! Y en a un peu plus dans la carrière de Cabassou que dans l'océan Manche ! Garde une énergie usante, spontanée & épuisante.

Marine : Pour le Sud-Ouest ! Pour ta capacité à cracher des flammes et à changer les lois de la physique ! On aura presque réussi à faire disparaître ton accent toulousain au profit de l'accent vosgien... Profite bien de Reims ! Derrière : retour abords garonnais ou nancéiens ?

Mélissa : Pour le Nord et le Sud ! Pour avoir relancé l'économie de la France à coup de robe à fleurs (attention il faut toujours le même motif !). On viendra te voir dans ta future ville ! Tes colocataires te manqueront ! Je sais ce que tu te dis : Viendront-Ils Encore Illuminer Le Lieu Elu ?

Quentin : Pour la Meuuuuuse ! Tu as finalement réussi à habiter pour de vrai dans notre immeuble ! (J'avoue, poltronesofa et ta PS5 ont bien aidé). Pour objectiver 1 de tes sublime soliloque : « Ammghhmg Rddpodghm Dmmgffh ».

A tous les co-internes avec qui j'ai eu la chance de partager un stage, ou pas : **Laurence** arrête de me suivre dans mes stages stp, **Anne Jul** (& **Clément** !) les 93 c'est surfait (ça se voit au bowling), **MélaAdriainen** meilleur site de comparateur de la moellosité des oreillers, **Stan** et tes pantalons en velours, **Emma** pour ta ponctualité légendaire (tu es probablement en retard pour le pot d'ailleurs), **Jean-Marie** fameux Major de sa promotion, **Marion** d'avoir encadré la meilleur promo d'internes ever, **Rachard** je connais toujours pas le nom de tes chiens mais prends soin de leur santé fragile, **Gaétan** qui essaye de nous faire grossir avec ses gaéaux nutella, **Jonas** pour montrer qu'il est jamais trop tard pour venir en bio, **Guillaume B** le meilleur caissier covid, **Guillaume P** sans qui l'affiche petit déjeuner n'aurais pas été possible, **Alizée & Christopher** les méchantes winx d'hémato, **Clara** petit ange parti trop tôt à The Killer, **Clémence** mon alliée lyonnaise de qualité, **Guillaume D** qu'on retrouve « par hasard » en bas des volcans en éruption et enfin à toute la team infectio **Steven/Caroline/Laura** qu'on retrouvera peut être un jour en bio (qui sait).

Aux copains de l'Internat Lobau : Solange, Paulain, Charles-Edouard, Maxime, Ludo, Thomas, ... Pour toutes les soirées tarot, Koh-Lantarot, GoTarot, les tonus, les parties estivales de mölkky...

A Versaille sans s : **Dono, Chloé, Alexandre, Julien, Ophélie** pour nous avoir fait découvrir les cultures vosgienne et meusienne, et pour le titre de champion de Meuse de beer-pong qu'on a raflé au passage.

A toute l'équipe des services de Biochimie, PharmacoTox, Génétique du CHRU de Nancy : Biologistes, ingénieurs, techniciens, secrétaires... Je vous remercie pour le temps que vous avez pris pour me former et me faire découvrir le métier de Biologiste.

Table des matières

Liste Tableaux	3
Liste Figures	3
Liste Annexes	3
Préambule	4
I) Introduction générale sur les maladies héréditaires.....	5
A) EPIDEMIOLOGIE DES MALADIES RARES HEREDITAIRES	5
1) Les maladies rares	5
2) Les maladies héréditaires.....	6
3) Contexte règlementaire du diagnostic des maladies constitutionnelles	8
B) STRATEGIE NATIONALE POUR LES MALADIES RARES	9
1) Plans nationaux « maladies rares ».....	9
2) Développement de la médecine moléculaire	11
C) APPROCHES MOLECULAIRES DIAGNOSTIQUES – LE SEQUENÇAGE	13
1) Séquençage Sanger	14
2) Séquençage haut débit (NGS) Illumina®.....	16
3) Séquençages de 3 ^{ème} génération.....	26
4) Classification des variants	27
II) Problématique de l'étude.....	29
A) CONTEXTE DE L'ETUDE	29
B) DESCRIPTION DE LA COHORTE.....	29
C) OBJECTIFS DE L'ETUDE	30
III) « Predictors of the Clinical Utility of Clinical Exome Sequencing as a First-Tier Genetic Test in Patients with Mendelian Phenotypes: Results from a Referral Center Study in 603 Consecutive Cases ».....	32
IV) Discussion.....	59
A) OBJECTIF PRINCIPAL : EFFICIENCE DIAGNOSTIQUE.....	59
B) OBJECTIF SECONDAIRE : PREDICTEUR DE LA MISE EN EVIDENCE D'UN VARIANT DE CLASSE 4 OU 5	61
C) OBJECTIF SECONDAIRE : PREDICTEURS DE LA CONFIRMATION DU DIAGNOSTIC SUSPECTE	62
D) OBJECTIF SECONDAIRE : THERAPIE GUIDEE PAR LE CES	63
E) LIMITES ET PERSPECTIVES	64
Bibliographie	66
Annexes.....	76

Abréviations

ACMG	American College of Medical Genetics and Genomics
ADN	Acide désoxyribonucléique
ALD	Affection Longue Durée
BA	Benign stand Alone
BAM	Binary Alignment Map
BBMN	Biochimie Biologie Moléculaire Nutrition
BP	Benign Poor or supporting
BS	Benign Strong
CES	Clinical Exome Sequencing
CHRU	Centre Hospitalier Régional et Universitaire
CRM	Centres de Référence Maladies Rares
CNV	Copy Number Variation
ENCODE	Encyclopedia of DNA Elements
GC	Guanine Cytosine
HGVS	Human Genome Variation Society
HNPCC	Cancer du Colon Hériditaire non polyposique
HPO	Human Phenotype Ontology
HRM	High Resolution Melt
IGV	Intergrative Genomics Viewer
MLPA	Multiplex Ligation-dependent Probe Amplification
NTP	Désoxyribonucléoside triphosphate
OMIM	Online Mendelian Inheritance in Man
OTC	Ornithine TransCarbamylase
PCR	Polymerisation Chain Reaction
PCU	Phénylcétonurie
PM	Pathogenic Moderate
PP	Pathogenic Poor or supporting
PS	Pathogenic Strong
PVS	Pathogenic Very Strong
SAM	Sequence Alignment Map
NGS	Next Generation Sequencing
SLR	Synthetic Long Reads
SMRT	Single-Molecule Real-Time Sequencing
TSA	Troubles du Spectre Autistique
VCF	Virtual Card File
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Liste Tableaux

Tableau I : Hétérogénéité de la définition de « maladie rare » à travers le monde	5
Tableau II : Mode de transmission des variants pathogènes (13,14)	7
Tableau III : Axes développés pour la 1 ^{ère} mobilisation du plan national « maladies rares » (8).....	10
Tableau III : Spécificité des techniques de NGS, d'après Stranneheim et Lundeberg (46)	16
Tableau IV : Exemple de pathologies et nombre de gènes impliqués d'après Lamoril et Bogard (55) 22	
Tableau V : Comparaison des différentes approches de séquençage (64)	24
Tableau VI : Articles de la littérature rapportant l'efficacité diagnostique chez des patients présentant un phénotype Mendélien	60

Liste Figures

Figure 1 : Maladies mendéliennes classiques versus phénotypes complexes (25).....	11
Figure 2 : Distribution des centres de références et de compétences des maladies rares du métabolisme (34)	13
Figure 3 : Principe de la PCR (36).....	14
Figure 4 : Principe du séquençage Sanger.....	15
Figure 5 : Etapes technique du NGS	17
Figure 6 : Structure d'un transposome (46)	17
Figure 7 : <i>Bridge PCR</i> , d'après Illumina (48)	18
Figure 8 : Principe du Séquençage par synthèse avec terminateurs de chaîne fluorescents, d'après Metzker (49)	19
Figure 9 : Exemple de profondeur et couverture d'un séquençage en 12q23.2 (sur IGV)	20
Figure 10 : Exemple de l'arborescence HPO pour l'HP:0002160, et gènes impliqués	21
Figure 11 : Homogénéisation de la classification des variants (73)	28
Figure 12 : Répartition des indications classés par famille de maladie selon les informations disponibles en pré-CES	31
Figure 13 : Répartition de la nature des mutations rapportées.....	61
Figure 14 : Adaptation thérapeutique due au CES.....	63

Liste Annexes

Annexe 1 : Modèle de consentement utilisé au CHRU de Nancy	76
Annexe 2 : Formulaire de compte-rendu des NGS prodigués dans le laboratoire BBMN du CHRU de Nancy.....	77
Annexe 3 : Liste des gènes couverts par les panels Trusight One et Trusight Expanded.....	78

Préambule

Ce document rapporte les résultats originaux d'un travail porté sur les « Prédicteurs de l'utilité clinique du diagnostic moléculaire par séquençage haut débit de l'exome réalisé en première ligne chez des patients présentant un phénotype mendélien ». Il s'agit d'une étude rétrospective observationnelle portant sur 603 patients consécutifs.

Ce manuscrit comporte quatre parties :

- La 1^{ère} est consacrée à une introduction générale sur les maladies héréditaires et la stratégie de diagnostic et de prise en charge en France ;
- La 2^{ème} partie introduira la genèse et la problématique de l'étude ;
- La 3^{ème} rapporte l'article découlant du projet (actuellement soumis dans la revue scientifique : *Journal of Medical Genetics*) ;
- La dernière partie sera consacrée à une discussion générale des résultats et leur mise en perspective par rapport à la pratique clinique.

I) Introduction générale sur les maladies héréditaires

A) EPIDEMIOLOGIE DES MALADIES RARES HEREDITAIRES

1) Les maladies rares

En Europe, une maladie rare est caractérisée par le fait qu'elle affecte moins de 50 individus sur 100 000 (1). Cette définition varie néanmoins en fonction des différents référentiels nationaux (Tableau I).

Tableau I : Hétérogénéité de la définition de « maladie rare » à travers le monde

Région géographique	Définition d'une maladie rare selon le référentiel national	Nombre d'individus atteints rapporté à 100 000 habitants*
Etats-Unis d'Amérique	< 200 000 individus atteints (2)	< 61 / 100 000
Australie	< 5 / 10 000 habitants (3)	< 50 / 100 000
Japon	< 50 000 individus atteints (4)	< 40 / 100 000

*En fonction du nombre d'habitants recensés dans le pays en 2021.

Les maladies rares sont donc regroupées en fonction de leur fréquence dans une population donnée. Elles représentent un ensemble de pathologies très hétérogènes de par leurs étiologies, leurs symptomatologies, et même de par leurs incidences (même si celles-ci sont par définition limitées). Bien que « rares », la multiplicité de pathologies rentrant dans la définition fait qu'elles affectent au total jusqu'à 10% de la population (5). Leur impact sur la santé publique est donc non négligeable. Orphanet® est une organisation européenne qui a pour objectif d'identifier, de classer, et de décrire les maladies rares. Elle fournit une base de données spécialisées sur le sujet qui recense en juin 2021 : 6172 pathologies rares (6).

Au niveau étiologique, les maladies rares peuvent être des pathologies infectieuses (exemples : Grippe aviaire ORPHA:454831), des cancers (Périneuriome des tissus mous ORPHA:100002), des maladies immunologiques (Lupus érythémateux discoïde ; ORPHA:90281), ou toxiques (Syndrome malin des neuroleptiques ; ORPHA:94093), etc... Mais 80% de ces pathologies ont une étiologie génétique identifiée soit en lien avec des anomalies chromosomiques (exemple : Syndrome 18,XXYY ; ORPHA:10) ou avec des anomalies géniques (exemple : Ataxie-Télangiectasie ; ORPHA:100) (7).

Le plus souvent, il s'agit de pathologies sévères d'évolution progressive : 65% des maladies rares sont considérées comme invalidantes. Elles provoquent des douleurs chroniques chez 1/5^{ème} des patients et des déficits moteurs, sensoriels ou intellectuels dans un cas sur deux (8). Généralement, ce sont des pathologies qui se déclarent précocement

dans la vie de l'individu : 50-75% des individus atteints sont des enfants (7) et la maladie se déclare dans 2/3 des cas avant l'âge de 2 ans (8). Elles sont une cause importante d'hospitalisations et de décès pédiatriques : 35% des décès au cours de la 1^{ère} année de vie sont dues à des maladies rares, et environ 30% des patients atteints meurent avant d'avoir 5 ans (9). Diagnostiquer les maladies rares est difficile, les errances diagnostiques sont fréquentes et pour la plupart de ces pathologies, les mécanismes physiopathologiques ne sont pas élucidés.

2) Les maladies héréditaires

Les maladies génétiques sont des affections dont la cause est une altération de la séquence d'ADN de l'individu. De nombreuses pathologies peuvent avoir une composante génétique, ainsi, seules les affections dans lesquelles la part de la génétique est particulièrement marquée peuvent être considérées comme telles (10). Elles peuvent correspondre à des anomalies chromosomiques de grandes tailles, ou à la modification d'une paire de base à une position précise du génome.

Bien que généralement « rares » certaines maladies héréditaires ne rentrent pas dans la définition. C'est le cas par exemple de l'hémochromatose génétique, de l'hypercholestérolémie familiale et de la drépanocytose dont les prévalences sont respectivement de 300, 100, et 58 (11) pour 100 000 habitants.

La majorité des maladies héréditaires se transmettent de manière Mendélienne, mais d'autres modes de transmissions non Mendéliens sont décrits, bien que plus rares (Tableau II). La maladie peut être héritée selon un mode récessif ou dominant, ou être de novo (le taux de mutation génique pathologique étant estimé à $3,2 \cdot 10^{-6}$ par génération (12)). Une fois l'anomalie présente chez l'individu, elle peut être transmise à sa descendance.

Tableau II : Mode de transmission des variants pathogènes (13,14)

	Mode de transmission		Exemples de pathologies OMIM / locus
Transmission mendélienne	Autosomique	Dominant	Neurofibromatose OMIM 162200 / <i>NF1</i>
		Récessif	Phénylcétonurie OMIM 261600 / <i>PAH</i>
	Lié à l'X	Dominant	Rachitisme hypophosphatémique OMIM 307800 / <i>PHEX</i>
		Récessif	Déficit en OTC OMIM : 311250) / <i>OTC</i>
Transmission non mendélienne	Pathologies mitochondriales	Transmission exclusivement maternelle via l'ADN mitochondrial	Déficit en cytochrome c oxydase OMIM 604377 / <i>SCO2</i>
	Maladies expansion de triplet	L'augmentation du nombre de triplets nucléotidiques de certains gènes est associée à la présence ou à une aggravation clinique de la pathologie	Syndrome de l'X Fragile OMIM 300624 / <i>FMR1</i>
	Empreinte parentale	La pathologie est due à une sous expression ou une surexpression d'un gène soumis à empreinte parentale en cas de disomie uniparentale.	Syndrome de Prader-Willi OMIM 176270 / 15q11-q13

La plupart des pathologies mendéliennes décrites sont dues à des anomalies monogéniques (15) sur les parties codantes des gènes. Seul 1,9 % du génome humain code pour des protéines (16) : cela correspond aux régions exoniques. Les autres séquences (introns, pseudogènes, séquences répétées, régions hétérochromatiques composées de chromatine extrêmement condensée), bien que correspondant à la grande majorité du génome, n'ont pas de rôle clairement élucidé. Néanmoins, elles semblent indispensables à l'expression des gènes en régulant par exemple l'activité transcriptionnelle. Jusqu'en 2021, la cartographie précise de l'ADN n'était pas entièrement connue. Le projet international ENCODE lancé en 2003 avait pour objectif de lever ces interrogations en délimitant de manière précise et exhaustive les segments du génome humain qui codent pour des éléments fonctionnels (17). Plus récemment, l'avènement des nouvelles technologies de séquençage haut débit de 2^e et de 3^e génération dans le cadre du consortium Consortium Telomere-to-Telomere (T2T) a permis en 2022 le séquençage intégral du génome humain incluant les 8 % du génome jusque-là non accessibles et correspondant aux régions hétérochromatiques (18). La séquence du génome rapportée par le consortium T2T comporte 3,055 milliards de paires de bases d'un génome

humain (T2T-CHM13) qui comprend des assemblages sans interruption pour tous les chromosomes sauf le chromosome Y ((18)). Les régions complétées comprennent tous les réseaux satellites centromériques, les duplications segmentaires récentes et les bras courts des cinq chromosomes acrocentriques (chromosomes 13, 14, 15, 21, 22) (18).

3) Contexte règlementaire du diagnostic des maladies constitutionnelles

Les maladies constitutionnelles, une fois diagnostiquées, sont définitives et peuvent avoir des conséquences chez la famille de la personne testée (en termes de pathologie ou de conseil génétique). Leur diagnostic est encadré par des lois et des organismes particuliers, comme l'agence de biomédecine. C'est pourquoi des règles de bonnes pratiques en génétique constitutionnelles ont été établies par la HAS (19) :

- La prescription : elle doit être justifiée par un retentissement clinique, psychologique, ou familial. Le fait qu'un test soit disponible et réalisable ne justifie pas à lui seul sa prescription. Le patient peut être symptomatique ou non. Dans tous les cas un consentement libre, éclairé et révocable à tout moment doit être signé (Annexe 1) par le patient avant le test et après l'avoir informé de manière complète et appropriée. Le prescripteur peut être un généticien, un conseiller en génétique sous la responsabilité d'un médecin qualifié en génétique, ou un médecin non-généticien connaissant la situation clinique et les conséquences d'un diagnostic constitutionnel, en collaboration avec une équipe de génétique clinique. Des protocoles nationaux de diagnostic et de soins, comprenant des arbres décisionnels et des recommandations diverses existent pour faciliter la prise en charge du patient.
- La réalisation du test : les laboratoires réalisant ces analyses doivent obtenir une autorisation de l'ARS. Ils ne réalisent que celles dont leur expertise est suffisante : c'est pourquoi une organisation en réseau s'est développée au niveau national. Un rapport d'activité annuel doit être rendu à l'ARS et à l'Agence de la Biomédecine. De même, le praticien qui valide le résultat doit être agréé par l'Agence de Biomédecine.

Le rendu des résultats : Un compte rendu doit être généré par le laboratoire exécutant. Il doit mentionner les variants ayant une conséquence clinique connue en rapport avec l'indication de la prescription ou sans rapport avec l'indication si le patient y a consenti, et les variants dont la conséquence clinique est inconnue (avec mention qu'ils ne doivent pas être utilisés pour la prise en charge en l'état actuel des connaissances). Les variants connus sans conséquences cliniques ne doivent pas être mentionnés. Le compte rendu est ensuite remis au prescripteur qui remettra lui-même les résultats au patient lors d'une consultation dédiée. Les connaissances du génome humain évoluant sans cesse, une réinterprétation du résultat peut être demandée par le médecin prescripteur. Le modèle du

compte-rendu utilisé au CHRU de Nancy par le laboratoire BBMN est retrouvé en Annexe 2.

B) STRATEGIE NATIONALE POUR LES MALADIES RARES

1) Plans nationaux « maladies rares »

Malgré l'hétérogénéité des maladies rares, la volonté des institutions de les regrouper a été motivée par la facilitation de la prise en charge de ces pathologies, et pour favoriser la recherche scientifique. Cela a également permis de définir la notion de « maladies orphelines » qui sont des maladies rares pour lesquelles aucun traitement spécifique n'existe (20).

Les maladies rares souffrent globalement d'un manque de connaissances médicales et scientifiques. Du fait de leur rareté et des faibles effectifs de malades, il est en effet difficile de conduire des études interventionnelles. De plus, elles ne sont apparues que récemment dans les politiques de recherche et de santé publique.

Les patients atteints rencontrent encore de nos jours des difficultés dans leur parcours de soin, tant sur le plan diagnostique, que pour leur suivi et leur orientation vers les professionnels compétents. En effet, seulement une personne atteinte d'une maladie rare sur deux est diagnostiquée ; et pour un quart d'entre elles, le diagnostic met en moyenne 5 ans à être posé (21). Ainsi, les maladies rares génèrent des répercussions non seulement médico-sociales, mais également économiques. Pour cela, une organisation nationale a été décidée puis mise en place avec la création de 3 « plans nationaux français : maladies rares » :

- 2005-2008 : Première mobilisation de l'ensemble des acteurs du domaine des maladies rares, et identification des centres de références et de compétence. Ce plan s'articule autour de 10 axes, détaillés dans le tableau III.

**Tableau III : Axes développés pour la 1^{ère} mobilisation du plan national « maladies rares »
(8)**

Axe	Objectif
1	Développer les connaissances épidémiologiques
2	Améliorer la reconnaissance des maladies rares en tant qu'ALD
3	Développer une information pour tous
4	Former les professionnels de santé à mieux identifier
5	Organiser le dépistage et l'accès aux tests diagnostiques
6	Améliorer l'accès aux soins et la qualité de la prise en charge
7	Poursuivre l'effort en faveur des médicaments orphelins
8	Répondre aux besoins spécifiques d'accompagnement des personnes atteintes de maladies rares
9	Promouvoir la recherche sur les maladies rares
10	Développer des partenariats nationaux et européens

- 2011-2016 : Cette seconde mobilisation s'inscrit dans la continuité du 1er, avec une volonté de poursuivre les actions engagées précédemment, et d'en amplifier les effets (22).
- 2018-2022 : Le 3^{ème} plan national maladies rares se base sur 5 objectifs principaux : accélérer le diagnostic de la maladie et ainsi réduire l'errance diagnostique (avec comme objectif de la réduire à 1 an) ; favoriser la recherche pour accroître des moyens thérapeutiques ; améliorer la qualité de vie et l'autonomie des personnes malades ; communiquer et former tous les acteurs en lien avec les maladies rares ; moderniser les organisations et optimiser les financements nationaux (23).

109 centres de références pour la prise en charge des maladies rares (CRM) ont été labellisés par les ministères chargés de la Santé, de la recherche et de l'innovation sur la période 2017-2022 (21). Ces centres rassemblent une équipe hospitalière spécialisée avec une expertise avérée pour une maladie rare ou un groupe de maladies rares. Cette équipe est pluridisciplinaire, avec des compétences médicales, paramédicales, psychologiques, médico-sociales, éducatives, et sociales. Des partenariats avec les associations de malades sont également nécessaires. Les centres ont pour objectif d'assurer la prise en charge et d'organiser le parcours de santé des individus concernés (24).

Ces plans s'inscrivent dans un contexte général d'une place croissante de la génomique dans la routine diagnostique des maladies héréditaires.

2) Développement de la médecine moléculaire

Une maladie héréditaire A n'est pas forcément due à la mutation d'un gène B. Hormis les maladies mendéliennes classiques (comme la dystrophie musculaire de Duchenne médiée par des mutations du gène *DMD* (25)), la majorité des pathologies implique l'interaction entre le bagage génétique de l'individu et son environnement dans des proportions variables. Généralement, plus la pathologie est complexe, plus le nombre de gènes impliqués est conséquent, et plus l'impact de l'environnement est présent (figure 1) (26).

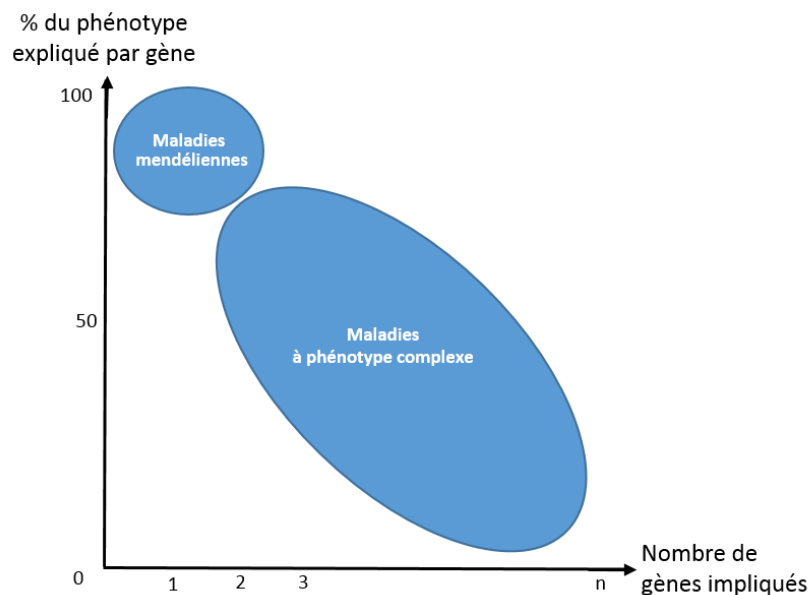


Figure 1 : Maladies mendéliennes classiques versus phénotypes complexes (26)

Les pathologies génétiques complexes ont ainsi une hétérogénéité génétique importante : ce qui va de pair avec l'hétérogénéité clinique observée dans de nombreux syndromes à première vue similaires (27–29). Ainsi, l'identification des mutations géniques, puis de la fonction du produit du gène, permet de révéler la biologie physiologique et celle de la maladie (26).

L'identification des gènes associés aux maladies complexes nécessite d'adopter des stratégies et des méthodes différentes, aussi bien en termes de recrutement des sujets de recherche qu'en termes des méthodologies utilisées (30).

C'est dans ce contexte, avec l'essor des nouvelles technologies et la volonté de se rapprocher au plus près de la médecine personnalisée que la médecine moléculaire a pu se développer. Celle-ci a pour objectif d'étudier les mécanismes moléculaires chez un individu, et d'en identifier les voies altérées par une mutation génétique. Les bénéfices des technologies de séquençage de nouvelle génération utilisant des « panels » de gènes et d'exomes à

moyen/haut débit en matière de diagnostic, sont aujourd'hui bien démontrés dans les domaines des maladies rares et de la cancérologie. Elles représentent ainsi un réel enjeu de santé publique. Le plan France Médecine Génomique 2025 a pour ambition de répondre à ce besoin par le biais de 4 objectifs : 1) définir la place et l'importance du séquençage dans la médecine actuelle et ses développements futurs ; 2) renforcer la position de la France dans le domaine de la recherche sur la génomique ; 3) évaluer les enjeux associés en matière d'innovation, de valorisation et de développement économique ; 4) proposer un modèle médico-économique viable sur le long terme, comprenant la prise en charge par l'Assurance maladie et le développement d'une filière industrielle adaptée (31).

De plus en plus de centres se dotent de plateformes de biologie moléculaire et de professionnels de santé / bio-informaticiens capables de générer d'importants panels de gènes, des exomes, voire des génomes entiers. En décembre 2016, le ministère chargé de la Santé a lancé un appel à projet qui a permis le financement des 2 premières plateformes de séquençage génomique à haut débit : SeqOIA (Sequencing, Omics, Information Analysis : AP-HP (32)) et AURAGEN (Auvergne Rhône-Alpes Génomique (33)).

Ce contexte tend à favoriser une médecine personnalisée pour permettre une meilleure prise en charge des patients, adaptée précisément aux mécanismes moléculaires anormaux. De nombreux centres utilisent les nouvelles technologies de séquençage, avec des panels de gènes de plus en plus larges, dans le cadre de la routine diagnostique.

La filière G2M (Maladies rares Héritaires du Métabolisme), labellisée en 2014 dans le cadre du 2ème plan National Maladies Rares a, entre autres, comme objectif de définir et nommer les différents Centres de Référence Maladies Rares (CRMR) (34). Le CHRU de Nancy est l'un des 8 CRMR du métabolisme (Figure 2).



Figure 2 : Distribution des centres de références et de compétences des maladies rares du métabolisme (35)

C) APPROCHES MOLECULAIRES DIAGNOSTIQUES – LE SEQUENÇAGE

Dans le cadre du diagnostic des maladies moléculaires, la caractérisation précise du phénotype permet d’orienter le laboratoire qui utilise ensuite la technique la plus appropriée. Le phénotypage est le point de départ essentiel au diagnostic moléculaire. Pour les troubles limités à un seul organe, la plupart des phénotypages doivent être effectués par des spécialistes de l’organe concerné (par exemple, des cardiologues pour les cardiopathies congénitales non syndromiques). Cependant, pour les troubles multi systémiques et pléiotropes, l’expertise du généticien médical sera essentielle. La plupart des troubles mendéliens présentent une hétérogénéité phénotypique importante avec une variabilité interindividuelle, intrafamiliale et interfamiliale de l’expressivité et même de la pénétrance. Un phénotypage effectué avec précision augmente la probabilité d’identifier le variant responsable (36).

Pour le diagnostic, de nombreuses techniques de biologie moléculaire existent, avec généralement une étape d’amplification de l’ADN par Polymerase Chain Reaction (PCR) afin de générer en quantité suffisante la séquence voulue.

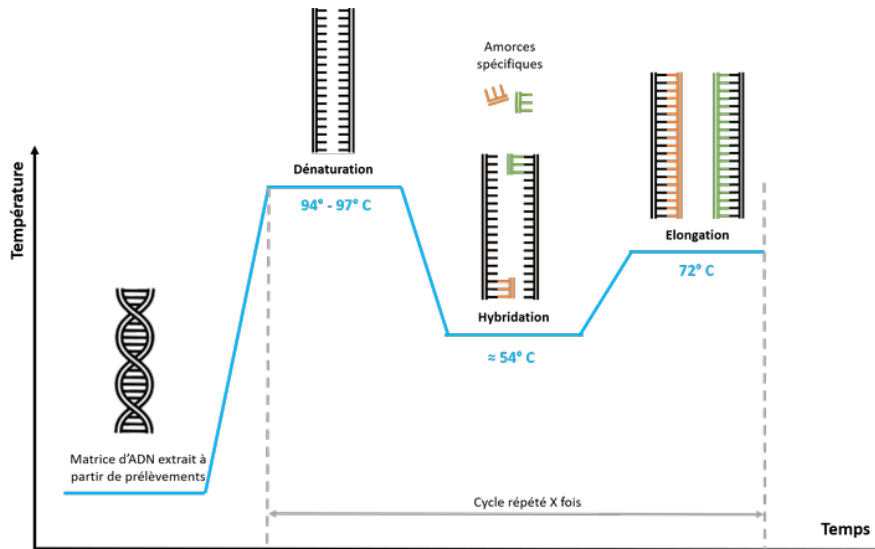


Figure 3 : Principe de la PCR (37)

La PCR est une technique de biologie moléculaire qui permet l'amplification d'une séquence génomique précise grâce à des amorces spécifiques s'hybridant autour de cette séquence. Trois étapes sont nécessaires à l'amplification et forment 1 cycle de PCR : 1) Dénaturation des doubles brins d'ADN à haute température (94-97°C), 2) Hybridation des amorces spécifiques (54°C), 3) Elongation des brins père à la température optimale de la Taq polymérase (environ 72°C). En fonction de la quantité d'ADN souhaitée, le cycle se répète n fois pour obtenir 2^n fragments.

Au cours des dernières décennies, les techniques de séquençage évoluent et se multiplient ; tout en devenant parallèlement de plus en plus abordables économiquement (entre 2009 et 2014, le prix du séquençage du génome humain a par exemple été divisé par 200 (38)). L'utilisation en routine de ces techniques a permis la mise en évidence de nombreux variants.

1) Séquençage Sanger

La première génération de séquençage a été imaginée dans les années 1970 par Sanger et son équipe (39,40). Celle-ci se base sur la réplification de l'ADN.

Avec des conditions physico-chimiques favorables, en présence de désoxyribonucléotides (dNTP) et d'une ADN polymérase, l'ADN simple brin va se répliquer. Les dNTPs se placent les uns à la suite des autres de manière complémentaire au brin père. L'enjeu du séquençage Sanger est de déterminer la séquence d'un brin d'ADN lors de son élongation. Pour cela, il utilise des amorces nucléotidiques qui servent de matrice à l'élongation ; et des didésoxyribonucléotides (ddNTP) qui sont capables de se substituer aux dNTP lors de l'élongation. Les ddNTP, du fait de leur effet terminateur de chaîne, bloquent l'élongation dès leur incorporation (41). Après migration sur gel de polyacrylamide en associant la nature du ddNTP intégré et la taille des fragments, il est possible d'en déduire la séquence du fragment (Figure 4).

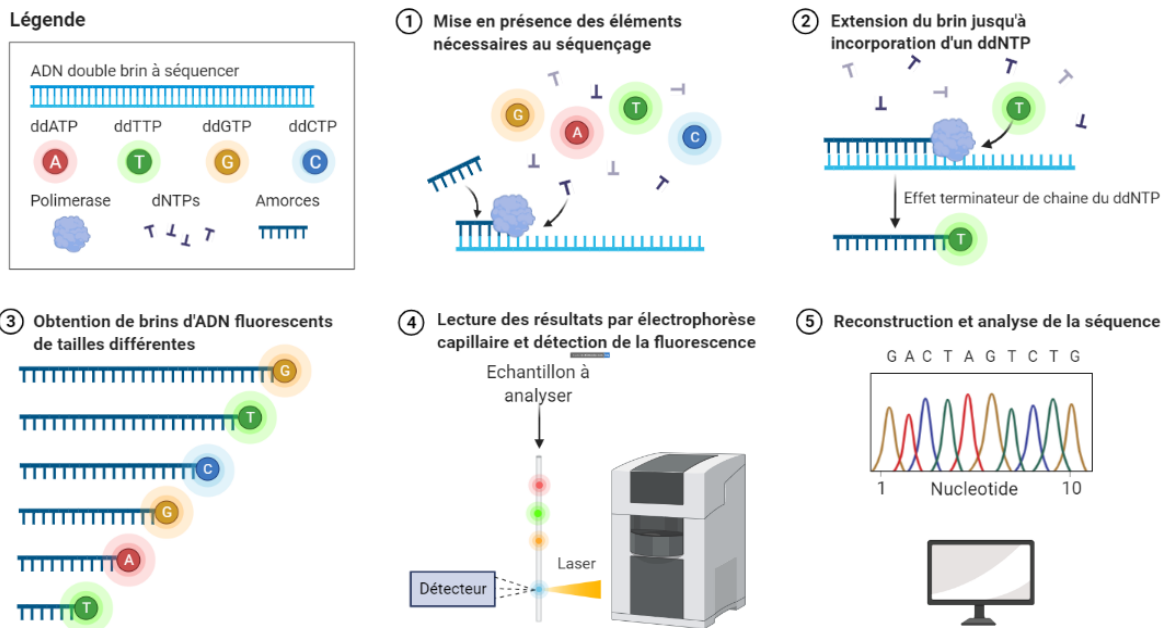


Figure 4 : Principe du séquençage Sanger

Le séquençage Sanger permet d'obtenir la séquence précise d'un fragment d'ADN. 1) Mise en contact d'ADN simple brin, d'ADN polymérase, d'amorces spécifiques, de dNTPs et de ddNTPs fluorescents avec effet terminateur de chaîne. 2) Extension des brins via la machinerie enzymatique disponible jusqu'à incorporation d'un ddNTP. 3) Obtention de différents fragments de tailles différentes. 4) Electrophorèse capillaire du résultat et lecture de la fluorescence. 5) Analyse et reconstitution de la séquence. Figure créée grâce à BioRender.com.

Le séquençage Sanger est une technique qui nécessite de connaître au moins une partie de la séquence de référence, que ce soit pour le design des amorces ou pour l'interprétation de la séquence obtenue.

L'automatisation actuelle du séquençage Sanger prend généralement en charge la génération de séquences allant jusqu'à 800-1000 pb avec une faible qualité de séquençage dans les 15 à 40 premiers pb en raison de la liaison à l'amorce. Prenant note de ces limitations, les logiciels d'analyse de séquences, commerciaux et non commerciaux, continuent d'évoluer pour aider l'utilisateur à évaluer et à rogner automatiquement les données de mauvaise qualité (42).

Le séquençage Sanger reste de nos jours la méthode de référence, même si de nombreuses études tendent à montrer que le séquençage de 2^{ème} génération peut s'en affranchir (43).

2) Séquençage haut débit (NGS) Illumina®

Le séquençage haut débit est une technique de séquençage de 2^{ème} génération. Il permet l'amplification rapide et simultanée d'un très grand nombre de séquences et ainsi, de séquencer le génome entier d'un individu très rapidement. Pour cela, il a été nécessaire de disposer d'une séquence de référence, obtenue dans le début des années 2000, au prix d'un effort commun international avec le « Human Genome Project » (44). Depuis, le génome de référence humain est régulièrement mis à jour en fonction des nouvelles connaissances, et est maintenu par un consortium international : le Genome Reference Consortium. La dernière version, datant de décembre 2013, est la GRCh38 (45).

Plusieurs techniques de séquençage haut débit ont été développées au cours des dernières années. Elles sont basées sur des principes différents (Tableau III). Dans le cadre de cette thèse, seule la technologie Illumina®, utilisée au CHRU de Nancy, sera développée. Le laboratoire BBMN du CHRU de Nancy possède un séquenceur NextSeq, et un MiSeq, tous deux utilisés en routine diagnostique.

Tableau III : Spécificité des techniques de NGS, d'après Stranneheim et Lundeberg (46)

Technique	Principe du séquençage	Méthode d'amplification	Principe de détection	Longueur d'un read	Débit d'analyse
Illumina®/Solexa	Séquençage par synthèse + terminateurs de chaîne fluorescents	Bridge PCR	Optique	100	Elevé
SOLID	Par ligation	Emulsion PCR	Optique	75	Elevé
Ion	Séquençage par synthèse + détection de protons	Emulsion PCR	Optique	200	Moyen
454	Pyroséquençage	Emulsion PCR	Electrique	700	Moyen
Complete Genomics	Séquençage par synthèse	Circularisation	Optique	35	Elevé

3.1) Versant technique

Le NGS peut se pratiquer sur de nombreuses matrices biologiques (sang, buvard, tumeurs, etc.). Suite au prélèvement, il se divise en deux grandes phases : tout d'abord le Wet Lab puis le Dry Lab (Figure 5).



Figure 5 : Etapes technique du NGS

3.1.1) Wet Lab

Extraction de l'ADN : Une fois le prélèvement effectué, il faut tout d'abord extraire l'ADN des cellules nucléées. L'extraction peut être manuelle ou automatisée, et le protocole dépend de la matrice.

Préparation de la librairie : Elle correspond à l'étape de sélection des régions ciblées par le séquençage. Elle aboutit in fine à un milieu enrichi en séquences cibles. Pour cela, l'ADN est d'abord fragmenté en séquences d'environ 300 pb soit par sonication soit par tagmentation. Cette dernière utilise des transposomes liés à des billes d'enrichissement. Les transposomes, en plus de fractionner l'ADN de manière relativement homogène, marquent chaque fragment avec des séquences adaptatrices (Figure 6). Les séquences marquées peuvent ensuite être amplifiées par PCR.

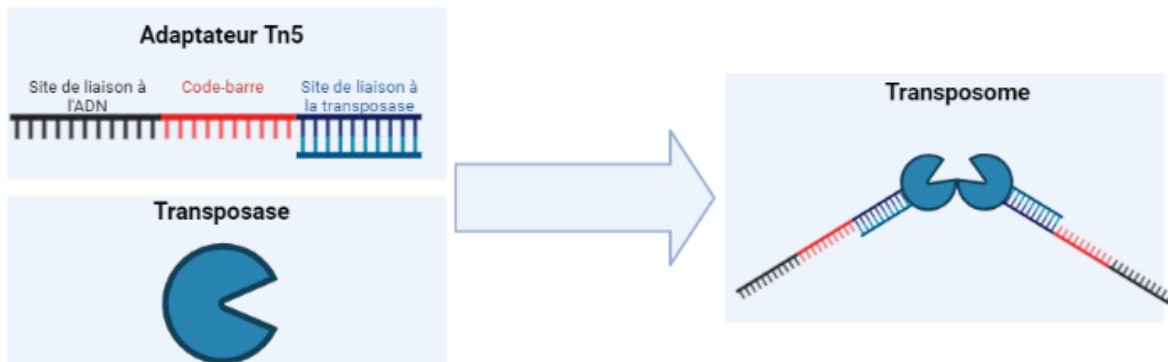


Figure 6 : Structure d'un transposome (47)

Transposome = Adaptateur Tn5 + Transposase. L'adaptateur est constitué de 3 parties : un site de liaison de transposase double brin de 19 pb, d'un code-barres simple brin et d'un site d'hybridation d'amorce de PCR. Les transposomes sont utilisés pour marquer et fragmenter l'ADN. Figure créée grâce à BioRender.com.

L'ADN de chaque patient est ensuite marqué par un index spécifique qui permettra par la suite de relier les reads au bon patient. Après cette étape d'indexage, les ADN des différents patients sont poolés mais restent identifiables grâce à ces « codes-barres ».

Une étape d'enrichissement pré-séquençage est ensuite nécessaire. Pour l'enrichissement par capture utilisé avec les technologies Illumina®, les régions génomiques d'intérêt sont capturées par hybridation sur des sondes spécifiques, avant d'être isolées grâce à des billes magnétiques, pour enfin être amplifiées. Par rapport au séquençage d'amplicons, cette technique peut cibler une plus grande quantité de contenu génique et prendre en charge un profilage plus complet de tous les types de variants (48). A ce stade, la librairie contient quasi exclusivement les séquences d'intérêt, mais en faible quantité. Avant le séquençage, une phase d'amplification clonale est nécessaire afin de pouvoir obtenir un signal de séquençage détectable.

Les kits de capture TruSight One et TruSight Expanded sont ceux utilisés au laboratoire de Biochimie du CHRU de Nancy. Ils permettent de cibler respectivement les régions exoniques et les jonctions introns/exons de 4813 gènes et de 6699 gènes (Annexe 3).

Tandis que les techniques SOLID, Ion et 454 utilisent des PCR à émulsion dans un mélange huile/eau ; Illumina® préconise l'amplification par bridge PCR (Figure 7). Dans ce cas, les librairies sont déposées sur un support solide (appelé flowcell) où est fixé un 1er type de sondes spécifiques des séquences retrouvées aux extrémités des fragments de la librairie. Chaque fragment peut alors se lier à la flowcell, avant de se replier pour accoler son extrémité libre spécifique du 2nd type de sonde avec ce dernier. Le fragment peut alors se répliquer grâce à une Taq polymérase, formant 2 fragments complémentaires. Cela permet la génération de clusters, au sein desquels un unique fragment se réplique jusqu'à saturation des sondes de la flowcell.

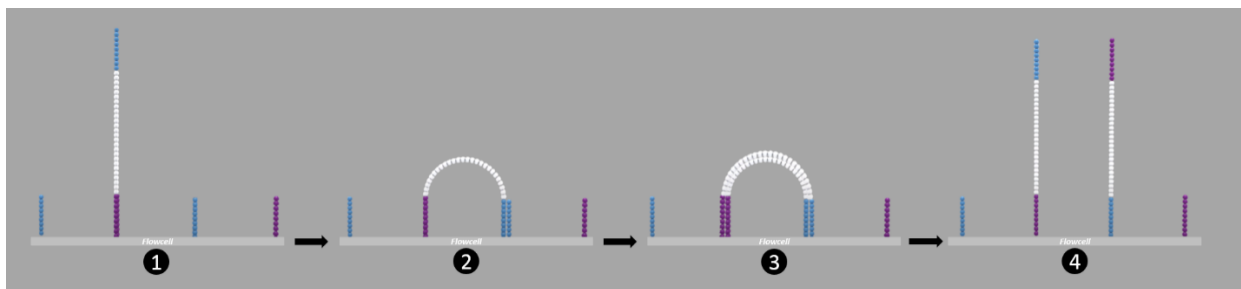


Figure 7 : Bridge PCR, d'après Illumina (49)

1) Liaison du fragment de librairie à la flowcell via des amorces spécifiques à des séquences préalablement ajoutées sur toute la librairie 2) Hybridation de l'extrémité libre sur un 2nd type d'amorce spécifique fixé à la flowcell 3) Réplication grâce à une Taq polymérase 4) Formation de 2 fragments complémentaires liés à la flowcell par une extrémité.

A l'issue de cette amplification, tous les fragments de la librairie sont théoriquement présents en quantité suffisante pour être détecté lors du séquençage.

Séquençage (production de reads) : La technologie Illumina® utilise le principe du séquençage par synthèse grâce à des nucléotides cycliques modifiés à effet terminateur de chaîne

réversible. Comme pour un séquençage Sanger, les nucléotides modifiés sont marqués par une molécule fluorescente et émettent un signal spécifique de la nature du nucléotide lors de leur incorporation dans la séquence. Mais contrairement au séquençage de 1ère génération, il est possible de réverser le blocage du nucléotide suivant grâce à un clivage qui élimine le groupe terminaison/colorant fluorescent (50). Les nucléotides modifiés sont donc incorporés un à un, de manière complémentaire à la séquence d'origine, puis la fluorescence est déterminée grâce à un lecteur optique. Après la lecture, le fluorochrome et le groupement qui bloque l'élongation sont clivés, puis éliminés par lavage. Un autre nucléotide modifié peut alors s'incorporer à la suite, et le schéma se répète (Figure 8).

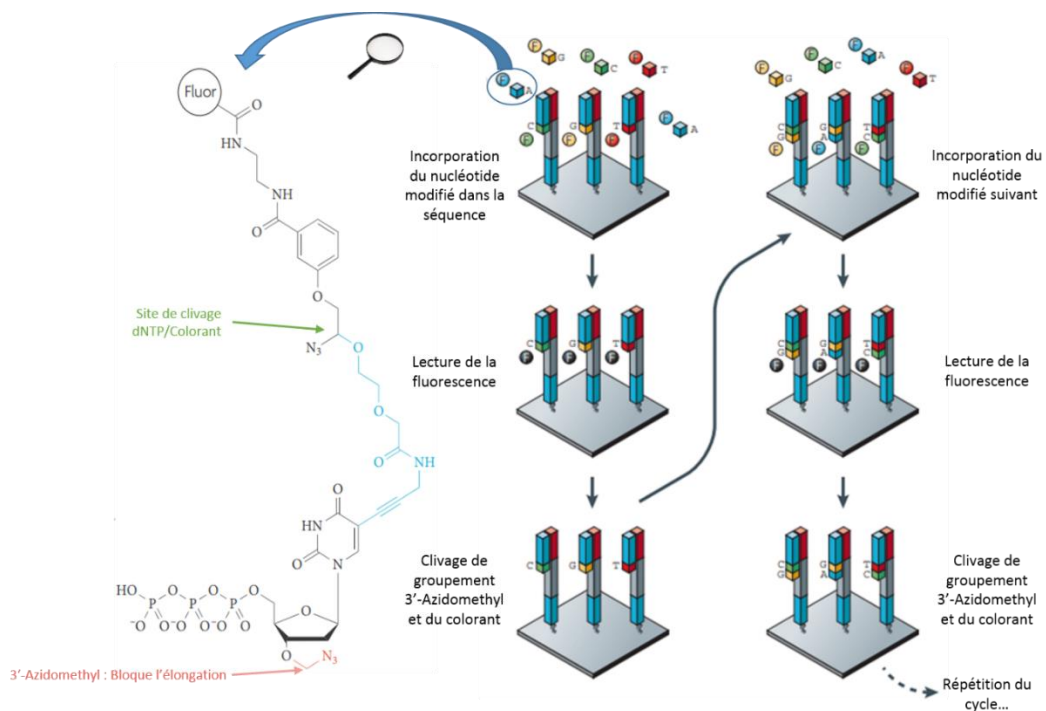


Figure 8 : Principe du Séquençage par synthèse avec terminateurs de chaîne fluorescents, d'après Metzker (51)

Ainsi, à chaque incorporation de nucléotide modifié, un signal lumineux est lu, et est converti en séquence dans un fichier FASTQ : on parle de base calling. En plus des séquences, ce fichier contient des scores de qualité et les informations nécessaires au démultiplexage.

Cette étape est la dernière étape de technique manuelle, avant le Dry Lab et l'analyse bio-informatique.

3.1.2) Dry Lab

Alignement : À la suite du Base Calling, les séquences des read peuvent être reliées au bon patient grâce à leur index. L'étape de l'alignement consiste ensuite à placer chaque séquence d'environ 100pb au bon endroit sur le génome de référence, et ainsi reconstituer la séquence complète du patient. Pour cela, différents algorithmes d'alignements ont été développés (algorithmes naïfs ; Burrows Wheeler ; algorithme de Needleman & Wanch ; ...). A l'issue de cette étape, un fichier SAM est obtenu, avant d'être compressé en fichier BAM. Dès l'alignement, des scores de qualité du séquençage sont disponibles, ainsi que la couverture et la profondeur pour chaque base.

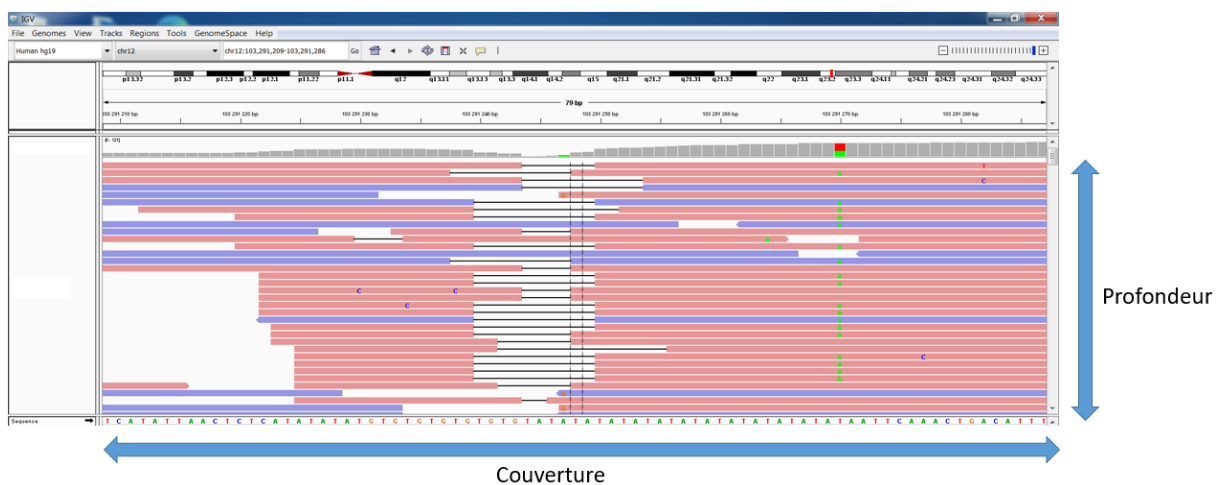


Figure 9 : Exemple de profondeur et couverture d'un séquençage en 12q23.2 (sur IGV)

Dans cet exemple : toute la séquence est couverte avec une profondeur minimum de 3.

Au CHRU de Nancy, les séquences obtenues sont alignées et analysées en comparaison avec la séquence de référence du génome humain version GRCh37 à l'aide de BWA (0.7.9a) et SAMTools (0.1.18). Les aspects méthodologiques relatifs à la procédure de séquençage et d'alignement sont rapportés dans le chapitre Matériel et Méthodes de la publication (cf. ci-dessous).

Annotation et appel des variants : La liste des variants détectés est alors établie. Le fichier obtenu est alors un VCF, et permet de décrire les variants de manière standardisée, avec 8 caractéristiques :

- Chromosome : Le chromosome sur lequel cette variation est rapportée.
- Position : La position à laquelle la variation est rapportée.
- Identifiant externe : A partir d'une liste d'identifiants de bases de données externes (dbSNP, Clinvar, Cosmic, etc.)
- Allèles de référence rapportés à cette position.
- Allèles alternatifs détectés dans nos échantillons.

- Score de qualité de la variation rapportée
- Filtres : Une liste de filtres que la variation n'a pas passée. Si tous les filtres sont passés, ce champ prend la valeur « PASS ». Si aucun filtre n'a été appliqué le champ prend la valeur « . ».
- Informations : Description de l'expérience ayant conduit à la génération de ce fichier.

Dans notre cas, l'annotation des variants est effectuée avec la suite d'outils GATK (v1.6-23).

Interprétation, Stockage, et Rendu des résultats : En moyenne, sur les 603 NGS effectués à Nancy entre 2016 et 2020, plus de 10 000 variants sont appelés par patient. L'enjeu de cette technique est de faire le tri dans la masse des données générées.

Les variants relevant cliniquement sont sélectionnés par un biologiste en fonction de l'indication du test et du contexte clinico-biologique. Plusieurs approches d'analyse bio-informatiques sont utilisées, incluant l'analyse par panels de gènes et l'utilisation d'algorithmes basés sur les ontologies HPO (*Human Phenotype Ontology*) que présente le patient. La base de données HPO (52) répertorie de façon systématisée les informations relatives aux ontologies, phénotypes, pathologies et gènes associés (Figure 10).

Ontologie	Phénotype	Gène
HP:0002160 Hyperhomocystinemia	OMIM:236200 Homocystinuria Due To Cystathionine Beta-synthase Deficiency	CBS
	OMIM:236250 Homocystinuria Due To Deficiency Of N(5,10)-methylenetetrahydrofolatereductase Activity	MTHFR
	ORPHA:395 Homocystinuria Due To Methylene Tetrahydrofolate Reductase Deficiency	MTHFR
	OMIM:236270 Homocystinuria-megaloblastic Anemia, Cbl E Type	MTRR
	OMIM:250940 Homocystinuria-megaloblastic Anemia, Cblg Complementation Type	MTR
	ORPHA:542643 Livedoid Vasculopathy	/
	OMIM:601775 Megaloblastic Anemia, Folate-responsive	SLC19A1
	ORPHA:2169 Methylcobalamin Deficiency Type Cble	MTRR
	ORPHA:79284 Methylmalonic Acidemia With Homocystinuria Type Cblf	LMBRD1
	ORPHA:79282 Methylmalonic Acidemia With Homocystinuria, Type Cblc	MMACHC
	OMIM:277400 Methylmalonic Aciduria And Homocystinuria, Cblc Type	PRDX1
	OMIM:277410 Methylmalonic Aciduria And Homocystinuria, Cblf Type	MMADHC
	OMIM:277380 Methylmalonic Aciduria And Homocystinuria, Cblf Type	LMBRD1
	OMIM:614857 Methylmalonic Aciduria And Homocystinuria, Cblj Type	ABCD4
	ORPHA:88618 S-adenosylhomocysteine Hydrolase Deficiency	AHIC

11 Gènes impliqués dans l'hyperhomocystéinémie : ABCD4, AHIC, CBS, LMBRD1, MMACHC, MMADHC, MTHFR, MTR, MTRR, PRDX1, SLC19A1

Figure 10 : Exemple de l'arborescence HPO pour l'HP:0002160, et gènes impliqués
 Pour l'ontologie « hyperhomocystéinémie » (HP:0002160), 15 phénotypes OMIM sont décrits en lien avec 11 gènes différents.

Les variants sont sélectionnés par le biologiste en fonction de la symptomatologie et des antécédents personnels et familiaux du patient. L'analyse des résultats du NGS est donc personnalisée et est fonction du patient et de son phénotype. La pathogénicité des variants retenus est ensuite déterminée selon la classification ACMG grâce à différentes bases de données en ligne (Varsome, Franklin, Clinvar, etc.). Sur le compte rendu, les variants sont décrits selon les recommandations HGVS (le A de l'ATG initiateur de la traduction, a été choisi comme n°1 de la référence) et selon les règles de bonnes pratiques de la HAS (19).

L'annotation des transcrits de gènes est basée sur RefSeq Genes 105v2, NCBI. Le résultat est finalement stocké de manière sécurisée et confidentielle par le laboratoire pour 30 ans comme le précise le code de la santé publique (53).

Les technologies de NGS ont considérablement amélioré les chances d'obtenir un diagnostic génétique, car elles sont génétiquement et phénotypiquement agnostiques dans leur approche. Le séquençage simultané de chaque gène du génome détecte la grande majorité des variations importantes présentes dans le génome d'un individu et permet potentiellement de poser des diagnostics sur une gamme large de phénotypes, même pour les affections rares et ultra-rares, y compris celles avec lesquelles le clinicien peut ne pas être familier (54). Dans ce cadre, plusieurs types d'approches ont été développées.

3.2) Approche ciblée, par panel, WES ou WGS

La multitude de données générées par le NGS rend son interprétation difficile. Pour chaque patient, plusieurs milliers de variants peuvent être décelés. De même, pour une indication précise, de nombreux gènes peuvent être impliqués (Tableau IV).

Tableau IV : Exemple de pathologies et nombre de gènes impliqués d'après Lamoril et Bogard (55)

Famille de pathologies	Exemple de pathologies	Nombre de gènes impliqués
Cancers	Cancers héréditaires (sein, syndrome de Lynch, HNPCC)	2-50
Maladies cardiaques	Cardiomyopathies	50-70
Troubles métaboliques	Troubles du métabolisme énergétique cellulaire	656
	Anomalies congénitales de la glycosylation	23-28
	Maladies mitochondriales	37-450
Divers	RASopathies	10
	Troubles de la croissance osseuse	12

Ainsi, plusieurs approches ont été développées et sont utilisées dans les différents centres spécialisés en fonction des besoins diagnostiques et les ressources disponibles dans les centres spécialisés. Le choix de l'approche dépend des coûts de l'analyse, de l'indication du test, du nombre de gènes potentiellement responsables, et des limites interprétatives de la multitude de données générées par l'analyse (Tableau V).

- Approche ciblée : elle ne permet l'analyse que d'un nombre restreint de gènes ou de loci. Plusieurs techniques sont disponibles et incluent le séquençage complet ou partiel d'un gène par Sanger, l'analyse des courbes de fusion par HRM, le comptage de répétitions de

triplets par PCR quantitative et la MLPA. Ce sont des techniques limitées de par leur faible couverture, mais qui sont souvent suffisantes dans le cadre de syndromes bien spécifiques, comme Prader-Willi, Angelman, X Fragile (56), etc. Elles permettent également de pallier les limites du NGS, peu performant dans l'analyse des séquences répétées, homologues ou riches en GC, et des CNV (57).

- Approche par panel : Pour de nombreux troubles, la variabilité clinique et l'hétérogénéité du locus génétique sont suffisamment importantes pour qu'une approche par panel de gènes soit appropriée pour un diagnostic moléculaire efficace et rapide. Les panels utilisés pour une indication ne sont pas forcément identiques d'un laboratoire à l'autre. Selon les étiologies et les habitudes diagnostiques du laboratoire, un panel peut couvrir moins d'une dizaine de gènes, à plusieurs milliers. Il existe des panels commerciaux (TruSight Cardio, TruSight One, TruSight Inherited Disease, etc.) ou des panels « faits maison », pour des indications particulières. Par rapport aux approches ciblées, les approches par panel ont une meilleure sensibilité et sont utiles dans le diagnostic différentiel dans le cas où l'altération de plusieurs gènes différents induit des présentations cliniques similaires (58). Néanmoins, plus le nombre de gènes séquencés augmente, plus le risque de découverte fortuite augmente.
- *Whole Exome Sequencing* : Cette approche ressemble à celle par panel, mais dans ce cas, la quasi-totalité de l'exome du patient est séquencée. La quantité de données générées est ainsi plus importante : augmentant la couverture du génome total et ainsi la probabilité que le gène responsable de la pathologie soit inclus dans les données ; mais rendant l'analyse plus complexe, avec des risques de découvertes fortuites encore augmentés, et des risques de faux négatifs (59,60). De plus, avec les technologies actuelles, quelle que soit le type de capture utilisé, il n'est statistiquement pas possible de capturer toutes les séquences codantes. En moyenne, 90 à 95% des exons peuvent être capturés, laissant des trous dans la couverture qui diminuent la sensibilité du test (55). Ainsi, malgré la meilleure couverture prodiguée par le WES, l'approche par panel reste utilisée en raison des faibles coûts de séquençage, des délais d'exécution courts, et d'un taux de découvertes fortuites qui reste correct, en ne manquant qu'environ 3-10 % des mutations détectables par l'exome entier séquençage (WES) (61,62).
- *Whole Genome Sequencing* : cette approche permet le séquençage complet du génome. Théoriquement le WGS est la meilleure approche car elle produit les données les plus complètes sur le génome d'un individu. Il permet de plus une couverture plus uniforme de l'exome (du fait de sa sensibilité moindre aux régions riches en GC (63)), permet la

détection de variants introniques et facilite l'appel des variants de structure. Mais, même compte tenu des récentes évolutions techniques, les coûts et temps d'analyse limitent encore sérieusement la mise en œuvre du WES dans le cadre des diagnostics de routine (64).

Tableau V : Comparaison des différentes approches de séquençage (64)

Approche	Ciblée	Panel	WES	WGS
Indications (Exemple)	<p>Hétérogénéité de locus minimale (<i>CFTR</i>)</p> <p>Clinique spécifique orientant vers un gène spécifique (<i>PAH</i> pour PCU)</p> <p>Répétition de triplets (<i>FMR1</i> pour X Fragile)</p>	<p>Diagnostic différentiel (Cardiomyopathie)</p> <p>Pathologies avec manifestations cliniques communes (ex : Epilepsie)</p> <p>Maladies associées à des gènes d'une voie ou d'une structure commune (RASopathies)</p>	<p>Hétérogénéité clinique importante (Autisme)</p> <p>Coexistence de plusieurs phénotypes non liés</p> <p>Clinique non spécifique, sans orientation diagnostique (Syndrome de Kabuki)</p> <p>Le phénotype est indistinct et la véritable cause sous-jacente n'est pas facile à identifier (Diarrhées congénitales)</p>	<p>Mêmes indications que WES</p> <p>Présentation clinique très évocatrice d'un trouble avec WES négatif (Suspicion de variant intronique)</p>
Couverture de la région ciblée	100%	100% (Si complété par Sanger)	97,5%	>97,5%
Profondeur	Non applicable	>500X	>50-100X	>30X
Découvertes fortuites et classe III	⊖	⊕	⊕⊕	⊕⊕⊕
Détection CNV	Oui	Généralement non	Si > 3 exons	Oui
Détection variants introniques (>30pb des sites donneurs et accepteurs)	Non	Non	Non	Oui
Prix	⊖⊖	⊖ (Selon les panels)	⊕	⊕⊕⊕ (3 fois plus que WES)

3.3) Principales limitations du NGS

Malgré les énormes quantités de données produites, les différentes approches disponibles et les performances des séquenceurs, le séquençage de 2ème génération présente des limitations.

- **Reads courts** : Quel que soit la technologie utilisée, elle aboutit à la production de reads relativement courts (moins de 1 kb). Ces faibles tailles ne permettent pas d'apprécier correctement les éléments du génome plus longs et complexes, comme les régions fortement répétées (dont les répétitions de triplets (65)) et les variations de structure. L'alignement des reads des régions répétées est en effet complexe et souvent incorrect en l'absence de point de repère sûr, d'autant que ces éléments représentent la moitié du génome humain. De même, en l'absence de chevauchement read-réarrangement de structure, ces derniers sont invisibles au séquençage (66). Cela pose problème, car les variants de structure sont fréquemment pathogènes. Par exemple, les translocations équilibrées peuvent être responsables de pathologie du point de cassure (67).
De plus, la faible longueur des reads ne permet pas de caractériser le caractère « cis » ou « trans » de 2 variants trop éloignés. Ainsi, en cas de présence de 2 variants pathogènes sur le même gène, il n'est pas forcément possible de conclure si les variants sont effectivement sur des allèles différents ; et l'interprétation de ceux-ci nécessite d'autres explorations complémentaires.
- **Régions riches en GC** : Ces régions sont difficiles à amplifier du fait de la possible formation de structure secondaires. Ainsi, une amplification clonale par PCR est nécessaire avant le séquençage pour augmenter le signal (68). Cette étape d'amplification supplémentaire est donc une source de biais supplémentaire.
- **Variation du nombre de copies (CNV, Copy number variation)** : une des limitations majeures de l'utilisation du NGS en diagnostic est sa capacité à détecter les CNV. Les techniques de référence pour la détection de ces anomalies de nombre restent la MLPA et les puces à ADN en microréseaux. De plus en plus d'outils se développent pour permettre d'améliorer la détectabilité des CNV par NGS. La plupart d'entre eux peuvent appeler de manière fiable des CNV de grande taille (dans l'ordre des mégabases) mais présentent des performances moyennes pour les petits CNV affectant seulement un ou quelques exons (69).

Ainsi, même si le séquençage haut débit a grandement facilité l'accès à la médecine moléculaire, d'autres technologies de séquençage ont dû être développées pour pallier ces limites techniques.

3) Séquençages de 3^{ème} génération

4.1) *Single-Molecule Real-Time Sequencing (Pacific Biosciences)*

Il s'agit de la 1^{ère} technique de séquençage de 3^{ème} génération. Le séquençage SMRT a introduit la capacité d'acquisition de séquences en temps réel pour des longueurs de lecture > 1 kb en utilisant le séquençage par synthèse et détection optique. Cette technologie utilise une matrice d'ADN circulaire simple brin formée à partir d'un fragment linéaire double brin, après ligature d'adaptateurs en épingle à cheveux sur les deux extrémités. Celles-ci sont ensuite dispersées individuellement dans des « nanopuits » avec au fond, une ADN polymérase qui effectue une amplification continue de la matrice circulaire. Les nucléotides incorporés, étant marqués, libèrent un signal fluorescent lors de leur intégration à la séquence. Le signal est ensuite enregistré par un lecteur optique en temps réel. En amplifiant continuellement la matrice circulaire, la séquence de la matrice et des brins complémentaires est capturée plusieurs fois, augmentant ainsi la profondeur de lecture et la précision de l'appel de base. L'amplification est continue jusqu'à expiration de la durée de vie fonctionnelle de la polymérase : ainsi, plus le read est long, plus la profondeur sera faible (70).

Cette technique ne permet que des lectures relativement courtes (par rapport aux autres techniques de séquençage de 3^{ème} génération) avec une très grande précision, tandis qu'une approche alternative de modèle linéaire permet des longueurs de lecture plus longues mais une précision inférieure (70).

4.2) *Nanopore Sequencing (Oxford Nanopore Technologies)*

Cette technologie utilise un séquenceur portable qui met à profit une grille de nanopores biologiques intégrés à une membrane. MinION® identifie les bases d'ADN en mesurant les changements de conductivité électrique générés lorsque les brins d'ADN traversent un pore biologique. Les molécules d'ADN sont préparées en ajoutant un adaptateur en épingle à cheveux à une extrémité de la molécule double brin. Puis, grâce à une hélicase et une protéine motrice attachée à la matrice, l'ADN (alors simple brin) peut traverser le nanopore. Les bases passent alors l'une après l'autre devant un capteur, qui enregistre les changements de courant électrique caractéristiques de chaque base. L'analyse de ces courants permet de déterminer la séquence d'ADN (71). Ainsi, aucune étape d'amplification n'est nécessaire, et théoriquement, la maîtrise de cette technique permettrait de séquencer chaque brin d'ADN en un seul read. De plus, comme le processus de séquençage utilise très peu de réactifs épuisables, l'analyse peut théoriquement se poursuivre efficacement jusqu'à ce qu'un résultat satisfaisant soit obtenu. Actuellement, le séquençage Nanopore a des taux d'erreurs

beaucoup trop élevés (~15%) pour être utilisé en routine diagnostique, mais la longueur des reads lus augmente rapidement. Des reads de plus 2 Mb ont déjà été enregistrés (72).

4.3) Synthetic Long Reads (Illumina and 10X Genomics)

La technologie SLR s'appuie sur un système de code-barres pour associer des fragments qui sont séquencés sur des séquenceurs courts. Actuellement, deux systèmes existent : la plateforme de séquençage Illumina SLR et le système à base d'émulsion 10X Genomics.

- Avec le système Illumina, l'ADN génomique est cisailé en fragments de 8 à 10 kb, suivi d'une ligature d'adaptateur « fragment long », qui peut être utilisée pour désigner les extrémités des contigs lors de l'assemblage de lecture courte en aval. Ces gros fragments sont ensuite répartis dans une plaque de microtitrage (plus de 3000 fragments par puits) et subissent un cisaillement supplémentaire et un ajout d'adaptateur par le biais d'un processus de tagmentation. Chaque puit contient un seul code-barres. L'ADN est ensuite regroupé et soumis à un séquençage Illumina classique suivi d'un assemblage local pour reconstruire les longs fragments d'origine (73).
- Avec le séquençage à base d'émulsion 10X Genomics, des fragments d'ADN naturels allant jusqu'à 100 kb sont mélangés dans des micelles dans une émulsion (« GEM ») avec des billes de gel contenant des séquences d'adaptateur et de code-barres. Dans chaque GEM, la bille de gel se dissout et des petits fragments d'ADN sont amplifiés à partir des gros fragments originaux, chacun avec un code-barres identifiant le GEM source. Les fragments sont ensuite regroupés avant de subir un séquençage classique Illumina®. Les lectures obtenues sont assemblées pour former une série de fragments ancrés sur une étendue de 50 kb (73).

4) Classification des variants

Les conséquences phénotypiques des différents variants ne sont pas équivalentes en termes de pathogénicité.

Face au nombre quasi infini de variations génomiques possibles, il a fallu mettre en place des règles permettant de classer et d'interpréter les variants. Celles-ci, de nos jours largement adoptées par la communauté, ont été décrites en 2015 par *l'American College of Medical Genetics and Genomics (ACMG)* (74). Les variants sont ainsi divisés en 5 classes : Bénin (classe 1, NGS diag), Probablement bénin (classe 2, NGS diag), Signification incertaine (classe 3, NGS diag), Probablement pathogène (classe 4, NGS diag) et Pathogène (classe 5, NGS diag) (74,75).

Les variants sont évalués selon des données épidémiologiques, structurales, fonctionnelles, de ségrégation, bibliographiques, cliniques et généalogiques. D'autres critères sont pris en compte, comme les variants associés et la prédiction sur la fonction protéique (via des outils de prédiction in silico). Puis chaque argument est étiqueté en fonction de son caractère plus ou moins pathogène (PVS ; PS ; PM ; PP ; BA ; BS ; BP) (74). Enfin, la classe du variant est attribuée selon les éléments décrits sur la figure 11.

Arguments en faveur de la pathogénicité du variant :				
Argument Très fort		Exemple : PVS1		
Fort		Exemples : PS4, PM3_fort		
Moyen		Exemples : PM2, PP1_moyen		
Faible		Exemples : PP3, PM3_faible, PS4_faible		
Arguments en faveur du caractère bénin du variant :				
Argument suffisant		Exemple : BA1		
Fort		Exemples : BS1, BS3		
Faible		Exemples : BP1, BP2		

<input type="checkbox"/> PATHOGENE	<input type="checkbox"/> PROBABLEMENT PATHOGENE	<input type="checkbox"/> DE SIGNIFICATION INCERTAINE	<input type="checkbox"/> PROBABLEMENT BENIN	<input type="checkbox"/> BENIN
<input type="checkbox"/> PVS1 + PS <input type="checkbox"/> PVS1 + 2PM <input type="checkbox"/> PVS1 + PM + PP <input type="checkbox"/> PVS1 + 2PP <input type="checkbox"/> 2PS <input type="checkbox"/> PS + 3PM <input type="checkbox"/> PS + 2PM + 2PP <input type="checkbox"/> PS + PM + 4PP	<input type="checkbox"/> PVS1 + PM <input type="checkbox"/> PS + PM <input type="checkbox"/> PS + 2PP <input type="checkbox"/> 3PM <input type="checkbox"/> 2PM + 2PP <input type="checkbox"/> 1PM + 4PP	<input type="checkbox"/> Critères insuffisants <input type="checkbox"/> Critères contradictoires	<input type="checkbox"/> BS + BP <input type="checkbox"/> 2BP	<input type="checkbox"/> BA1 <input type="checkbox"/> 2BS

Figure 11 : Homogénéisation de la classification des variants (75)

Ces éléments permettent de classer les variants de manière standardisée. Cependant, les variants inconnus ou localisés dans des gènes sans lien clairement identifié avec une pathologie, seront souvent classés comme ayant une signification incertaine (classe III). Dans ce cas, pour permettre d'étudier l'impact sur la protéine, il faudra réaliser des études complémentaires (études fonctionnelles, études de ségrégation, ...).

Enfin, l'interprétation finale des variants reste à l'appréciation du biologiste en charge du dossier, éventuellement à la suite d'une réunion pluridisciplinaire, et en fonction de l'état actuel des connaissances et des informations disponibles. Suivant l'évolution de celles-ci, un variant peut être amené à être reclassé, et son implication dans la pathogénicité d'un patient, réévaluée. (76)

II) Problématique de l'étude

A) CONTEXTE DE L'ETUDE

Le CHRU de Nancy est l'un des 8 CRMR du métabolisme en France. Le Département de Médecine moléculaire du Laboratoire de BBMN fonctionne en collaboration avec différents services cliniques, incluant : les centres de référence des maladies héréditaires du métabolisme et de compétence en médecine vasculaire ; et les différents pôles cliniques du CHRU. Il permet de proposer une démarche d'expertise et de diagnostic en médecine moléculaire au niveau local, régional, national et international sur les thématiques reconnues du Laboratoire (Anomalies héréditaires du métabolisme, en particulier le métabolisme des monocarbones).

Une augmentation constante de l'activité de Médecine Moléculaire du Laboratoire de BBMN au cours des 5 dernières années est observée du fait de l'intégration en clinique de routine du diagnostic moléculaire dans les démarches diagnostiques des pathologies présentant une forte suspicion d'étiologie génétique. Cette croissance permet au laboratoire de conserver les données clinico-biologiques d'un nombre important de patients ayant bénéficié d'un diagnostic moléculaire grâce au séquençage haut débit.

B) DESCRIPTION DE LA COHORTE

Entre janvier 2016 et décembre 2019, 603 séquençages d'exome clinique ont été réalisés au laboratoire BBMN du CHRU de Nancy. 426 concernent des patients « index » présentant un phénotype de maladie mendélienne. Les CES restant ont été réalisés chez des apparentés de ces cas index. Cette étude rétrospective observationnelle décrit les résultats de 603 CES consécutifs, réalisés au CHRU de Nancy, en tant que test génétique de 1ère ligne dans le diagnostic de maladie rare.

Les prescriptions (dans le cadre du diagnostic phénotypique et génétique) ont été réalisées par un clinicien disposant des connaissances et compétences nécessaires au diagnostic.

Les données administratives et cliniques des patients ont été récupérées par revue systématique des dossiers médicaux électroniques disponibles (sur le logiciel DxCare® : Dedalus France, Le Plessis Robinson, France) pour les patients suivis au CHRU de Nancy ou grâce aux informations renseignées sur les prescriptions médicales pour les patients hors CHRU de Nancy.

Les différentes indications des prescriptions recueillies ont ensuite été classées en 14 familles, elles même divisées en « sous famille », lorsque cela était possible et pertinent en fonction des effectifs. Le détail des indications est disponible sur la figure 12.

C) OBJECTIFS DE L'ETUDE

Cette étude rapporte les résultats d'expérience de plus de 600 séquençages d'exomes cliniques effectués en situation réelle dans un centre de référence en tant que test génétique de 1ère ligne, chez des patients adultes et pédiatriques suspectés d'être atteints de troubles mendéliens.

L'objectif principal est de déterminer l'efficacité du CES en termes de rendement diagnostique.

Les objectifs secondaires sont d'identifier des facteurs prédictifs de l'utilité clinique du CES pour établir un diagnostic moléculaire, confirmer le diagnostic suspecté et orienter la stratégie thérapeutique.

Nombre de NGS réalisés par indication

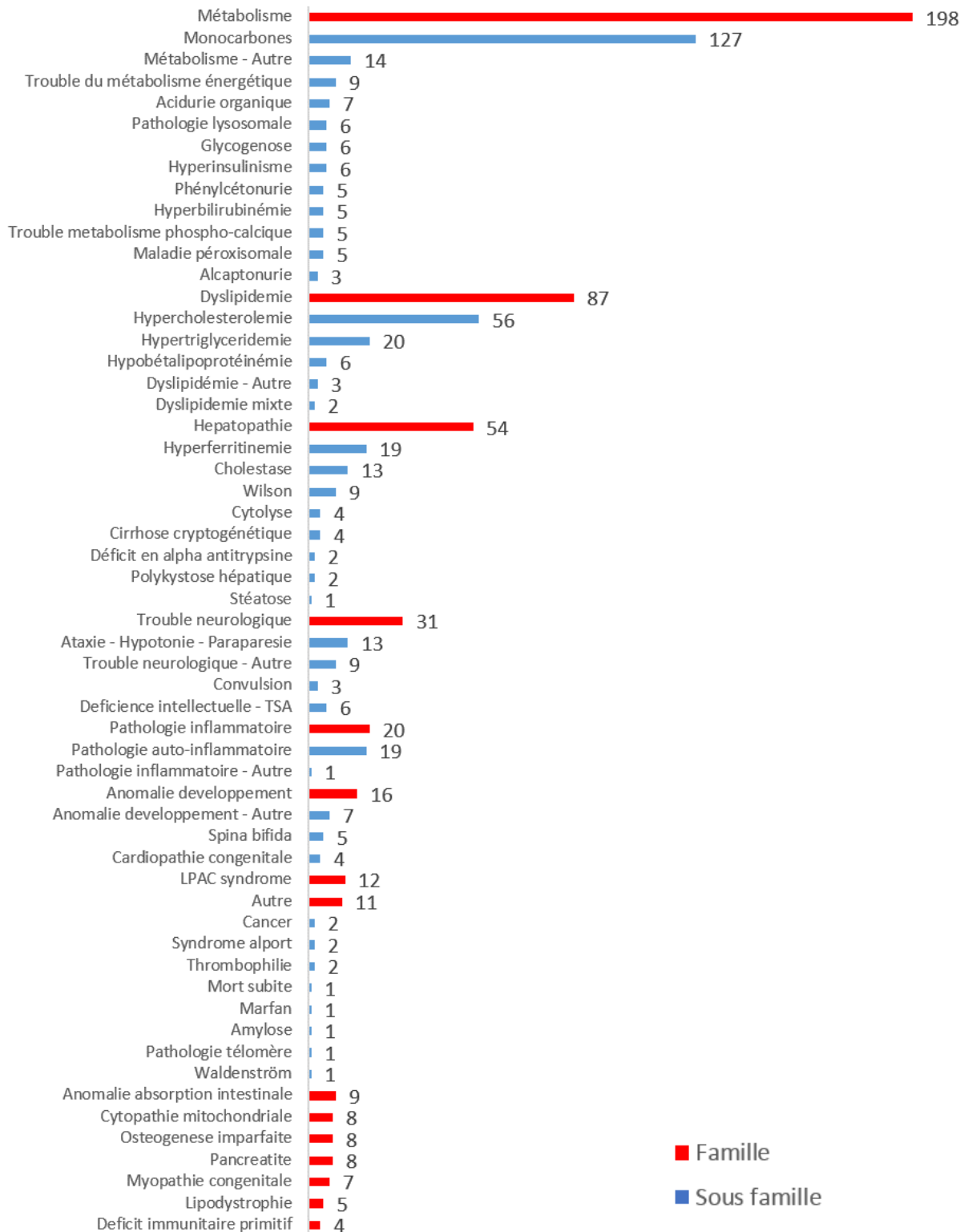


Figure 12 : Répartition des indications classés par famille de maladie selon les informations disponibles en pré-CES

III) « Predictors of the Clinical Utility of Clinical Exome Sequencing as a First-Tier Genetic Test in Patients with Mendelian Phenotypes: Results from a Referral Center Study in 603 Consecutive Cases »

Tom Alix, PharmD, MScs¹; Céline Chéry, PhD¹⁻³; Thomas Josse, MSc¹;
Jean-Pierre Bronowicki, MD, PhD^{2,4}; François Feillet, MD, PhD^{2,3,5,6};
Rosa-Maria Guéant-Rodriguez, MD, PhD¹⁻³; Farès Namour, MD, PhD¹⁻³;
Jean-Louis Guéant, MD, DSc, AGAF¹⁻³; and Abderrahim Oussalah, MD, PhD¹⁻³

1 Department of Molecular Medicine, Division of Biochemistry, Molecular Biology, Nutrition, and Metabolism, University Hospital of Nancy, F-54000 Nancy, France.

2 University of Lorraine, INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), Faculty of Medicine of Nancy, F-54000 Nancy, France.

3 Reference Center for Inborn Errors of Metabolism (ORPHA67872), University Hospital of Nancy, F-54000 Nancy, France.

4 Department of Gastroenterology and Liver Diseases, University Hospital of Nancy, F-54000 Nancy, France.

5 Department of Pediatrics, University Hospital of Nancy, F-54000 Nancy, France.

Corresponding Authors:

Prof. Abderrahim Oussalah, MD, PhD

Department of Molecular Medicine, Division of Biochemistry, Molecular Biology, and Nutrition, University Hospital of Nancy, and INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), Faculty of Medicine of Nancy, 9 Avenue de la Forêt 25 de Haye, F-54000, Nancy, France.

Telephone: +33 3 83 15 36 29

E-mail address: abderrahim.oussalah@univ-lorraine.fr

&

Prof. Jean-Louis Guéant, MD, DSc, AGAF

Department of Molecular Medicine, Division of Biochemistry, Molecular Biology, and Nutrition, University Hospital of Nancy, and INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), Faculty of Medicine of Nancy, 9 Avenue de la Forêt de Haye, F-54000, Nancy, France.

Telephone: +33 3 83 15 44 83

E-mail address: jean-louis.gueant@univ-lorraine.fr

Word count: 3,705

ABSTRACT

Importance: While several studies have focused on the diagnostic yield of clinical exome sequencing (CES), no study assessed the predictors of CES utility among patients with various Mendelian phenotypes, such as those observed in the setting of a molecular medicine department.

Objective: To assess the effectiveness of CES as a first-level genetic test for making a molecular diagnosis in patients with a Mendelian phenotype and to seek independent predictors of the clinical utility of CES.

Design: Retrospective observational study on consecutive patients.

Setting: Referral center in the North-East of France within the “French National Plan for Rare Diseases” framework.

Participants: Adult and pediatric patients with Mendelian phenotypes.

Exposure: Clinical exome sequencing.

Main Outcome(s) and Measure(s): The study outcomes were the discovery of at least one pathogenic or likely pathogenic variant, the “Confirmation of the suspected clinical diagnosis by CES”, and “CES-guided treatment strategy”.

Results: Between January 2016 and December 2019, 603 patients (426 probands and 177 siblings) underwent CES at the Department of Molecular Medicine of the University Hospital of Nancy. The median age of the probands was 34 years (IQR, 12–48) and the proportion of males was 46.9% (200/426). Adults and pediatric populations represented 64.8% (276/426) and 35.2% (150/426), respectively. The median test-to-report time was 5.6 months (IQR, 4.1–7.2). The CES revealed 203 pathogenic or likely pathogenic (P/LP) variants in 160 patients, corresponding to a diagnostic yield of 37.6% (160/426). This figure was 60.8% (259/426) when VUS variants were included in the diagnostic yield analysis. Independent predictors of CES utility were criteria strongly suggestive of an extreme phenotype, including pediatric presentations and patients’ phenotypes associated with an increased risk of a priori probability of a monogenic disorder, the inclusion of at least one family member in addition to the proband, and a CES prescription performed by an expert in the field of rare genetic disorders.

Conclusions and Relevance: Using a large dataset of consecutive patients with various Mendelian phenotypes referred for CES as a first-tier genetic test, we report a diagnostic yield of ~40% and several independent predictors of CES utility that could improve CES diagnostic efficiency.

Fundings: INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure.

Keywords: clinical exome sequencing; diagnostic yield; predictors of clinical utility; Mendelian phenotype; reference center; case-consecutive series.

Key Points

Question: Are there any factors that could predict the effectiveness of clinical exome sequencing (CES) in patients with Mendelian phenotype?

Findings: In this retrospective observational study on consecutive adult and pediatric patients with Mendelian phenotypes, independent predictors of CES utility were criteria strongly suggestive of an extreme phenotype, the inclusion of at least one family member in addition to the proband, and a CES prescription performed by an expert in the field of rare genetic disorders.

Meaning: Our results could be translated into several perspectives from the clinicians' point of view and pave the way toward better CES prescription strategies to optimize CES diagnostic efficiency.

1. INTRODUCTION

Rare diseases affect more than 400 million people worldwide, corresponding to an estimated cumulative population prevalence of \approx 3.5 to 5.9%, according to the Orphanet database (77). Genetic disorders represent 72% of rare diseases and 70% have an exclusive pediatric onset (77). The advent of high throughput sequencing methods for DNA analysis has revolutionized the diagnostic approach of patients with suspected genetic disorders (78). In recent years, several studies have reported the usefulness of clinical exome sequencing (CES) (78–85), whole-exome sequencing (WES) (86–91), and whole-genome sequencing (WGS) (90–96) in patients with highly suggestive Mendelian phenotypes and reported diagnostic yields ranging from 9 to 61%, 32 to 74%, and 16 to 42%, respectively. In neonates and infants with critically ill conditions, CES, WES, or WGS approaches achieved a molecular diagnostic rate of 37 to 72% (97–105).

All studies that assessed high throughput sequencing methods for obtaining molecular diagnoses have focused on the diagnostic rate as the primary outcome. Few studies assessed whether baseline patients' characteristics could predict exome sequencing efficiency for establishing a molecular diagnosis. However, to our knowledge, no study has assessed the predictors of exome sequencing clinical utility, including the diagnostic yield, the confirmation of the suspected clinical diagnosis, and therapeutic guidance based on the molecular diagnosis report.

Here, we report a real-life setting experience on more than 600 CESs performed at a referral center as a first-tier genetic test among adult and pediatric patients with various Mendelian phenotypes, including inherited metabolic disorders. We assessed the CES efficiency for achieving a molecular diagnosis and reported independent predictors of CES clinical utility.

2. PATIENTS AND METHODS

2.1. French framework for the diagnosis and management of genetic disorders

As previously reported (78), the management of genetic disorders in France is organized around specialized reference centers within the “French National Plan for Rare Diseases” framework. One of these centers is located at the University Hospital of Nancy (ORPHA67872) and receives patients referred by their treating physician when a genetic disease is suspected. Genetic testing is performed after the formal written consent from the adult patient or proband’s parents in the case of pediatric patients. In France, in the setting of the diagnosis of genetic disorders and inborn errors of metabolism, the CES is 100% covered by the French health care system (78).

2.2. Study design, setting, and patients’ selection criteria

We carried out a retrospective observational study on consecutive patients who underwent CES at the Department of Molecular Medicine at the University Hospital of Nancy (Nancy, France), as a first-tier genetic test to investigate a Mendelian phenotype. The inclusion criteria were: i) CES prescribed between January 2016 and December 2019; ii) adult or pediatric patients; and iii) CES prescribed for a suspected genetic disorder other than phenylketonuria or hyperphenylalaninemia for whom a dedicated molecular diagnostic approach is available via Sanger and/or targeted gene panel sequencing. There were no exclusion criteria. The study was observational, meaning that all clinical evaluations, biochemical investigations, imaging examinations, and clinical diagnoses were performed at the discretion of the treating physicians. The institutional review board of the University Hospital of Nancy approved the study.

2.3. Data collected for the study

The following CES-related administrative and clinical data were retrieved through electronic chart review using DxCare® software (Dedalus France, Le Plessis Robinson, France) for patients followed at the University Hospital of Nancy or medical prescriptions for patients outside the University Hospital of Nancy: institution (university hospital, regional hospital, or private practice); setting (outpatient clinic, hospital department at the University Hospital of Nancy, medical day hospital, or outside the University Hospital of Nancy); adult or pediatric department; geographical region; department of CES prescription; proband vs. sibling; trio analysis (yes/no); gender; date of birth; date of CES prescription (age at CES prescription expressed in years); date of CES clinical report; date of the molecular diagnosis report established by the molecular medicine physician (test-to-report time expressed in months); suspected diagnoses that were regrouped and classified in main categories based on the

recruitment of the Department of Molecular Medicine and the Reference Center for Inborn Errors of Metabolism (metabolic disorders; dyslipidemia; liver and biliary tract disorders; neurological disorder; inflammatory and auto-inflammatory disease; developmental abnormality; mitochondrial cytopathy; pancreatitis; intestinal absorption disorders; myopathy; osteogenesis imperfecta; lipodystrophy; primary immunodeficiencies; and miscellaneous conditions). CES-related genomic data were extracted from the molecular diagnosis report established by the molecular medicine physician and included: the number of retained variants; variant annotation of each genomic variant (#1 to #4 in the order reported in the CES report) according to the Human Genome Variation Society nomenclatures using both coding DNA (c.) and protein reference sequences (p.); gene; variant classification according to the American College of Medical Genetics and Genomics (ACMG) criteria (74); patients' genotype status. We assessed patients' follow-up data regarding the clinical utility of the CES as reported by the prescribing physician: 1) confirmation or contribution to the clinically suspected diagnosis and 2) patient's treatment adaptation based on the conclusions of the CES report.

2.4. Next-generation sequencing

We performed CES using the TruSight One panel on an Illumina MiSeq platform (Illumina, Evry, France) or TruSight One expanded panel on an Illumina NextSeq 550 platform, in compliance with the French Accreditation Committee requirements, at the University Hospital of Nancy and the Functional Genomics Facility of the INSERM unit UMR_S 1256 (NGERE), as previously described (78,106,107). At the Department of Molecular Medicine at the University Hospital of Nancy and the National Reference Center for Inborn Errors of Metabolism, we opted for a strategy based on Illumina TruSight captures that cover a large number of disease-associated genes instead of choosing a full-exome-based approach. The TruSight One Sequencing Panel provides comprehensive coverage of >4800 disease-associated genes, while the TruSight One Expanded Sequencing Panel targets ~1900 additional genes with recent disease associations in the scientific literature (**Supplemental Table 1**). We used Nextera and Nextera Flex enrichment solutions with the TruSight One and TruSight One Expanded captures, respectively. The TruSight One Expanded and TruSight One panels cover 100% and 82% (145/176) of the 176 mitochondrial nuclear genes reported in **Supplemental Table 1**. Bioinformatics analysis was performed as previously described (78,107–109). The Burrows-Wheeler Aligner (BWA) software package was used to map all reads to GRCh37/hg19 annotation using the BWA-MEM algorithm. PCR duplicates were removed from alignments using the Picard (MarkDuplicates) tool (<http://broadinstitute.github.io/picard>). Indels were realigned using the Genome Analysis Toolkit (GATK) (110). SNVs and short indels were called using the GATK HaplotypeCaller software (110). We used the following strategy for filtering and prioritizing the genetic variants:

Step #1: Import all retrieved genetic variants; Step #2: Retain only low frequency and rare variants by excluding variants with a reported alternative allele frequency (AAF) >5% in the following databases: The Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD Exomes). The aim of step #2 is to retain rare (AAF<1%) and low-frequency (AAF<5%) variants that exhibit relatively large effects on disease risk, such as those involved in severe monogenic diseases (111); Step #3: Exclude intronic variants other than splice donor, splice acceptor or splice region variants or synonymous variants; Step #4: Prioritization of genetic variants for their association with the studied phenotype using the SVS-PhoRank gene ranking algorithm based on the Phevor algorithm (111). The Phevor algorithm scores ontology terms based on their proximity to patient's phenotypes. The SVS-PhoRank score was calculated according to the RefSeq Genes 105v2, NCBI gene source, using the following ontologies: Human Phenotype Ontology, Gene Ontology, and OMIM Phenotype Ontology. Functional variant annotation was performed using the following tools: MutationTaster (112), Functional Analysis through Hidden Markov Models (FATHMM) (113), MetaSVM / MetaLR scores (114), Sorting Tolerant From Intolerant (SIFT) (115), Protein Variation Effect Analyzer (Provean) (116), and deleterious annotation of genetic variants using neural networks (DANN) (117). Variant pathogenicity was reported according to the ACMG criteria (74). All retained and top-prioritized variants were systematically assessed for their ClinVar annotation (118). All the genetic variants retained in the molecular diagnosis report had to meet high-quality metric criteria (read depth [RD] >20, alt-read ratio [ARR] >0.2, genotype quality score [GQ] of 99). All the genetic variants retained in the bioinformatics analysis were confirmed by visual inspection of the read-level data of BAM files. Sanger confirmatory sequencing has been performed in rare cases of variants with suboptimal quality metrics indicators (e.g., splice donor or acceptor variants, homozygous variants). Large structural variations were not assessed. All bioinformatics analyses, including the SVS-PhoRank gene ranking algorithm, were performed using the SNP & Variation Suite (v8.9.0; Golden Helix, Inc., Bozeman, MT, USA).

2.5. Study aims and outcomes

The aims of the study were: i) to assess the CES efficiency for achieving a molecular diagnosis as a first-tier genetic test in patients with a Mendelian phenotype in the whole cohort and according to the suspected diagnoses; ii) to look for predictors of discovering at least one variant classified as pathogenic or likely pathogenic (LP) (P/LP); iii) to look for predictors of confirming the suspected clinical diagnosis; and iv) to look for predictors of a CES-guided treatment strategy by the prescribing physician.

The study outcomes were the discovery of at least one P/LP variant, the “Confirmation of the suspected clinical diagnosis by CES”, and “CES-guided treatment strategy”. In the diagnostic yield analysis, we considered patients for whom the CES report retained at least one genetic variant classified as P/LP or by using a broader definition by adding patients with at least one genetic variant classified as of uncertain significance (VUS). The assessment of the “Confirmation of the suspected clinical diagnosis by CES” and “CES-guided treatment strategy”, as outcomes, was based on chart review and the medical consultation report made by the prescribing physician following the molecular diagnosis restitution to the patient or the parents’ proband.

2.6. Statistical Analysis

Categorical variables were summarized as frequency counts and percentages with the 95% confidence interval (95% CI). Quantitative variables were expressed as medians and interquartile range (IQR, 25th and 75th percentiles). We used univariate logistic regression to look for predictors of: i) discovering at least one variant classified as P/LP; ii) confirming the suspected clinical diagnosis; iii) a CES-guided treatment strategy by the prescribing physician. All the significant variables retrieved in univariate analyses were integrated into a multivariable logistic regression model to look for the variables independently associated with the studied outcomes. All variables with $P < 0.1$ were included in the model and the variables with $P < 0.05$ were retained in the model. Results are shown as odds ratios (ORs) and 95% CI for each independent predictor, and the percentage of cases correctly classified by the logistic regression model. We assessed model discrimination using ROC analysis and the percentage of cases correctly classified by the model. We assessed the goodness of fit of the model using Nagelkerke R^2 and Cox & Snell R^2 statistics (119). All statistical analyses were conducted using MedCalc, version 19.5.3 (MedCalc Software, Ostend, Belgium) and SVS (v8.8.1; Golden Helix, Inc., Bozeman, MT, USA).

3. RESULTS

3.1. Study population

Between January 2016 and December 2019, 603 patients (426 probands and 177 siblings) underwent CES at the Department of Molecular Medicine of the University Hospital of Nancy (**Table 1**). The median age of the probands was 34 years (IQR, 12–48) and the proportion of males was 46.9% (200/426). The majority of CES prescriptions originated from the University Hospital of Nancy (96.7%, 412/426), mainly from outpatient clinics (83.3%, 355/426) and hospital departments (7.5%, 32/426). Adults and pediatric populations represented 64.8% (276/426) and 35.2% (150/426), respectively. Eighty percent of CESs were performed for patients followed at the Reference Centre for Inborn Errors of Metabolism (53.8%, 229/426),

the Department of Gastrointestinal & Liver diseases (15.7%, 67/426), or the Department of Endocrinology and Nutrition (13.4%, 57/426) (**Table 1**). The main suspected diagnoses in relation to CES prescription were metabolic disorders (40.6%, 173/426), dyslipidemia (17.8%, 76/426), liver and biliary tract disorders (15.3%, 65/426), and neurological disorders (6.3%, 27/426) (**Table 2**). One-carbon metabolism disorders represented 24.4% of CES indications (104/426) and more than 60% of CES performed in the setting of 'Metabolic disorders' (104/173). Among patients with dyslipidemia, more than 85% (66/76) of CESs were performed in patients exhibiting hypercholesterolemia (46/66, 69.7%) or hypertriglyceridemia (20/66, 30.3%) (**Table 2**). Among patients with liver or biliary tract disorders, CESs were mainly performed for exploring hyperferritinemia (19/426, 4.5%), a suspicion of low phospholipid-associated cholelithiasis syndrome (12/426, 2.8%), or a cholestatic disorder (12/426, 2.8%), totaling 66% (43/65) of CES indications in this subgroup (**Table 2**).

3.2. Diagnostic yield of the clinical exome sequencing

The median test-to-report time was 5.6 months (IQR, 4.1–7.2). The median number of genetic variants reported per patient in the CES was 1 (IQR, 0–2). The same figures were 1 (IQR, 0–1) and 0 (IQR, 0–1) when P/LP/VUS variants or only P/LP variants were considered, respectively. The proportions of patients with genetic variants according to the ACMG classification are reported in **Table 3**, **Figure 1**, and **Supplemental Figure S1**. The CES revealed 203 P/LP variants in 160 patients, corresponding to a diagnostic yield of 37.6% (160/426) (**Table 3** and **Supplemental Figure S2**). This figure was 60.8% (259/426) when VUS variants were included in the diagnostic yield analysis (**Table 3**). The first two variants (variants #1 and #2) classified as P/LP represented 96.1% (195/203) of all P/LP variants reported (**Figure 2** and **Supplemental Figure S2**).

The diagnostic yield of CES varied according to the suspected diagnosis (**Table 4**). When P/LP variants were considered, the CES diagnostic yield was >40% for patients with neurological disorders, dyslipidemia, developmental abnormalities, osteogenesis imperfecta, intestinal absorption disorders, and lipodystrophy. When VUS variants were included in the evaluation of CES efficiency, all suspected diagnoses subgroups had a diagnostic yield above 40%, with the top items represented by osteogenesis imperfecta, neurological disorders, mitochondrial cytopathy, dyslipidemia, inherited metabolic disorders, and lipodystrophy (**Figure 3**).

3.3. Predictors of CES efficiency in univariate and multivariable analyses

3.3.1. Predictors of discovering at least one P/LP variant

In univariate analyses, several items were associated with discovering at least one genetic variant classified as P/LP (**Table 5**). Among them, three maintained their significance in the

multivariable analysis: 'Suspected diagnosis subgroup, Lysosomal disorder' (OR, 11.81 [95% CI, 1.35– 103.33]; P = 0.03), 'Suspected diagnosis subgroup, Hypercholesterolemia' (OR, 4.14 [95% CI, 2.15–7.96]; P<0.0001), and 'Pediatric care department' (OR, 1.96 [95% CI, 1.28–2.99]; P = 0.002) (**Table 5**). The univariate and multivariate analyses assessing the predictors of retrieving at least one P/LP/VUS variant are reported in **Supplemental Table S2**.

3.3.2. Predictors of confirming the suspected clinical diagnosis

The CES report confirmed the suspected clinical diagnosis in 43.0% (183/426) of the patients and contributed to the diagnosis in 7.5% (32/426) of them (**Table 3**). In univariate analyses, several items were associated with a confirmation or a contribution to the suspected clinical diagnosis (**Table 6**). Among them, three maintained their significance in the multivariable analysis: 'Suspected diagnosis subgroup, Hypercholesterolemia' (OR, 2.94 [95% CI, 1.52–5.70]; P = 0.001), 'Exome sequencing on the proband and at least one family member' (OR, 1.90 [95% CI, 1.20–3.01]; P = 0.007), and 'Suspected diagnosis group, Liver disorder' (OR, 0.30 [95% CI, 0.14–0.65]; P = 0.002).

3.3.3. Predictors of a CES-guided Treatment Strategy in Univariate and Multivariable

Sixty-two patients (14.6%) had their therapy guided by their physician based on NGS results. In univariate analyses, several items were associated with a CES-guided treatment strategy (**Table 7**). Among them, four maintained their significance in the multivariable analysis: 'Confirmation of the suspected diagnosis, Yes or Contributive' (OR, 3.75 [95% CI, 1.74–8.12]; P = 0.0008), 'Suspected diagnosis subgroup, Hypercholesterolemia' (OR, 3.45 [95% CI, 1.56–7.66]; P = 0.002), 'Number of genetic variants reported in the CES report per patient' (OR, 1.44 [95% CI, 1.06–1.97]; P = 0.02), and 'Department, Reference Center for Inborn Errors of Metabolism' (OR, 1.94 [95% CI, 1.02–3.69]; P = 0.04) (**Table 7**).

4. DISCUSSION

Using a retrospective observational study on more than 603 consecutive subjects, including 426 probands, who underwent CES as a first-tier genetic test to investigate a Mendelian phenotype, we report a diagnostic yield \approx 38% for reporting P/LP variants. The CES diagnostic yield was greater than 40% among patients with neurological disorders, dyslipidemia, developmental abnormalities, osteogenesis imperfecta, intestinal absorption disorders, and lipodystrophy.

The predictors independently associated with detecting at least one plausible P/LP variant included a CES prescribed in a pediatric care department, a suspected lysosomal disorder, or

hypercholesterolemia. Regarding the confirmation of the suspected clinical diagnosis, independent predictors were a CES prescription for hypercholesterolemia and the inclusion of at least one family member in addition to the proband. Finally, several predictors were associated with a CES-guided therapy. They included the confirmation of the suspected diagnosis by CES, a CES prescription originating from the Reference Center for Inborn Errors of Metabolism, and the total number of genetic variants reported per patient.

Exome and WGS have been evaluated in genetic disorders with more evidence for CES, which exhibited a molecular diagnostic yield ranging from 9 to 51% (78–105). The place of WGS in the genetic diagnosis armamentarium is unclear, mainly due to the lack of reliable data comparing these two approaches in terms of diagnostic efficiency. A study on 108 patients suggested that WGS offered additional but limited clinical utility in this setting (120). A recent study from the U.K. 100,000 Genomes Project reported a diagnostic yield of 25% for WGS on a cohort of 2183 probands with a broad spectrum of rare diseases (96). Interestingly, the diagnostic yield was higher among patients with suspected monogenic disorders (35%) when compared to those with complex phenotypes (11%) (96). In our study, we identified three predictors that were independently associated with CES diagnostic yield, including lysosomal disorders and hypercholesterolemia, which correspond to monogenic diseases and CES prescription from a pediatric care department. In this latter category, younger age at the time of disease presentation may be considered a surrogate marker of extreme phenotype, thereby increasing the risk of a priori probability of a monogenic disorder. Furthermore, and consistently with the results from the U.K. 100,000 Genomes Project (96), the CES diagnostic yield in our study was > 40% among patients with osteogenesis imperfecta, developmental abnormalities, dyslipidemia, and neurological disorders, while it was < 25% among patients with more complex presentations including immunodeficiencies and inflammatory diseases.

A limited number of studies assessed the predictors of CES or WGS utility among patients with various Mendelian phenotypes (85,96,121). Bonaventura *et al.* reported the performance of the Mayo hypertrophic cardiomyopathy genotype predictor score on 336 consecutive patients with hypertrophic cardiomyopathy and reported a positive association between the clinical score and the diagnostic rate of an NGS-based panel of 229 genes (121). A recently published Australian study investigated the predictors of CES utility among 204 patients assessed in multidisciplinary renal genetics clinics and reported a younger age at presentation as an independent predictor of CES efficiency (85). The results from the U.K. 100,000 Genomes Project confirmed that the diagnostic yields of WGS were the highest among family trios and families with larger pedigrees (96). In line with these results, in our study, the inclusion of at least one family member in addition to the proband was independently associated with the confirmation of the suspected clinical diagnosis by the CES.

Using a systematic approach and predefined outcomes, we assessed the clinical utility of CES as a first-tier genetic test on a large dataset of consecutive adult and pediatric patients with various Mendelian phenotypes. Our results could be translated into several perspectives from the clinician's point of view and a well-structured dialog framework between prescribing clinicians and molecular medicine physicians. First, the clinical utility of CES could be assessed on well-powered prospective studies using predefined inclusion criteria to allow a better inference of the results on a large population. In addition, a randomized trial could be designed to evaluate a first-line CES-based strategy in patients with high expected diagnostic yield compared to a first-line WGS-based strategy in patients with low expected diagnostic yield from CES, using a diagnostic yield cut-off \approx 25-30%, based on the results from the U.K. 100,000 Genomes Project (96). Second, in line with previous studies, our results reinforce the added value of family-based CES prescription, notably when VUS variants are discovered (122). Third, our study showed that a CES prescription performed in a reference center is more likely to lead to a molecular diagnosis. Notably, the confirmation of the suspected diagnosis by CES represented an independent predictor of CES-guided therapy. Our results support the importance of prescribing CESs in close interaction between referring physicians with expertise in genetic disorders and molecular medicine physicians.

We acknowledge several potential limitations of the study that should be considered in interpreting our results. First, we report results from a retrospective single-center study that need to be confirmed in independent cohorts (78). Second, given the constantly evolving knowledge of gene and variant annotations, non-contributory CES may in the future reveal variants upgraded for their pathogenicity (78). Third, the efficiency of CES is suboptimal in detecting deep intronic or regulatory mutations and copy number variants (CNV) (78). In this setting, WGS exhibited an improved diagnostic yield compared with targeted gene sequencing panels, particularly by highlighting structural and non-exonic sequence variants not detectable with whole-exome sequencing (123). Our study has several strengths. First, we reported one of the most extensive European series of consecutive patients with a Mendelian phenotype that assessed the diagnostic yield of a large CES panel in the routine practice of a Molecular Medicine Department. Second, we reported a median CES test-to-report time of 5 months, consistent with effective patient care. Third, in our bioinformatics analysis pipeline, we systematically combined classical methods of variant filtration and annotation with the SVS-PhoRank phenotype-driven computational algorithm, in line with our molecular diagnostic strategy based on gene captures that spanned several thousand disease-associated genes.

CONCLUSIONS

In conclusion, using one of the most extensive series of consecutive patients with various Mendelian phenotypes, we evaluated the diagnostic yield of a large CES panel as a first-line diagnostic strategy in a real-life setting. Independent predictors of CES utility in terms of diagnostic performance and confirmation of the suspected diagnosis were criteria strongly suggestive of an extreme phenotype, including pediatric presentations and patients' phenotypes associated with an increased risk of a priori probability of a monogenic disorder, the inclusion of at least one family member in addition to the proband, and a CES prescription performed by an expert in the field of rare genetic disorders. Our results could be translated into several perspectives from the clinicians' point of view and pave the way toward better CES prescription strategies to optimize CES diagnostic efficiency.

Conflict of Interest statement

The authors who have taken part in this study declare that they do not have anything to disclose regarding conflicts of interest concerning this manuscript.

Funding

INSERM UMR_S 1256, Nutrition, Genetics, and Environmental Risk Exposure.

Data availability statement

Anonymized patient data are available for use in collaborative studies to researchers upon reasonable request (abderrahim.oussalah@univ-lorraine.fr). Data will be provided following the review and approval of a research proposal (including a statistical analysis plan) and completion of a data-sharing agreement. Responses to the request for the raw data will be judged by the IRB of the INSERM UMR_S 1256 and the University Hospital of Nancy.

Authors' contributions (according to CRediT statement)

T.A.: Formal analysis; Writing – original draft preparation; Writing – review and editing; Visualization; C.C.: Investigation; Resources; Writing – review and editing; T.J.: Investigation; Resources; Writing – review and editing; J-P.B.: Investigation; Resources; Writing – review and editing; F.F.: Investigation; Resources; Writing – review and editing; R-M.G.R.: Investigation; Resources; Writing – review and editing; F.N.: Investigation; Resources; Writing – review and editing; J-L.G.: Conceptualization; Validation; Investigation; Resources; Writing – original draft preparation; Writing – review and editing; Visualization; Supervision; Project administration; Funding acquisition; A.O.: Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Resources; Writing – original draft preparation; Writing – review and editing; Visualization; Supervision; Project administration; Funding acquisition.

TABLES

Table 1. Description of 426 Patients Assessed Using Clinical Exome Sequencing at the Department of Molecular Medicine of the University Hospital of Nancy.

Demographics			
Age at clinical exome sequencing (years) — n, median (IQR)	426	34	(12–48)
Female — n/N, % (95% CI)	226/426	53.1	(48.3–57.8)
Male — n/N, % (95% CI)	200/426	46.9	(42.2–51.7)
Institution — n/N, % (95% CI)			
University hospital	412/426	96.7	(95.0–98.4)
Regional hospital	13/426	3.1	(1.4–4.7)
Private practice	1/426	0.2	(0*–0.7)
Setting — n/N, % (95% CI)			
Outpatient clinic, University Hospital of Nancy	355/426	83.3	(79.8–86.9)
Hospital department, University Hospital of Nancy	32/426	7.5	(5.0–10.0)
Medical day hospital, University Hospital of Nancy	7/426	1.7	(0.4–2.9)
Outside the University Hospital of Nancy	32/426	7.5	(5.0–10.0)
Adult, pediatric Departments — n/N, % (95% CI)			
Adult care department	276/426	64.8	(60.2–69.3)
Pediatric care department	150/426	35.2	(30.7–39.8)
Geographical region and city — n/N, % (95% CI)			
Nancy	394/426	92.5	(90.0–95.0)
North-East region, outside Nancy	16/426	3.8	(1.9–5.6)
Other	16/426	3.8	(1.9–5.6)
Department — n/N, % (95% CI)			
Reference Centre for Inborn Errors of Metabolism	229/426	53.8	(49.0–58.5)
Gastrointestinal & Liver diseases	67/426	15.7	(12.3–19.2)
Endocrinology and nutrition	57/426	13.4	(10.1–16.6)
Pediatrics	20/426	4.7	(2.7–6.7)
Neurology	12/426	2.8	(1.2–4.4)
Internal Medicine	11/426	2.6	(1.1–4.1)
Clinical genetics	8/426	1.9	(0.6–3.2)
Vascular medicine	5/426	1.2	(0.1–2.2)
Cardiology	4/426	0.9	(0–1.9)
Hematology	4/426	0.9	(0–1.9)
Nephrology	3/426	0.7	(0*–1.5)
Hematology-oncology	2/426	0.5	(0*–1.1)
Geriatrics	1/426	0.2	(0*–0.7)
Gynecology and Obstetrics	1/426	0.2	(0*–0.7)
Orthopedics	1/426	0.2	(0*–0.7)
Intensive care	1/426	0.2	(0*–0.7)

Table 2. Description of the Suspected Diagnoses Associated with the Clinical Exome Sequencing Prescriptions

Suspected diagnoses — n/N, % (95% CI)			
Metabolic disorders	173/426	40.6	(35.9–45.3)
One-carbon metabolism disorders	104/426	24.4	(20.3–28.5)
Energy metabolism disorders	9/426	2.1	(0.7–3.5)
Organic acidurias	6/426	1.4	(0.3–2.5)
Congenital hyperinsulinisms	6/426	1.4	(0.3–2.5)
Lysosomal storage disorders*	6/426	1.4	(0.3–2.5)
Glycogen storage diseases	5/426	1.2	(0.1–2.2)
Hyperbilirubinemias	5/426	1.2	(0.1–2.2)
Peroxisomal disorders	5/426	1.2	(0.1–2.2)
Phenylketonuria	5/426	1.2	(0.1–2.2)
Calcium and phosphorus metabolic disorders	5/426	1.2	(0.1–2.2)
Alkaptonuria	3/426	0.7	(0+–1.5)
Metabolic disorders, Miscellaneous	14/426	3.3	(1.6–5.0)
Dyslipidemia	76/426	17.8	(14.2–21.5)
Hypercholesterolemia	46/426	10.8	(7.8–13.8)
Hypertriglyceridemia	20/426	4.7	(2.7–6.7)
Hypolipoproteinemia	5/426	1.2	(0.1–2.2)
Mixed dyslipidemia	2/426	0.5	(0+–1.1)
Dyslipidemia, Other	3/426	0.7	(0+–1.5)
Liver and biliary tract disorders	65/426	15.2	(12.0–19.0)
Hyperferritinemia	19/426	4.5	(2.5–6.4)
Low phospholipid-associated cholelithiasis	12/426	2.8	(1.2–4.4)
Cholestatic disorders	12/426	2.8	(1.2–4.4)
Wilson’s disease	9/426	2.1	(0.7–3.5)
Cryptogenic cirrhosis	4/426	0.9	(0–1.9)
Chronic liver cytolysis	4/426	0.9	(0–1.9)
Alpha-1-antitrypsin deficiency	2/426	0.5	(0+–1.1)
Polycystic liver disease	2/426	0.5	(0+–1.1)
Liver steatosis	1/426	0.2	(0+–0.7)
Neurological disorder	27/426	6.3	(4.0–8.7)
Ataxia, hypotonia, paraparesis	13/426	3.1	(1.4–4.7)
Mental retardation with or without autism	5/426	1.2	(0.1–2.2)
Epilepsy	3/426	0.7	(0+–1.5)
Neurological disorders, Other	6/426	1.4	(0.3–2.5)
Inflammatory and auto-inflammatory disease	18/426	4.2	(2.3–6.1)
Auto-inflammatory diseases	17/426	4.0	(2.1–5.9)
Inflammatory diseases, Other	1/426	0.2	(0+–0.7)
Developmental abnormality	13/426	3.1	(1.4–4.7)
Heart defects	4/426	0.9	(0–1.9)
Neural tube defects	3/426	0.7	(0+–1.5)
Developmental abnormality, Other	6/426	1.4	(0.3–2.5)
Mitochondrial cytopathy	8/426	1.9	(0.6–3.2)
Pancreatitis	8/426	1.9	(0.6–3.2)

Intestinal absorption disorders	7/426	1.6	(0.4–2.9)
Myopathy	7/426	1.6	(0.4–2.9)
Osteogenesis imperfecta	7/426	1.6	(0.4–2.9)
Lipodystrophy	5/426	1.2	(0.1–2.2)
Primary immunodeficiencies	4/426	0.9	(0–2.2)
Other	8/426	1.9	(0.6–3.2)
Thrombophilia	2/426	0.5	(0+ –1.1)
Amyloidosis	1/426	0.2	(0+ –0.7)
Cancer	2/426	0.5	(0+ –1.1)
Marfan syndrome	1/426	0.2	(0+ –0.7)
Sudden death	1/426	0.2	(0+ –0.7)
Telomere Diseases	1/426	0.2	(0+ –0.7)

Note. n: number of observations; N: total number of patients; 95% CI: 95% confidence interval.

* Lysosomal storage disorders other than glycogen storage disease type II.

† The 95% CI was truncated at the left margin.

Table 3. Results of the Clinical Exome Sequencing Clinical Report in the 426 Patients That Were Assessed.

Time delay between CES and report (Months) — n, median (IQR)	419	5.6	(4.1–7.2)
Description of the genetic variants reported in the clinical exome sequencing report			
Proportion of patients with at least one variant classified as LP or P — n/N, % (95% CI)	160/426	37.6%	(32.9–42.2)
Proportion of patients with at least one variant classified as VUS, LP, or P — n/N, % (95% CI)	259/426	60.8%	(56.1–65.5)
Number of genetic variants reported in the CES report per patient — n, median (IQR)	426	1	(0–2)
Number of genetic variants classified as VUS, LP, or P per patient — n, median (IQR)	426	1	(0–1)
Number of genetic variants classified as LP, or P per patient — n, median (IQR)	426	0	(0–1)
Variant#1			
ACMG, pathogenic — n/N, % (95% CI)	78/426	18.3%	(14.6–22)
ACMG, likely pathogenic — n/N, % (95% CI)	72/426	16.9%	(13.3–20.5)
ACMG, uncertain significance — n/N, % (95% CI)	104/426	24.4%	(20.3–28.5)
ACMG, likely benign — n/N, % (95% CI)	8/426	1.9%	(0.6–3.2)
ACMG, benign — n/N, % (95% CI)	22/426	5.2%	(3.1–7.3)
No variant retrieved — n/N, % (95% CI)	142/426	33.2%	(28.8–37.8)
NGS-diag network scoring system — n, median (IQR)	284	4	(3–5)
Variant#2			
ACMG, pathogenic — n/N, % (95% CI)	24/426	5.6%	(3.4–7.8)
ACMG, likely pathogenic — n/N, % (95% CI)	21/426	4.9%	(2.9–7.0)
ACMG, uncertain significance — n/N, % (95% CI)	52/426	12.2%	(9.1–15.3)
ACMG, likely benign — n/N, % (95% CI)	3/426	0.7%	(0–1.5)
ACMG, benign — n/N, % (95% CI)	17/426	4.0%	(2.1–5.9)
No variant retrieved — n/N, % (95% CI)	309/426	72.5%	(68.3–76.8)
NGS-diag network scoring system — n, median (IQR)	117	3	(3–4)
Variant#3			
ACMG, pathogenic — n/N, % (95% CI)	1/426	0.2%	(0–0.6)
ACMG, likely pathogenic — n/N, % (95% CI)	6/426	1.4%	(0.3–2.5)
ACMG, uncertain significance — n/N, % (95% CI)	15/426	3.5%	(1.8–5.3)
ACMG, likely benign — n/N, % (95% CI)	1/426	0.2%	(0–0.7)
ACMG, benign — n/N, % (95% CI)	8/426	1.9%	(0.6–3.2)
No variant retrieved — n/N, % (95% CI)	395/426	92.7%	(89.8–95.0)

NGS-diag network scoring system — n, median (IQR)	31	3	(1–3)
Variant#4			
ACMG, pathogenic — n/N, % (95% CI)	0/426	0%	(—)
ACMG, likely pathogenic — n/N, % (95% CI)	1/426	0.2%	(0–0.6)
ACMG, uncertain significance — n/N, % (95% CI)	6/426	1.4%	(0.3–2.5)
ACMG, likely benign — n/N, % (95% CI)	1/426	0.2%	(0–0.6)
ACMG, benign — n/N, % (95% CI)	5/426	1.2%	(0.1–2.2)
No variant retrieved — n/N, % (95% CI)	413/426	96.9%	(94.7–98.3)
NGS-diag network scoring system — n, median (IQR)	13	3	(1–3)
Confirmation of the clinically-suspected diagnosis — n/N, % (95% CI)			
Confirmation of the suspected diagnosis	183/426	43.0%	(38.2–47.7)
Contribution to the suspected diagnosis	32/426	7.5%	(5.0–10.0)
No contributive variant to the suspected diagnosis	211/426	49.5%	(44.8–54.3)
Therapy guided by NGS results — n/N, % (95% CI)			
Yes	62/426	14.6%	(11.2–17.9)
No	263/426	61.7%	(57.1–66.4)
No follow-up information	101/426	23.7%	(19.7–27.8)

Note. n: number of observations; N: total number of patients; 95% CI: 95% confidence interval; IQR: interquartile range, 25th–75th.

Table 4. Diagnostic Yield of the Clinical Exome Sequencing According to the Suspected Diagnosis.

Metabolic disorders			
At least one variant classified as VUS, LP, or P	108/173	62.4	(55.1–69.7)
At least one variant classified as LP or P	61/173	35.3	(28.1–42.5)
Dyslipidemia			
At least one variant classified as VUS, LP, or P	51/76	67.1	(56.3–77.9)
At least one variant classified as LP or P	37/76	48.7	(37.2–60.2)
Liver and biliary tract disorders			
At least one variant classified as VUS, LP, or P	30/65	46.2	(33.7–59.0)
At least one variant classified as LP or P	20/65	30.8	(19.9–43.5)
Neurological disorders			
At least one variant classified as VUS, LP, or P	22/27	81.5	(65.8–97.1)
At least one variant classified as LP or P	15/27	55.6	(35.5–75.6)
Inflammatory and auto-inflammatory diseases			
At least one variant classified as VUS, LP, or P	8/18	44.4	(19.0–69.9)
At least one variant classified as LP or P	4/18	22.2	(0.9–43.5)
Developmental abnormalities			
At least one variant classified as VUS, LP, or P	7/13	53.8	(22.5–85.2)
At least one variant classified as LP or P	6/13	46.2	(14.8–77.5)
Mitochondrial cytopathy			
At least one variant classified as VUS, LP, or P	6/8	75	(36.3–100*)
At least one variant classified as LP or P	3/8	37.5	(0–80.8)
Pancreatitis			
At least one variant classified as VUS, LP, or P	4/8	50	(5.3–94.7)
At least one variant classified as LP or P	2/8	25	(0–63.7)
Myopathy			
At least one variant classified as VUS, LP, or P	4/7	57.1	(7.7–100*)
At least one variant classified as LP or P	1/7	14.3	(0–49.2)
Osteogenesis imperfecta			
At least one variant classified as VUS, LP, or P	6/7	85.7	(50.8–100*)
At least one variant classified as LP or P	3/7	42.9	(0–92.3)
Intestinal absorption disorders			
At least one variant classified as VUS, LP, or P	4/7	57.1	(7.7–100)
At least one variant classified as LP or P	3/7	42.9	(†0–92.3)
Lipodystrophy			
At least one variant classified as VUS, LP, or P	3/5	60	(0–100*)
At least one variant classified as LP or P	2/5	40	(0–100*)
Primary immunodeficiencies			
At least one variant classified as VUS, LP, or P	2/4	50	(0–100)
At least one variant classified as LP or P	1/4	25	(0–100)
Other			
At least one variant classified as VUS, LP, or P	4/8	50	(5.3–94.7)
At least one variant classified as LP or P	2/8	25	(0–63.7)

Note. n: number of observations; N: total number of patients; 95% CI: 95% confidence interval.

* The 95% CI was truncated at the right margin;

† The 95% CI was truncated at the left margin.

Table 5. Predictors of Discovering at Least One Variant Classified as Likely Pathogenic or Pathogenic in Univariate and Multivariable Analyses.

Predictor	Univariate analysis			Multivariable analysis†		
	Beta (SE)	Odds ratio (95% CI)	P-value*	Beta (SE)	Odds ratio (95% CI)	P-Value‡
Suspected diagnosis subgroup, Lysosomal disorder	2.15 (1.1)	8.55 (0.99–73.84)	2.01×10 ⁻²	2.47 (1.11)	11.81 (1.35– 103.33)	0.03
Suspected diagnosis subgroup, Hypercholesterolemia	1.28 (0.33)	3.61 (1.9–6.86)	5.62×10 ⁻⁵	1.42 (0.33)	4.14 (2.15–7.96)	<0.0001
Pediatric care department	0.55 (0.21)	1.73 (1.15–2.6)	8.29×10 ⁻³	0.67 (0.22)	1.96 (1.28–2.99)	0.002
Suspected diagnosis subgroup, cholestatic disorder	-1.93 (1.05)	0.15 (0.02–1.14)	1.84×10 ⁻²	Not retained§		
Suspected diagnosis group, Dyslipidemia	0.56 (0.26)	1.75 (1.06–2.89)	2.89×10 ⁻²	Not used in the model		
Adult care department	-0.55 (0.21)	0.58 (0.38–0.87)	8.29×10 ⁻³	Not used in the model		

Note. 95% CI: 95% confidence interval; Beta: beta coefficient; SE: standard error.

* Univariate logistic regression analysis.

† Cox & Snell R², 0.07; Nagelkerke R², 0.10; Percent of cases correctly classified, 67%; AUROC, 0.629 (95% CI, 0.581 to 0.675).

‡ Multivariable logistic regression analysis using the stepwise method.

§ Not retained in the multivariate logistic regression model.

|| The following items were not used in multivariable logistic regression analysis to avoid collinearity with other items reported in the multivariable model (e.g., 'Adult care department' vs. 'Pediatric care department').

Table 6. Predictors of confirming the suspected clinical diagnosis using clinical exome sequencing In Univariate and Multivariable Analyses.

	Univariate analysis			Multivariable analysis†		
	Beta (SE)	Odds ratio (95% CI)	P-value*	Beta (SE)	Odds ratio (95% CI)	P-Value‡
Suspected diagnosis subgroup, Hypercholesterolemia	1.02 (0.34)	2.76 (1.41–5.41)	1.92×10 ⁻³	1.08 (0.34)	2.94 (1.51–5.70)	0.001
Exome sequencing on the proband and at least one family member	0.81 (0.23)	2.25 (1.42–3.56)	4.27×10 ⁻⁴	0.64 (0.24)	1.89 (1.20–3.01)	0.007
Suspected diagnosis group, Liver disorder	-1.29 (0.34)	0.28 (0.14–0.53)	3.81×10 ⁻⁵	-1.19 (0.39)	0.30 (0.14–0.65)	0.002
Suspected diagnosis group, Osteogenesis imperfecta	1.8 (1.08)	6.03 (0.72–50.51)	4.75×10 ⁻²		Not retained§	
Suspected diagnosis subgroup, One-carbon metabolism disorder	0.65 (0.23)	1.91 (1.21–3.01)	4.59×10 ⁻³		Not retained§	
Department, Reference Center for Inborn Errors of Metabolism	0.59 (0.2)	1.8 (1.22–2.64)	2.67×10 ⁻³		Not retained§	
Suspected diagnosis subgroup, Metabolic disorder	0.42 (0.2)	1.52 (1.03–2.24)	3.47×10 ⁻²		Not retained§	
Suspected diagnosis group, Inflammatory and auto-inflammatory disease	-1.01 (0.54)	0.36 (0.13–1.04)	4.55×10 ⁻²		Not retained§	
Department, Gastrointestinal and Liver diseases	-1.2 (0.3)	0.3 (0.17–0.54)	1.90×10 ⁻⁵		Not retained§	
Setting, Medical day hospital	-1.83 (1.08)	0.16 (0.02–1.34)	4.24×10 ⁻²		Not retained§	
Suspected diagnosis group, Pancreatitis	-1.99 (1.07)	0.14 (0.02–1.12)	2.18×10 ⁻²		Not retained§	
Suspected diagnosis subgroup, cholestatic disorder	-2.47 (1.05)	0.08 (0.01–0.66)	1.41×10 ⁻³		Not retained§	

Note. 95% CI: 95% confidence interval; Beta: beta coefficient; SE: standard error.

* Univariate logistic regression analysis.

† Cox & Snell R², 0.08; Nagelkerke R², 0.11; Percent of cases correctly classified, 62%; AUROC, 0.637 (95% CI, 0.589 to 0.683).

‡ Multivariable logistic regression analysis using the stepwise method.

§ Not retained in the multivariate logistic regression model

Table 7. Predictors of an NGS-guided Treatment Strategy in Univariate and Multivariable Analyses.

Predictor	Univariate analysis			Multivariable analysis†		
	Beta (SE)	Odds ratio (95% CI)	P-value*	Beta (SE)	Odds ratio (95% CI)	P-Value‡
Confirmation of the suspected diagnosis, Yes or Contributive	1.86 (0.36)	6.41 (3.16–13.01)	3.13×10 ⁻⁹	1.32 (0.39)	3.75 (1.74–8.12)	0.0008
Suspected diagnosis subgroup, Hypercholesterolemia	1.11 (0.36)	3.03 (1.51–6.08)	3.23×10 ⁻³	1.24 (0.41)	3.45 (1.56–7.66)	0.002
Number of genetic variants reported in the CES report per patient	0.5 (0.13)	1.77 (1.38–2.27)	8.24×10 ⁻⁶	0.37 (0.16)	1.44 (1.06–1.97)	0.02
Department, Reference Center for Inborn Errors of Metabolism	0.69 (0.29)	1.99 (1.12–3.52)	1.56×10 ⁻²	0.66 (0.33)	1.94 (1.02–3.69)	0.04
Setting, University Hospital of Nancy	1.74 (1.03)	5.68 (0.76–42.38)	2.57×10 ⁻²		Not retained§	
Number of genetic variants classified as LP, or P per patient	0.59 (0.22)	1.8 (1.17–2.77)	8.62×10 ⁻³		Not retained§	
Number of genetic variants classified as VUS, LP, or P per patient	0.56 (0.15)	1.76 (1.31–2.36)	2.04×10 ⁻⁴		Not retained§	
Exome sequencing on the proband and at least one family member	0.64 (0.29)	1.89 (1.06–3.36)	3.39×10 ⁻²		Not retained§	
Suspected diagnosis subgroup, One-carbon metabolism disorder	0.64 (0.29)	1.89 (1.06–3.36)	3.39×10 ⁻²		Not retained§	
Suspected diagnosis group, Liver disorder	-1.14 (0.61)	0.32 (0.1–1.06)	3.01×10 ⁻²		Not retained§	

0 Note. 95% CI: 95% confidence interval; Beta: beta coefficient; SE: standard error.

* Univariate logistic regression analysis. 36

† Cox & Snell R², 0.11; Nagelkerke R², 0.20; Percent of cases correctly classified, 85%; AUROC, 0.768 (95% CI, 0.725 to 0.807).

‡ Multivariable logistic regression analysis using the stepwise method.

§ Not retained in the multivariate logistic regression model.

FIGURES LEGENDS

Figure 1. Distribution of genetic variants according to their pathogenicity. Variants #1 to #4 647 are ordered as reported in the CES medical report.

Figure 2. Distribution of the 195 pathogenic or likely pathogenic (P/LP) variants among the 158 patients exhibiting at least one P/LP variant (variants #1 and #2 correspond to the first two variants reported in the CES medical report). The numbers outside the Venn diagram correspond to the numbers of genetic variants. The numbers inside the Venn diagram correspond to the numbers of patients. Panels A, B, and C, report figures for P, LP, and P/LP variants.

Figure 3. Diagnostic yield of CES according to the suspected diagnosis when only pathogenic or likely pathogenic (P/LP) variants were considered (panel A) or when variants of uncertain significance (VUS) were included in the diagnostic yield analysis (panel B). Vertical bars denote truncated 95% confidence intervals at the right or the left margin.

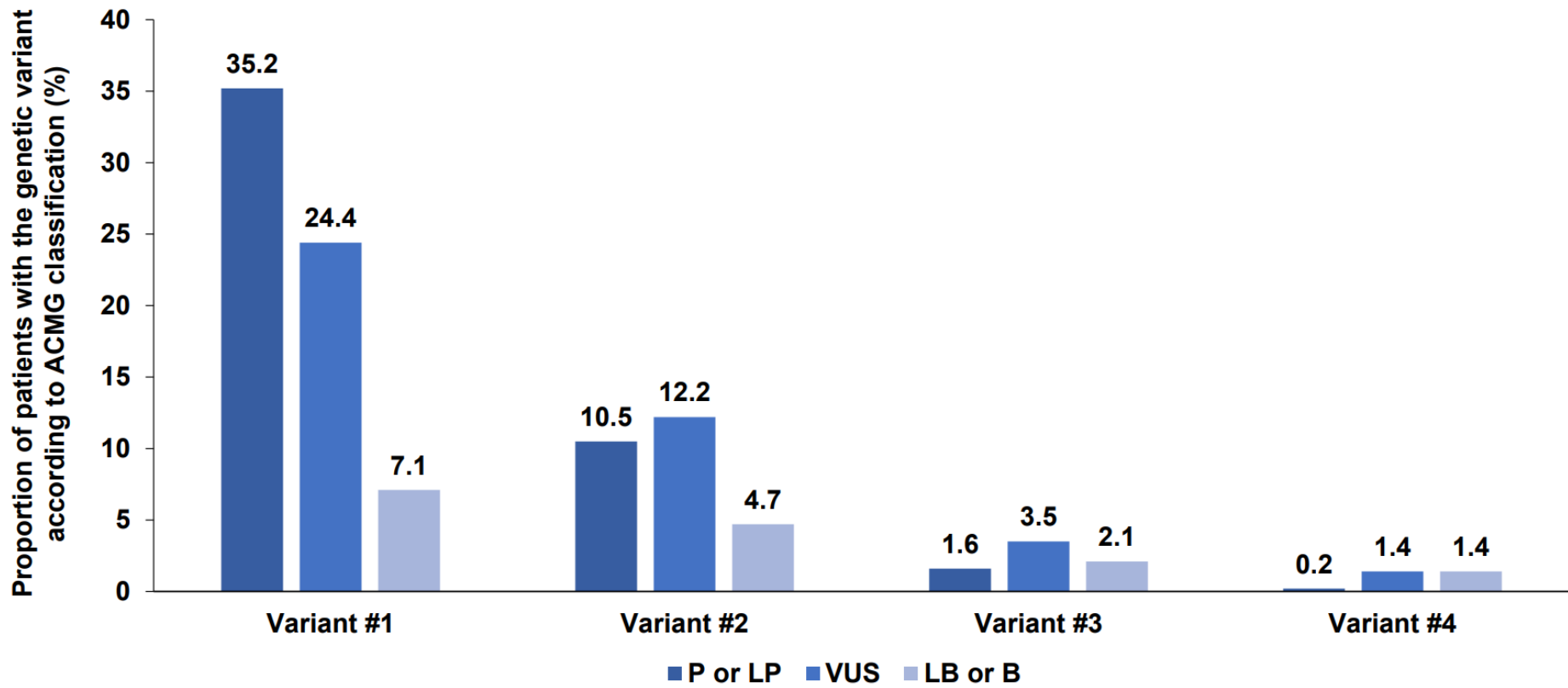


Figure 1.

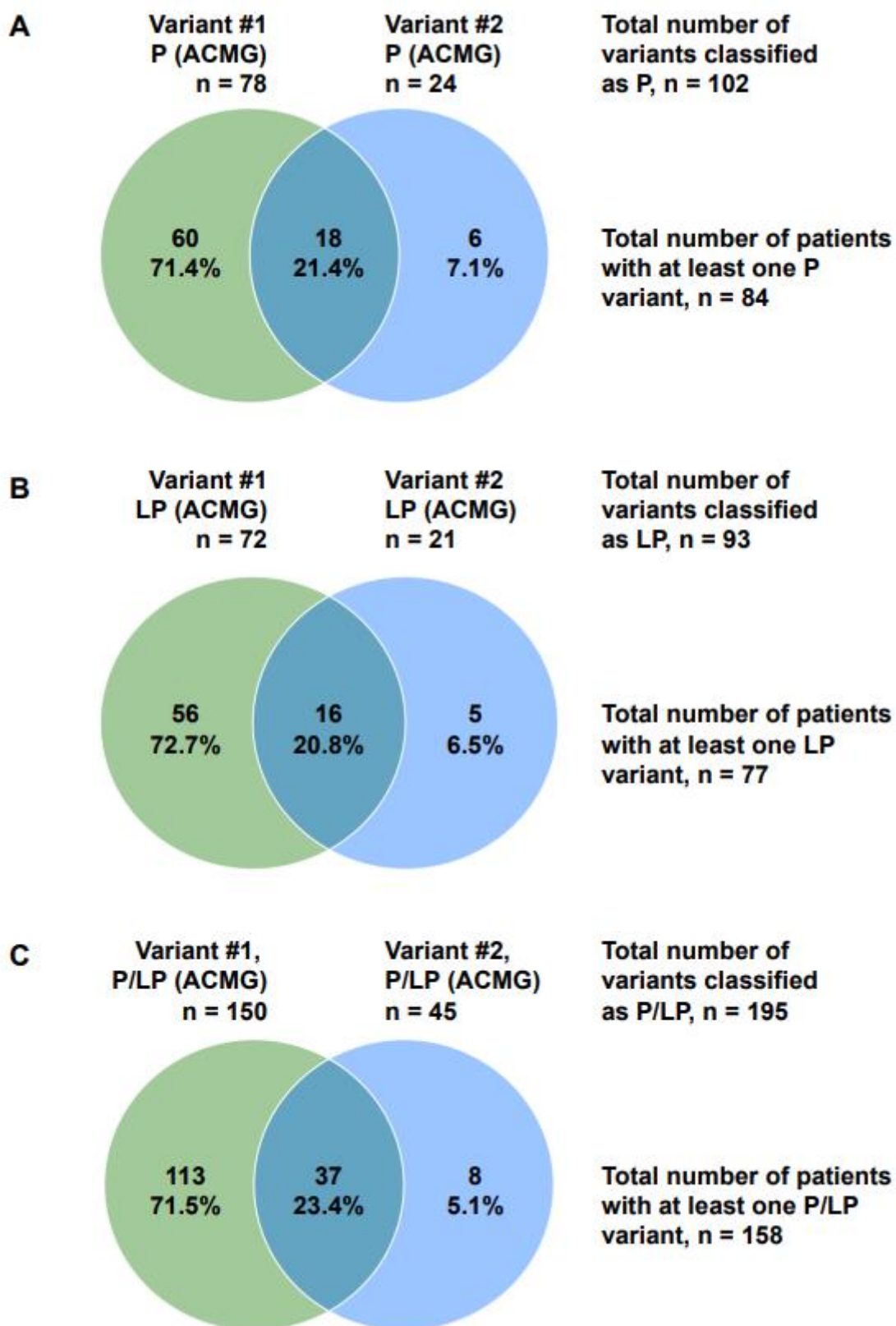


Figure 2.

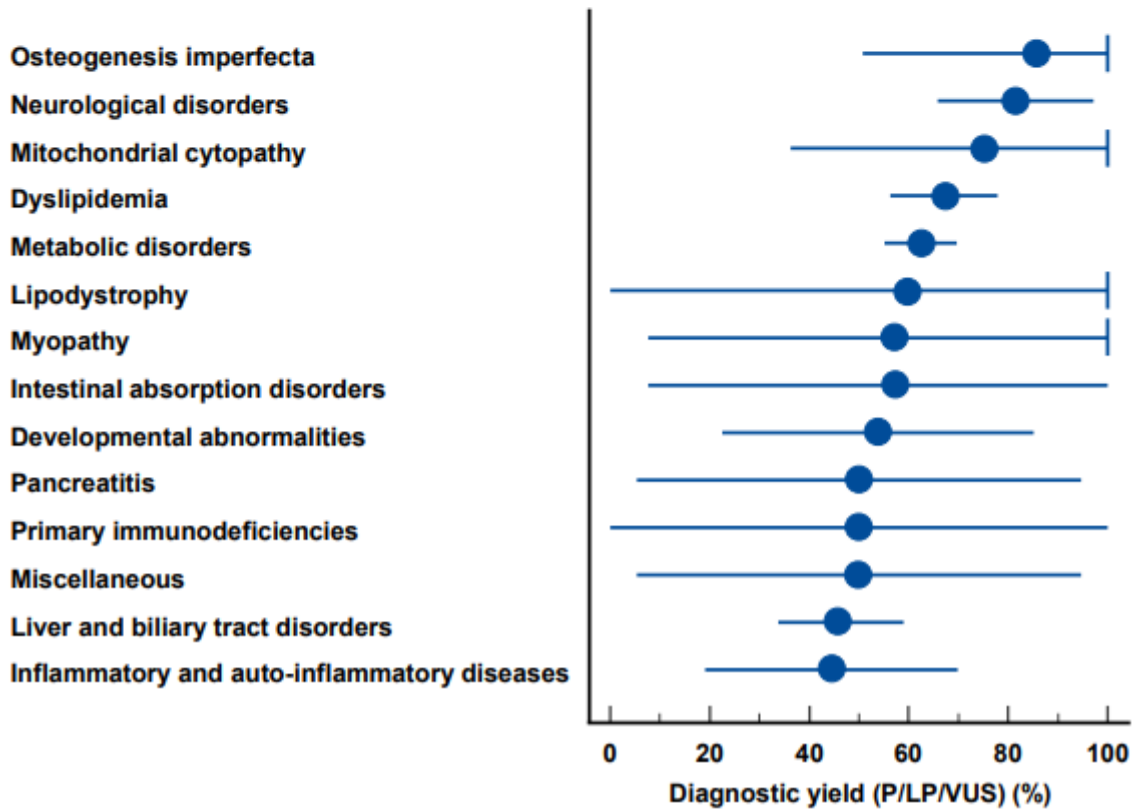
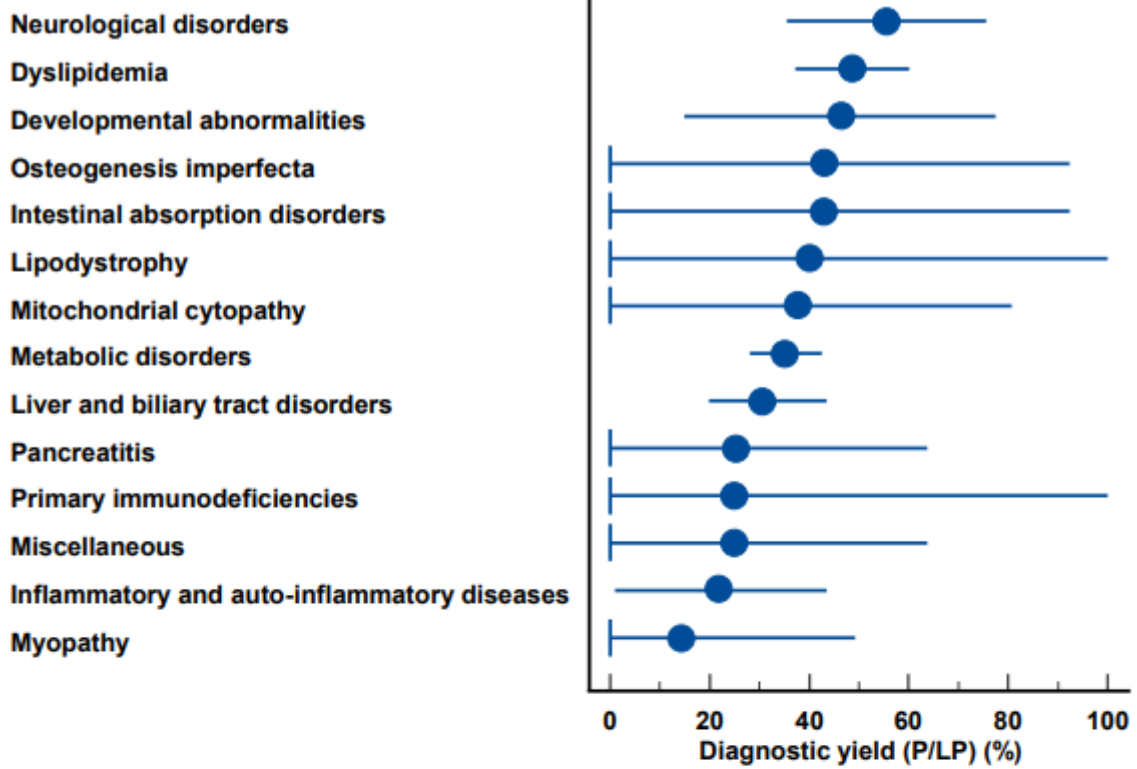


Figure 3.

IV) Discussion

Cette étude observationnelle rétrospective portant sur 603 sujets consécutifs (dont 426 cas index) a permis d'évaluer le rendement diagnostique du CES comme stratégie diagnostique de 1^{ère} ligne chez des patients présentant un phénotype mendélien dans la pratique courante du laboratoire de BBMN du CHRU de Nancy.

Celle-ci a également permis de mettre en évidence des facteurs prédictifs de la mise en évidence de variants pathogènes ou probablement pathogène chez ces patients, ainsi que d'observer l'impact du résultat du NGS sur le traitement des patients

A) OBJECTIF PRINCIPAL : EFFICIENCE DIAGNOSTIQUE

La plupart des études publiées qui évaluent le rendement diagnostique du NGS se focalisent sur un type de pathologie particulier : comme par exemple les troubles neurologiques (124,125) (déficiences intellectuelles, TSA, épilepsie), les déficits immunitaires (126), etc.

Dans la littérature, peu d'études rapportent l'efficacité diagnostique du séquençage haut débit utilisé en 1^{ère} ligne pour le diagnostic de patient présentant des phénotypes mendéliens. Les principales études et leurs résultats sont rapportés dans le tableau VI.

Tableau VI : Articles de la littérature rapportant l'efficacité diagnostique chez des patients présentant un phénotype Mendélien

Année	Pays	NGS : technologie utilisée	Nombre de patients	Efficacité diagnostique	Reference
2021	France	CES	426	160 (38%)	Alix et al.
2021	France	CES	128	70 (55%)	(78)
2021	Australia	CES	204	80 (39%)	(85)
2021	Singapore	CES	216	87 (40%)	(79)
2021	Canada	WES/WGS	47	8 (17%)	(90)
2021	US	WGS	99	16 (16%)	(92)
2021	Singapore	WES WGS	172 24	65 (38%) 8 (33%)	(91)
2021	Germany; Saudi Arabia	WGS	1007	212 (21%)	(93)
2020	Australia	WGS	92	38 (42%)	(94)
2020	Brazil	WES	500	158 (32%)	(86)
2020	UAE	CES	51	21 (41%)	(80)
2020	Japan	CES	360	171 (48%)	(81)
2020	Canada	WGS	49 familles	11 (22%)	(95)
2020	Pakistan	CES	349	214 (61%)	(82)
2019	25 nations	CES	3315	307 (9%)	(84)
2019	US	WES	66	23 (35%)	(87)
2019	China	WES	98	72 (74%)	(88)
2019	Qatar	CES	509	246 (48%)	(83)
2017	Australia	WES	44	23 (52%)	(89)

Parmi les 426 cas index de la cohorte de Nancy, l'efficacité diagnostique est de 37,6% (160/426) en ne considérant que les patients présentant au moins un variant de classe 4 ou 5, et de 60,8% (259/426) en incluant les variants de classe 3. Ces résultats sont cohérents avec ceux de la littérature où l'efficacité fluctue entre 9 et 74%. Ces grandes fluctuations peuvent s'expliquer par la grande variété des phénotypes explorés dans les différents centres, et par la rareté des pathologies recherchées.

L'efficacité diagnostique de notre cohorte est favorisée par certaines familles d'indications (troubles neurologiques, dyslipidémie, anomalies développementales) pour lesquelles l'efficacité diagnostique est excellente, tandis que d'autres indications ont des résultats plus modestes (myopathies, maladies inflammatoires).

Chez les 259 patients pour lesquels au moins un variant de classe 3 a été détecté, 349 variants de classe 3, 4 et 5 ont été mis en évidence sur 154 gènes différents répartis sur tout le génome. Ces mutations sont de natures variées, mais sont très majoritairement dues à des faux-sens (Figure 13).

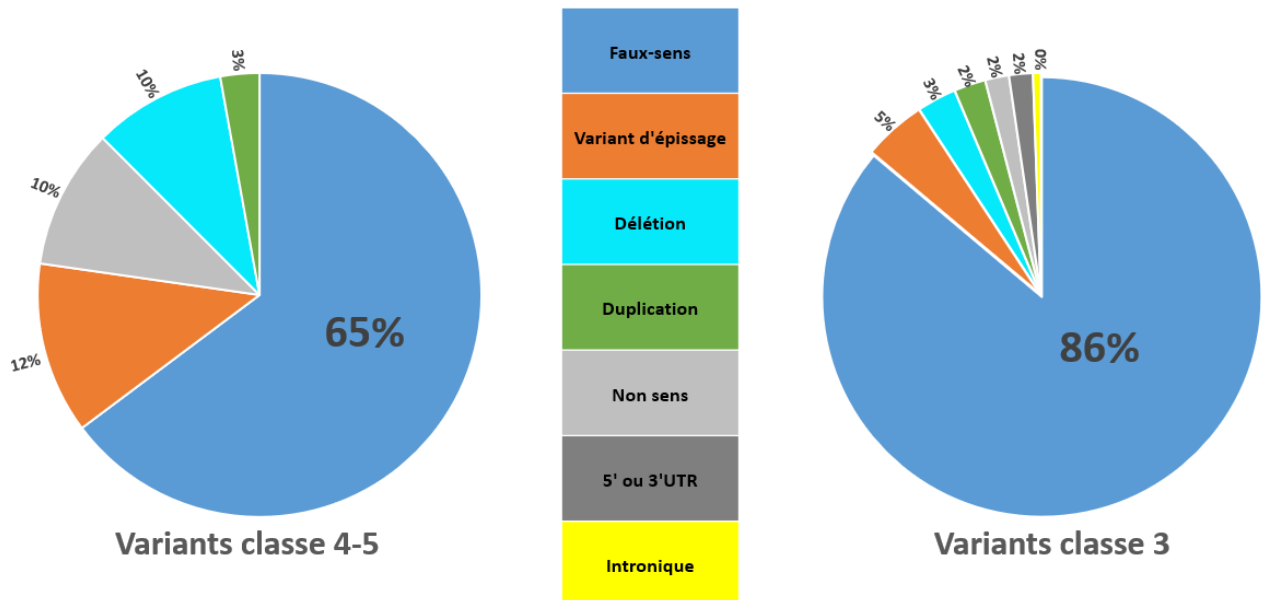


Figure 13 : Répartition de la nature des mutations rapportées

Cette répartition est comparable à celle retrouvée dans la littérature, les substitutions étant l'anomalie la plus fréquemment retrouvée, suivie des délétions, puis des insertions (127).

La plupart de ces variations ont été retrouvées à l'état hétérozygote, mais 5 d'entre elles sont à l'état hémizygotique et 34 à l'état homozygote. Certains variants sont associés à des transmissions dominantes, et peuvent donc facilement être considérés comme la cause de la pathologie (ex : variation de MEFV et fièvre méditerranéenne OMIM 134610). D'autres variations, comme celles associées à des transmissions récessives, rapportées avec des effets dominants négatifs ou dont le mécanisme physiopathologique est plus complexe (digénique ou polygénique) sont plus difficilement directement imputables.

B) OBJECTIF SECONDAIRE : PREDICTEUR DE LA MISE EN EVIDENCE D'UN VARIANT DE CLASSE 4 OU 5

A ce jour, aucune étude n'avait pour objectif d'identifier des facteurs prédictifs de l'utilité clinique du CES pour établir un diagnostic moléculaire chez des patients présentant des phénotypes mendéliens variés.

Dans cette étude, une approche systématique de tous les paramètres recueillis pour les patients inclus dans l'étude a permis de mettre en évidence 3 facteurs significativement associés à la découverte d'un variant de classe 4 ou 5. Ceux-ci sont :

- Prescription de l'exome clinique réalisée par un service pédiatrique : 2/3 des patients présentant des maladies rares sont des enfants (8). Leur phénotype, parfois sévère, est généralement cliniquement visible dès les premières années de vie.
- Suspicion de pathologie lysosomale : ces pathologies sont dues à des déficits enzymatiques qui entraînent des dysfonctions cellulaires. Dans notre cohorte, 6 patients présentaient un phénotype compatible avec ce type de dysfonction (2 suspicions de maladie de Fabry, 2 de maladie de Gaucher, 1 de Niemann Pick et 1 de maladie de Krabbe). Pour chacune de ces pathologies, le diagnostic passe par le dosage d'enzyme spécifique permettant d'orienter grandement l'analyse bio-informatique.
- Suspicion d'hypercholestérolémie familiale : il s'agit de pathologies dont on dispose de marqueurs phénotypiques facile à doser en routine (bilan lipidique). Malgré leur caractère souvent polygénique et environnemental (128), l'anamnèse approfondie du patient et de sa famille par un médecin spécialisé permet l'obtention de rendement diagnostique satisfaisant.

C) OBJECTIF SECONDAIRE : PREDICTEURS DE LA CONFIRMATION DU DIAGNOSTIC SUSPECTE

Le deuxième objectif secondaire était de mettre en évidence des facteurs prédictifs de la confirmation du diagnostic suspecté.

Dans notre cohorte, deux facteurs sont significativement associés à la confirmation du diagnostic émis par le prescripteur : la suspicion d'hypercholestérolémie familiale, et l'inclusion d'au moins un apparenté dans l'analyse NGS.

Cette analyse renforce donc le fait que comparer les résultats du cas index à des apparentés (notamment à ses parents) facilitent l'interprétation des différents variants en permettant de confirmer ou d'infirmer le caractère de novo d'un variant, ou d'en effectuer la co-ségrégation. Ces informations sont extrêmement utiles pour les variants dont la signification est inconnue (classe 3) et permettent parfois de les reclasser en variants (probablement) pathogènes ou (probablement) bénins (129).

D) OBJECTIF SECONDAIRE : THERAPIE GUIDEE PAR LE CES

Enfin, le dernier objectif était d'observer l'impact du résultat du NGS sur le traitement des patients. Parmi les 426 patients, ces informations étaient disponibles chez 325 patients (Figure 14). Des adaptations thérapeutiques ont été réalisées chez 19,1% (62/325) des patients. Pour les autres patients (263/325), le traitement n'a pas été adapté suite au NGS : soit parce qu'il n'a pas permis de diagnostiquer la pathologie, soit parce que le traitement n'était pas nécessaire / déjà mis en œuvre sur la base du phénotype / ou parce qu'il n'y avait pas de traitement spécifique existant.

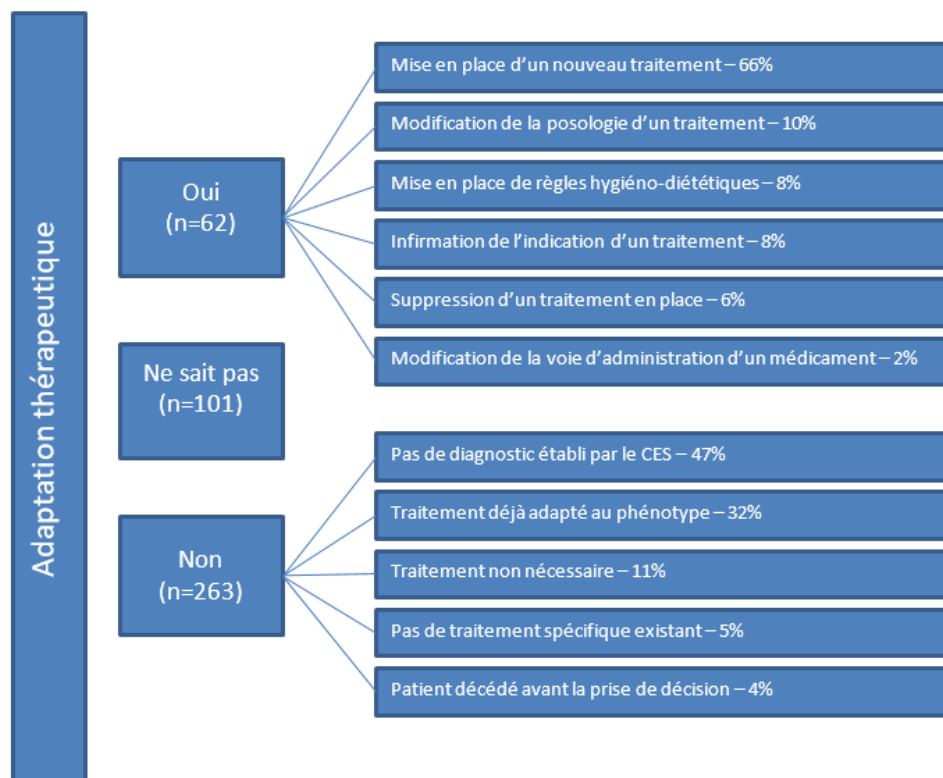


Figure 14 : Adaptation thérapeutique due au CES

Trois facteurs ont montré une association significative avec la modification du traitement :

- Confirmation du diagnostic suspecté : la confirmation du diagnostic suspecté permet d'adapter plus facilement le traitement à l'étiologie moléculaire.
- Prescription provenant du centre de référence des maladies du métabolisme : la prise en charge du patient par un centre de référence, spécialisé dans le diagnostic de maladie héréditaire favorise la modification rapide du traitement et son adaptation précise en fonction du génotype.

- Nombre total de variant rapporté par patient : Un nombre important de variant rapporté peut être en faveur de pathologies complexes, digéniques ou multigénique, qui nécessite un traitement adapté au vu du contexte génétique global.

E) LIMITES ET PERSPECTIVES

Cette étude permet de décrire l'efficacité clinique du CES en tant que test diagnostique génétique de 1ère ligne. L'efficacité clinique a été évaluée en termes de rendement diagnostic, de confirmation du diagnostic initialement proposé par le clinicien, et d'adaptation thérapeutique que le CES a impliquée.

Certaines limites de l'étude doivent tout de même être prises en compte dans l'interprétation de nos résultats.

- Premièrement, nous rapportons les résultats d'une étude rétrospective monocentrique qui doivent être confirmés dans des cohortes indépendantes. Les études rétrospectives génèrent de nombreux biais du fait de l'accès parfois compliqué aux informations à recueillir. Les données manquantes sont généralement nombreuses et peuvent générer des biais statistiques.
- Deuxièmement, étant donné la connaissance en constante évolution dans le domaine de la biologie moléculaire, les CES non contributifs peuvent être réévalués et réanalysés. Certains variants peuvent alors être amenés à être reclassés. L'ACMG recommande par ailleurs une réanalyse régulière (tous les 2 ans) des variants et des cas (130). En pratique, cette réévaluation est compliquée car chronophage, et ne s'intègre pas facilement dans la routine du laboratoire. Les réanalyses, au CHRU de Nancy, sont ainsi réalisées à la demande des cliniciens, notamment en cas de nouveaux symptômes à intégrer à l'analyse.
- Troisièmement, l'efficacité du CES est sous-optimale pour détecter les mutations introniques ou régulatrices profondes ainsi que les variations du nombre de copies (CNV). Le WGS a présenté un meilleur rendement de diagnostic par rapport aux panels de séquençage (123).

Malgré ces limitations, cette étude rapporte l'une des plus importantes séries européennes de patients consécutifs présentant un phénotype mendélien pour évaluer le rendement diagnostique d'un large panel de CES dans la pratique de routine.

Deuxièmement, le délai médian entre le test CES et le rendu des résultats est d'environ 5 mois. Il s'agit d'un temps relativement court, compatible avec la routine clinique, qui comprend les étapes techniques et le délai d'analyse d'un panel génique large. Le délai minimum observé est de 17 jours. En cas d'urgence diagnostic, le délai de rendu des résultats

peut donc être drastiquement réduit afin d'être compatible avec une prise en charge efficace et adapté des différentes pathologies.

Troisièmement, l'analyse bio-informatique utilisée combine systématiquement des méthodes classiques de filtrage et d'annotation de variants avec un algorithme de calcul basé sur le phénotype SVS-3PhoRank, conformément à la stratégie de diagnostic moléculaire choisie par le laboratoire, basée sur des captures de plusieurs milliers de gènes associés à des pathologies. Ce type d'analyse permet d'augmenter la sensibilité des tests, en recherchant les variants responsables par plusieurs approches.

Différentes perspectives sont envisageables grâce à cette étude. La création d'une étude prospective permettant d'évaluer l'utilité clinique de la CES permettrait d'augmenter la puissance statistique et de confirmer les conclusions de cette étude. L'utilisation de critères d'inclusion prédéfinis permettrait ainsi d'obtenir des résultats plus robustes, sur une large population. De plus, un essai randomisé pourrait être conçu pour comparer les différentes stratégies diagnostiques existantes (CES, WES, WGS, étude ciblée, ...).

Notre étude confirme l'importance de la prescription des CES en interaction étroite entre les médecins référents ayant une expertise dans les troubles génétiques et les biologistes de médecine moléculaire. Un dialogue clinico-biologique bien structuré permet d'optimiser la prise en charge du patient. Notre étude a montré que lorsque la prescription de cet examen biologique est effectuée un centre de référence expert, le traitement du patient est réadapté à l'anomalie moléculaire découverte. La création de centres experts, spécialisé dans la prise en charge des pathologies moléculaires, et la mise en place de réunion clinico-biologiques pluridisciplinaires permettrait de rationaliser la prescription, et ainsi d'optimiser la prise en charge globale du patient.

Bibliographie

1. Plaiasu V, Nanu M, Matei D. Rare Disease Day – at a glance. *Mædica*. janv 2010;5(1):65-6.
2. Commissioner O of the. Rare Diseases at FDA [Internet]. FDA. FDA; 2020 [cité 28 août 2021]. Disponible sur: <https://www.fda.gov/patients/rare-diseases-fda>
3. Health AGD of. What we're doing about rare diseases [Internet]. Australian Government Department of Health. Australian Government Department of Health; 2020 [cité 28 août 2021]. Disponible sur: <https://www.health.gov.au/health-topics/chronic-conditions/what-were-doing-about-chronic-conditions/what-were-doing-about-rare-diseases>
4. Hayashi S, Umeda T. 35 years of Japanese policy on rare diseases. *The Lancet*. 13 sept 2008;372(9642):889-90.
5. Pogue RE, Cavalcanti DP, Shanker S, Andrade RV, Aguiar LR, de Carvalho JL, et al. Rare genetic diseases: update on diagnosis, treatment and online resources. *Drug Discov Today*. 1 janv 2018;23(1):187-95.
6. RESERVES IUTD. Orphanet [Internet]. [cité 1 juin 2021]. Disponible sur: <http://www.orpha.net/consor/www/cgi-bin/index.php?lng=FR>
7. Rode J. Rare Diseases: Understanding this Public Health Priority. *Rare Dis*. 2005;14.
8. Plan national maladies rares 2005-2008 [Internet]. Ministère de la santé et des solidarités; [cité 31 août 2021]. Disponible sur: https://solidarites-sante.gouv.fr/IMG/pdf/plan_national_maladies_rares_2005-2008.pdf
9. Dodge JA, Chigladze T, Donadieu J, Grossman Z, Ramos F, Serlicorni A, et al. The importance of rare diseases: from the gene to society. *Arch Dis Child*. sept 2011;96(9):791-2.
10. Lutte contre les maladies génétiques : Rapport du Secrétariat [Internet]. OMS; 2005 avr [cité 4 juill 2021]. Report No.: EB116/3 Point 4.1. Disponible sur: https://apps.who.int/gb/archive/pdf_files/EB116/B116_3-fr.pdf
11. Drépanocytose [Internet]. Inserm - La science pour la santé. [cité 4 juill 2021]. Disponible sur: <https://www.inserm.fr/information-en-sante/dossiers-information/drepanocytose>
12. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 1 sept 2001;17(9):502-10.
13. Orphaschool. Transmission des maladies génétiques [Internet]. [cité 28 mai 2021]. Disponible sur: <https://www.orpha.net/orphaschool/formations/transmission/ExternData/InfoTransmission-Dreamweaver/Transmission.pdf>
14. OMIM - Online Mendelian Inheritance in Man [Internet]. [cité 28 mai 2021]. Disponible sur: <https://www.omim.org/>
15. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*. 4 mai 2017;100(5):695-705.
16. de Souza N. The ENCODE project. *Nat Methods*. nov 2012;9(11):1046.

17. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. juill 2020;583(7818):699-710.
18. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. avr 2022;376(6588):44-53.
19. Règles de bonnes pratiques en génétique constitutionnelle à des fins médicales (Hors diagnostic prénatal) [Internet]. Haute Autorité de Santé. 2013 [cité 14 août 2021]. Disponible sur: https://www.has-sante.fr/jcms/c_1360718/fr/regles-de-bonnes-pratiques-en-genetique-constitutionnelle-a-des-fins-medicales-hors-diagnostic-prenatal
20. Règlement (CE) n° 141/2000 du Parlement européen et du Conseil, du 16 décembre 1999, concernant les médicaments orphelins [Internet]. Journal officiel n° L 018 du 22/01/2000 p. 0001 - 0005; OPOCE; [cité 31 août 2021]. Disponible sur: <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32000R0141&from=FR>
21. DGOS. Les maladies rares [Internet]. Ministère des Solidarités et de la Santé. 2021 [cité 28 août 2021]. Disponible sur: <https://solidarites-sante.gouv.fr/soins-et-maladies/prises-en-charge-specialisees/maladies-rares/article/les-maladies-rares>
22. Ministère de la Santé et des Solidarités. Plan national maladies rares 2011-2014 [Internet]. [cité 3 sept 2021]. Disponible sur: https://solidarites-sante.gouv.fr/IMG/pdf/plan_national_maladies_rares_2011-2014.pdf
23. Ministère de la Santé et des Solidarités. Plan national maladies rares 2018-2022 [Internet]. [cité 3 sept 2021]. Disponible sur: https://solidarites-sante.gouv.fr/IMG/pdf/pnmr_3_v25-09pdf.pdf
24. Ministère de la Santé et des Solidarités. Rapport d'activité 2019 - Filières de Santé - Maladies Rares [Internet]. 2020 oct [cité 3 sept 2021]. Disponible sur: https://solidarites-sante.gouv.fr/IMG/pdf/ra_2019_vf_2021-02-22.pdf
25. Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*. 31 juill 1987;50(3):509-17.
26. Rousseau F, Laflamme N. Génétique moléculaire humaine : des maladies monogéniques aux maladies complexes. *médecine/sciences*. 1 oct 2003;19(10):950-4.
27. Bouzigon E, Nadif R, Le Moual N, Dizier MH, Aschard H, Boudier A, et al. [Genetic and environmental factors of asthma and allergy: Results of the EGEA study]. *Rev Mal Respir*. oct 2015;32(8):822-40.
28. Sabatelli M, Conte A, Zollino M. Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clin Genet*. mai 2013;83(5):408-16.
29. von Coelln R, Shulman LM. Clinical subtypes and genetic heterogeneity: of lumping and splitting in Parkinson disease. *Curr Opin Neurol*. déc 2016;29(6):727-34.
30. Rao DC. Genetic dissection of complex traits: an overview. *Adv Genet*. 2001;42:13-34.
31. Aviesan - L'Alliance pour les sciences de la vie et de la santé. France Médecine Génomique 2025 [Internet]. [cité 7 sept 2021]. Disponible sur:

https://www.gouvernement.fr/sites/default/files/document/document/2016/06/22.06.2016_remise_du_rapport_dyves_levy_-_france_medecine_genomique_2025.pdf

32. Assistance Publique – Hôpitaux de Paris. Sélection du projet SeqOIA porté par l'Assistance Publique – Hôpitaux de Paris, l'Institut Curie et Gustave Roussy pour la constitution d'une plateforme nationale de séquençage génomique très haut débit dans le cadre du Plan France Médecine Génomique 2025 [Internet]. 2017 [cité 7 sept 2021]. Disponible sur: <https://www.gustaveroussy.fr/sites/default/files/2017-07-17-cp-sequoia.pdf>
33. AURAGEN | Auvergne Rhône-Alpes Génomique [Internet]. AURAGEN | Auvergne Rhône-Alpes Génomique. [cité 7 sept 2021]. Disponible sur: <https://www.auragen.fr/>
34. Présentation de la Filière G2M [Internet]. Filière G2M groupement des maladies héréditaires du métabolisme. [cité 28 déc 2021]. Disponible sur: <http://www.filiere-g2m.fr/filiere/presentation-de-la-filiere-g2m/>
35. Domenighetti L. La liste des hôpitaux labellisés Centres de Référence et Centres de Compétence pour les Maladies Héréditaires du Métabolisme [Internet]. Les Feux Follets. [cité 3 sept 2021]. Disponible sur: <https://phenylcetonurie.org/la-liste-des-hopitaux-labellises-centres-de-reference-et-centres-de-competence-pour-les-maladies-hereditaires-du-metabolisme-11889/>
36. Hennekam RCM, Biesecker LG. Next-generation sequencing demands next-generation phenotyping. *Hum Mutat.* mai 2012;33(5):884-6.
37. PCR [Internet]. Biomnigene. [cité 4 août 2021]. Disponible sur: <https://www.biomnigene.fr/fr/nos-solutions/pcr.html>
38. Illumina. The \$1,000 Genome [Internet]. 2014 [cité 15 août 2021]. Disponible sur: <https://emea.illumina.com/content/dam/illumina-marketing/documents/company/featured-articles/the-1000-dollar-genome.pdf>
39. Sanger F, Donelson JE, Coulson AR, Kössel H, Fischer D. Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA. *Proc Natl Acad Sci U S A.* avr 1973;70(4):1209-13.
40. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 25 mai 1975;94(3):441-8.
41. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* déc 1977;74(12):5463-7.
42. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, et al. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Investig Off Publ Am Assoc Vet Lab Diagn Inc.* nov 2020;32(6):767-75.
43. Beck TF, Mullikin JC, Biesecker LG. Systematic Evaluation of Sanger Validation of NextGen Sequencing Variants. *Clin Chem.* avr 2016;62(4):647-54.
44. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* oct 2004;431(7011):931-45.
45. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. *Nucleic Acids Res.* 28 janv 2015;43(Database issue):D662-9.

46. Stranneheim H, Lundeberg J. Stepping stones in DNA sequencing. *Biotechnol J*. sept 2012;7(9):1063-73.
47. Wu J, Dai W, Wu L, Wang J. SALP, a new single-stranded DNA library preparation method especially useful for the high-throughput characterization of chromatin openness states. *BMC Genomics*. 13 févr 2018;19(1):143.
48. Target Enrichment | NGS-based hybrid capture method [Internet]. [cité 5 août 2021]. Disponible sur: <https://emea.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/target-enrichment.html>
49. Illumina. Illumina Sequencing by Synthesis [Internet]. 2016 [cité 8 août 2021]. Disponible sur: <https://www.youtube.com/watch?v=fCd6B5HRaZ8>
50. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 6 nov 2008;456(7218):53-9.
51. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. janv 2010;11(1):31-46.
52. Human Phenotype Ontology [Internet]. [cité 13 août 2021]. Disponible sur: <https://hpo.jax.org/app/>
53. Article R1131-20 - Code de la santé publique - Légifrance [Internet]. [cité 14 août 2021]. Disponible sur: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000018615509/
54. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet*. mai 2018;19(5):253-68.
55. Lamoril J, Bogard M. Genomic medicine: the new way of thinking medicine Present and future – Part two. *Ann Biol Clin (Paris)*. janv 2014;72(1):25-48.
56. Crawford DC, Acuña JM, Sherman SL. FMR1 and the Fragile X Syndrome: Human Genome Epidemiology Review. *Genet Med Off J Am Coll Med Genet*. 2001;3(5):359-71.
57. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med*. 7 août 2017;141(11):1544-57.
58. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing | *Genetics in Medicine* [Internet]. [cité 15 août 2021]. Disponible sur: <https://www.nature.com/articles/gim2014122>
59. Bloch-Zupan A, Jamet X, Etard C, Laugel V, Muller J, Geoffroy V, et al. Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects. *Am J Hum Genet*. 9 déc 2011;89(6):773-81.
60. Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PLoS One*. 2013;8(11):e78496.

61. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? *Hum Genet.* mars 2016;135(3):359-62.
62. LaDuca H, Farwell KD, Vuong H, Lu HM, Mu W, Shahmirzadi L, et al. Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS ONE.* 2 févr 2017;12(2):e0170843.
63. Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, et al. New insights into the performance of human whole-exome capture platforms. *Nucleic Acids Res.* 23 juin 2015;43(11):e76.
64. Sun Y, Ruivenkamp CAL, Hoffer MJV, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-generation diagnostics: gene panel, exome, or whole genome? *Hum Mutat.* juin 2015;36(6):648-55.
65. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet.* 6 mai 2021;108(5):764-85.
66. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics.* oct 2015;13(5):278-89.
67. Zamariolli M, Di-Battista A, Moysés-Oliveira M, de Mello CB, de Paula Ramos MA, Liehr T, et al. Disruption of PCDH10 and TNRC18 Genes due to a Balanced Translocation. *Cytogenet Genome Res.* 2020;160(6):321-8.
68. Tilak MK, Botero-Castro F, Galtier N, Nabholz B. Illumina Library Preparation for Sequencing the GC-Rich Fraction of Heterogeneous Genomic DNA. *Genome Biol Evol.* 1 févr 2018;10(2):616-22.
69. Moreno-Cabrera JM, del Valle J, Castellanos E, Feliubadaló L, Pineda M, Brunet J, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur J Hum Genet.* déc 2020;28(12):1645-55.
70. Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. *Semin Thromb Hemost.* oct 2019;45(7):661-73.
71. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics.* oct 2016;14(5):265-79.
72. Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinforma Oxf Engl.* 1 juill 2019;35(13):2193-8.
73. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet TIG.* sept 2018;34(9):666-81.
74. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med Off J Am Coll Med Genet.* mai 2015;17(5):405-24.
75. ACQUAVIVA-BOURDAIN C, BAERT-DESURMONT S, CAPUTO S, CHATRON N, COULET F, DOCO-FENZY M, et al. Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS. Réseau NGS-Diagnostic; 2021 févr. Report No.: 2.

76. Gorokhova S. Systématisation d'interprétation des données de séquençage à haut débit : application au diagnostic des maladies neuromusculaires [DES Génétique Médicale]. Faculté des sciences médicales et paramédicales de Marseille; 2020.
77. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet.* févr 2020;28(2):165-73.
78. Mergnac JP, Wiedemann A, Chery C, Ravel JM, Namour F, Guéant JL, et al. Diagnostic yield of clinical exome sequencing as a first-tier genetic test for the diagnosis of genetic disorders in pediatric patients: results from a referral center study. *Hum Genet.* 8 sept 2021;
79. Wei H, Lai A, Tan ES, Koh MJA, Ng I, Ting TW, et al. Genetic landscape of congenital disorders in patients from Southeast Asia: results from sequencing using a gene panel for Mendelian phenotypes. *Arch Dis Child.* janv 2021;106(1):38-43.
80. Mahfouz NA, Kizhakkedath P, Ibrahim A, El Naofal M, Ramaswamy S, Harilal D, et al. Utility of clinical exome sequencing in a complex Emirati pediatric cohort. *Comput Struct Biotechnol J.* 2020;18:1020-7.
81. Kosaki R, Kubota M, Uehara T, Suzuki H, Takenouchi T, Kosaki K. Consecutive medical exome analysis at a tertiary center: Diagnostic and health-economic outcomes. *Am J Med Genet A.* juill 2020;182(7):1601-7.
82. Cheema H, Bertoli-Avella AM, Skrahina V, Anjum MN, Waheed N, Saeed A, et al. Genomic testing in 1019 individuals from 349 Pakistani families results in high diagnostic yield and clinical utility. *NPJ Genomic Med.* 5 oct 2020;5(1):44.
83. Al-Dewik N, Mohd H, Al-Mureikhi M, Ali R, Al-Mesaifri F, Mahmoud L, et al. Clinical exome sequencing in 509 Middle Eastern families with suspected Mendelian diseases: The Qatari experience. *Am J Med Genet A.* juin 2019;179(6):927-35.
84. Groopman EE, Marasa M, Cameron-Christie S, Petrovski S, Aggarwal VS, Milo-Rasouly H, et al. Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med.* 10 janv 2019;380(2):142-51.
85. Jayasinghe K, Stark Z, Kerr PG, Gaff C, Martyn M, Whitlam J, et al. Clinical impact of genomic testing in patients with suspected monogenic kidney disease. *Genet Med Off J Am Coll Med Genet.* janv 2021;23(1):183-91.
86. Quaio CRDC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. *Am J Med Genet C Semin Med Genet.* déc 2020;184(4):955-64.
87. Reuter CM, Kohler JN, Bonner D, Zastrow D, Fernandez L, Dries A, et al. Yield of whole exome sequencing in undiagnosed patients facing insurance coverage barriers to genetic testing. *J Genet Couns.* déc 2019;28(6):1107-18.
88. Hong S, Wang L, Zhao D, Zhang Y, Chen Y, Tan J, et al. Clinical utility in infants with suspected monogenic conditions through next-generation sequencing. *Mol Genet Genomic Med.* juin 2019;7(6):e684.

89. Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, et al. Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions. *JAMA Pediatr.* 1 sept 2017;171(9):855-62.
90. Shickh S, Gutierrez Salazar M, Zakoore KR, Lázaro C, Gu J, Goltz J, et al. Exome and genome sequencing in adults with undiagnosed disease: a prospective cohort study. *J Med Genet.* avr 2021;58(4):275-83.
91. Bhatia NS, Lim JY, Bonnard C, Kuan JL, Brett M, Wei H, et al. Singapore Undiagnosed Disease Program: Genomic Analysis aids Diagnosis and Clinical Management. *Arch Dis Child.* janv 2021;106(1):31-7.
92. Brockman DG, Austin-Tse CA, Pelletier RC, Harley C, Patterson C, Head H, et al. Randomized prospective evaluation of genome sequencing versus standard-of-care as a first molecular diagnostic test. *Genet Med Off J Am Coll Med Genet.* sept 2021;23(9):1689-96.
93. Bertoli-Avella AM, Beetz C, Ameziane N, Rocha ME, Guatibonza P, Pereira C, et al. Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet EJHG.* janv 2021;29(1):141-53.
94. Yeung A, Tan NB, Tan TY, Stark Z, Brown N, Hunter MF, et al. A cost-effectiveness analysis of genomic sequencing in a prospective versus historical cohort of complex pediatric patients. *Genet Med Off J Am Coll Med Genet.* déc 2020;22(12):1986-93.
95. Costain G, Walker S, Marano M, Veenma D, Snell M, Curtis M, et al. Genome Sequencing as a Diagnostic Test in Children With Unexplained Medical Complexity. *JAMA Netw Open.* 1 sept 2020;3(9):e2018109.
96. 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med.* 11 nov 2021;385(20):1868-80.
97. Australian Genomics Health Alliance Acute Care Flagship, Lunke S, Eggers S, Wilson M, Patel C, Barnett CP, et al. Feasibility of Ultra-Rapid Exome Sequencing in Critically Ill Infants and Children With Suspected Monogenic Conditions in the Australian Public Health Care System. *JAMA.* 23 juin 2020;323(24):2503-11.
98. Gubbels CS, VanNoy GE, Madden JA, Copenheaver D, Yang S, Wojcik MH, et al. Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. *Genet Med Off J Am Coll Med Genet.* avr 2020;22(4):736-44.
99. Brunelli L, Jenkins SM, Gudgeon JM, Bleyl SB, Miller CE, Tvrdik T, et al. Targeted gene panel sequencing for the rapid diagnosis of acutely ill infants. *Mol Genet Genomic Med.* juill 2019;7(7):e00796.
100. Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, et al. Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *JAMA Pediatr.* 4 déc 2017;171(12):e173438.
101. Wu B, Kang W, Wang Y, Zhuang D, Chen L, Li L, et al. Application of Full-Spectrum Rapid Clinical Genome Sequencing Improves Diagnostic Rate and Clinical Outcomes in Critically Ill Infants in the China Neonatal Genomes Project. *Crit Care Med.* 1 oct 2021;49(10):1674-83.

102. Scholz T, Blohm ME, Kortüm F, Bierhals T, Lessel D, van der Ven AT, et al. Whole-Exome Sequencing in Critically Ill Neonates and Infants: Diagnostic Yield and Predictability of Monogenic Diagnosis. *Neonatology*. 2021;118(4):454-61.
103. Śmigiel R, Biela M, Szmyd K, Błoch M, Szmida E, Skiba P, et al. Rapid Whole-Exome Sequencing as a Diagnostic Tool in a Neonatal/Pediatric Intensive Care Unit. *J Clin Med*. 13 juill 2020;9(7):E2220.
104. Wang H, Lu Y, Dong X, Lu G, Cheng G, Qian Y, et al. Optimized trio genome sequencing (OTGS) as a first-tier genetic test in critically ill infants: practice in China. *Hum Genet*. avr 2020;139(4):473-82.
105. French CE, Delon I, Dolling H, Sanchis-Juan A, Shamardina O, Mégy K, et al. Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med*. mai 2019;45(5):627-36.
106. Oussalah A, Jeannesson-Thivisol E, Chéry C, Perrin P, Rouyer P, Josse T, et al. Population and evolutionary genetics of the PAH locus to uncover overdominance and adaptive mechanisms in phenylketonuria: Results from a multiethnic study. *EBioMedicine*. janv 2020;51:102623.
107. Wiedemann A, Chery C, Coelho D, Flayac J, Gueguen N, Desquiret-Dumas V, et al. Mutations in MTHFR and POLG impaired activity of the mitochondrial respiratory chain in 46-year-old twins with spastic paraparesis. *J Hum Genet*. janv 2020;65(2):91-8.
108. Guéant JL, Chéry C, Oussalah A, Nadaf J, Coelho D, Josse T, et al. APRDX1 mutant allele causes a MMACHC secondary epimutation in cblC patients. *Nat Commun*. 4 janv 2018;9(1):67.
109. Renard E, Chéry C, Oussalah A, Josse T, Perrin P, Tramoy D, et al. Exome sequencing of cases with neural tube defects identifies candidate genes involved in one-carbon/vitamin B12 metabolisms and Sonic Hedgehog pathway. *Hum Genet*. juill 2019;138(7):703-13.
110. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma*. 2013;43:11.10.1-11.10.33.
111. Singleton MV, Guthery SL, Voelkerding KV, Chen K, Kennedy B, Margraf RL, et al. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet*. 3 avr 2014;94(4):599-610.
112. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. avr 2014;11(4):361-2.
113. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. janv 2013;34(1):57-65.
114. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 15 avr 2015;24(8):2125-37.
115. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-81.

116. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinforma Oxf Engl*. 15 août 2015;31(16):2745-7.
117. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinforma Oxf Engl*. 1 mars 2015;31(5):761-3.
118. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 4 janv 2018;46(D1):D1062-7.
119. Nagelkerke N. A note on a general definition of the coefficient of determination. *Biometrika*. 1991;78(3):691-2.
120. Alfares A, Aloraini T, Subaie LA, Alissa A, Qudsi AA, Alahmad A, et al. Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet Med Off J Am Coll Med Genet*. nov 2018;20(11):1328-33.
121. Bonaventura J, Norambuena P, Tomašov P, Jindrová D, Šedivá H, Macek M, et al. The utility of the Mayo Score for predicting the yield of genetic testing in patients with hypertrophic cardiomyopathy. *Arch Med Sci AMS*. mai 2019;15(3):641-9.
122. Mariathasan S, Andrews KA, Thompson E, Challis BG, Wilcox S, Pierce H, et al. Genetic testing for hereditary hyperparathyroidism and familial hypocalciuric hypercalcaemia in a large UK cohort. *Clin Endocrinol (Oxf)*. oct 2020;93(4):409-18.
123. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med Off J Am Coll Med Genet*. avr 2018;20(4):435-43.
124. Stefanski A, Calle-López Y, Leu C, Pérez-Palma E, Pestana-Knight E, Lal D. Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. *Epilepsia*. janv 2021;62(1):143-51.
125. Canafoglia L, Franceschetti S, Gambardella A, Striano P, Giallonardo AT, Tinuper P, et al. Progressive Myoclonus Epilepsies: Diagnostic Yield With Next-Generation Sequencing in Previously Unsolved Cases. *Neurol Genet*. déc 2021;7(6):e641.
126. Elsink K, Huibers MMH, Hollink IHIM, Simons A, Zonneveld-Huijssoon E, van der Veken LT, et al. Implementation of Early Next-Generation Sequencing for Inborn Errors of Immunity: A Prospective Observational Cohort Study of Diagnostic Yield and Clinical Implications in Dutch Genome Diagnostic Centers. *Front Immunol*. 2021;12:780134.
127. Hanna N, Parfait B, Vidaud D, Vidaud M. Mécanismes et conséquences des mutations. *médecine/sciences*. 1 nov 2005;21(11):969-80.
128. Jaber M. Mise en place d'une étude rétrospective sur les patients atteints d'hyperlipidémies combinées familiales au sein du laboratoire de biologie médicale multisite (LBMMS) des Hospices Civils de Lyon (HCL) [Thèse d'exercice de Pharmacie]. [Lyon]: Université Claude Bernard Lyon 1; 2021.
129. Caputo S, Damiola F, Léoné M, et al. Etude de COségrégation familiale des VARiants nucléotidiques (COVAR) dans les gènes BRCA1/2 pour valider leur utilisation en conseil génétique. [Internet]. Assises de génétique humaine et médicale 2018; 2018 [cité 7 janv 2022]. Disponible


sur:

https://www.researchgate.net/publication/323546714_Etude_de_COsegregation_familiale_des_VARIants_nucleotidiques_COVAR_dans_les_genes_BRCA12_pour_valider_leur_utilisation_en_conseil_genetique

130. Deignan JL, Chung WK, Kearney HM, Monaghan KG, Rehder CW, Chao EC. Points to Consider in the Reevaluation and Reanalysis of Genomic Test Results: A Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med Off J Am Coll Med Genet.* juin 2019;21(6):1267-70.

Annexes

Annexe 1 : Modèle de consentement utilisé au CHRU de Nancy

	Formulaire BIOC-BM-Consentement pour Étude de Génétique Constitutionnelle	Référence	FORM-01615
		Version	02 Page 1 sur 1
		Applicable le	21/12/2020

FORMULAIRE DE CONSENTEMENT POUR ÉTUDE DE GÉNÉTIQUE CONSTITUTIONNELLE Laboratoire de Biochimie Biologie Moléculaire Nutrition Centre de Référence des Maladies Héréditaires du Métabolisme

Conformément aux articles R.1131-4 et R.1131-5 du code de la santé publique (décret n°2008-321 du 4 avril 2008 – arrêté du 27 mai 2013). (Établi en 3 exemplaires : Le 1^{er} remis à l'intéressé(e) le 2^e reste dans le dossier du patient et le 3^e est envoyé au laboratoire avec le prélèvement).

Je soussigné(e), Nom : _____ Prénom : _____ Date de naissance : ___/___/_____	Ou agissant en qualité de parent/ représentant légal de l'enfant mineur ou de la personne majeure sous tutelle ⁽¹⁾ Nom : _____ Prénom : _____ Date de naissance : ___/___/_____
--	---

Certifie avoir été informé par le Docteur / Professeur⁽¹⁾ Nom : _____ Prénom : _____ des raisons et des conditions de l'étude génétique me concernant ou concernant mon enfant mineur et portant sur :

Afin :

- de confirmer ou d'infirmier le diagnostic d'une maladie génétique en relation avec mes symptômes.
- de confirmer ou d'infirmier le diagnostic pré-symptomatique d'une maladie génétique.
- d'identifier un statut de porteur sain (recherche d'hétérozygote).
- d'évaluer ma susceptibilité génétique à une maladie ou à un traitement médicamenteux.

Je reconnais également avoir été informé(e) :

- que les résultats obtenus ont un caractère strictement confidentiel.
- que les examens seront réalisés par un laboratoire spécialisé et agréé.
- que je pourrais avoir connaissance, par l'intermédiaire du médecin prescripteur, des résultats ainsi que toute information confidentielle me concernant ou concernant mon enfant mineur.
- que les données nominatives me concernant peuvent faire l'objet d'un traitement automatisé. Le droit d'accès et de rectification prévu par la Loi informatique et Libertés s'exerce à tout moment auprès du Directeur des Affaires Générales de l'hôpital où le prélèvement a été effectué. L'accès aux informations de nature médicale peut être direct ou par l'intermédiaire du médecin de mon choix.

Donne mon accord pour l'enregistrement des données médicales nécessaires à ces examens.

Consens à effectuer ou à faire effectuer un prélèvement :

Sanguin Tissulaire Autre : _____

Chez moi

Chez mon enfant mineur ou la personne majeure sous tutelle concernée

Déclare que si une partie du prélèvement reste inutilisé après l'examen, j'accepte qu'elle soit conservée et intégrée, le cas échéant, dans un projet à des fins de recherche scientifique dans le laboratoire de Biochimie Biologie Moléculaire Nutrition du CHRU de Nancy (Pr. J.L. Guéant). Dans ce cas, l'ensemble des données médicales me concernant seront protégées grâce à une anonymisation totale. En conséquence, je suis conscient que ces études scientifiques effectuées ne seront sans aucun bénéfice ni préjudice pour moi.

Fait à : _____ le : ___/___/_____	Signature du patient adulte ou du représentant légal de l'enfant mineur ou de la personne majeure sous tutelle :
--	---

(1) Rayer la mention inutile

ATTESTATION DE CONSULTATION


Préalable à la réalisation des examens des caractéristiques génétiques d'une personne à des fins médicales
Je soussigné, **Docteur / Professeur** ⁽¹⁾ _____, conformément aux articles R.1131-4 et R.1131-5 du code de la santé publique, certifie avoir reçu en consultation ce jour le(la) patient(e) susnommé(e) afin de lui apporter les informations sur les caractéristiques de la maladie explorée, des moyens de la détecter, des possibilités de prévention et de traitement.

Fait à : _____ **le :** ___/___/_____ **Signature et tampon du médecin :**

(1) Rayer la mention inutile

CHRU de NANCY – Hôpitaux de Brabois – Rue du Morvan – 54511 VANDOEUVRE-LES-NANCY CEDEX, Tél : +33 3 83 85 85 85

Annexe 2 : Formulaire de compte-rendu des NGS prodigués dans le laboratoire BBMN du CHRU de Nancy

	Formulaire	Référence	FORM-02677
	BIOC-BM-Compte-rendu Diagnostic moléculaire par séquençage haut débit (Analyse des régions exoniques d'un panel de 4813 gènes, TruSight One)	Version	03 Page 1 sur 1
		Applicable le	15/10/2020



CENTRE HOSPITALIER REGIONAL UNIVERSITAIRE DE NANCY
POLE LABORATOIRES – LABORATOIRE DE BIOLOGIE MEDICALE
 Responsable du laboratoire : Dr Patricia FRANCK



Biochimie – Biologie Moléculaire – Nutrition et
 Centre de Référence des Maladies Héréditaires du Métabolisme

Pr Jean-Louis GUEANT
 Tél : 03 83 15 44 84 – 44 83
 Chef de structure - Resp UF
 jl.gueant@chru-nancy.fr

Pr Bernard NAMOUR
 Tél : 03 83 15 34 56
 Responsable UF
 b.namour@chru-nancy.fr

Pr Rosa-Maria RODRIGUEZ-GUEANT
 Tél : 03 83 15 36 30
 rm.rodriguez@chru-nancy.fr

Dr Isabelle AIMONE-GASTIN
 Tél : 03 83 15 35 13
 i.gastin@chru-nancy.fr

Pr Abderrahim OUSSALAH
 Tél : 03 83 15 36 29
 a.oussalah@chru-nancy.fr

UF Biologie Moléculaire : 03 83 15 36 71

Diagnostic moléculaire par séquençage haut débit (Analyse des régions exoniques d'un panel de 4813 gènes, TruSight One)

<p>NOM : Prénom : Sexe : Date de naissance :</p> <p>Statut de l'individu : N° IPP : N° Venue : Code Laboratoire :</p>	<p>Indication du test :</p> <p>Nature du prélèvement : Matériel biologique :</p> <p>Date du prélèvement : Date de réception : Date du résultat :</p>
---	--

PRESCRIPTEUR :

TECHNIQUES :
 Le kit de capture TruSight One sequencing panel (Illumina) est utilisé pour cibler les régions exoniques de 4813 gènes (y compris les jonctions exon/intron). Une fois isolées, les régions cibles du génome sont séquençées après amplification clonale sur le système MiSeq (Illumina®) sur une longueur de lecture de 150 pb en sens et en anti-sens. Les séquences obtenues sont alignées et analysées en comparaison avec la séquence référence du génome humain version GRCh37 à l'aide de BWA (0.7.9a) et SAMTools (0.1.18). L'annotation des variants est effectuée avec la suite d'outils GATK (v1.6-23). Les variants sont décrits selon les recommandations HGVS (le A de l'ATG initiateur de la traduction, a été choisi comme n°1 de la référence). L'annotation des transcrits de gènes est basée sur RefSeq Genes 105v2, NCBI.

RESULTATS :

Gène	Description du variant	Génotype du patient	Fréquence de l'allèle alternatif ¹	Nomenclature HGVS ² (rs)	Prédiction ³

1. La fréquence de l'allèle alternatif est basée sur l'interrogation des bases de données internationales.

2. Les variants sont décrits selon la nomenclature HGVS (<https://varnomen.hgvs.org>)

3. Selon les recommandations d'homogénéisation de l'interprétation de variants de séquence générés par analyses en NGS (Réseau NGS-Diag, sur la base des recommandations ACMG-AMP, Richards *et al.*, Genetics in Medicine 2015), dans leur version 1, dans les limites des connaissances actuelles (www.anpgm.fr/index.php/recommandations-professionnelles). Classe 1 : bénin ; Classe 2 : probablement bénin ; Classe 3 : signification inconnue ; Classe 4 : probablement pathogène ; Classe 5 : pathogène.

Pr JL. GUÉANT

Pr RM. RODRIGUEZ-GUÉANT

Pr B. NAMOUR

Dr I. AIMONE-GASTIN

Pr A. OUSSALAH

CLCN5, TSEN2, SGCA, MAGI2, NLRP2, FLT3, NELL1, PCDH19, TSP0, ADORA1, HES6, UCP1, POLE2, FADD, FGF9, FOXA1, IL2, VNN1, CACNA1A, SPG21, CDH1, OSRI, GRIN1, ITPA, EMC4, EDA, CHRNA4, HSPA5, SERPINB6, TNFSF13B, CMPK1, SCN2A, WNT7A, UPK3A, FIP1L1, SSH1, RHOB, ENSA, H6PD, GNAQ, XPA, MMAA, PRB1, APOBEC3H, AMH, ZNF526, GNPTAB, IRF5, ACAT2, GNMT, HADH, TRPM1, PPP2R1A, CCKAR, TMEM173, COL2A1, ARL14EP, RPA1, TROVE2, IFI30, PLEKHM1, GLUL, MSX1, WNT10A, HK2, RDH12, MICB, EPHX1, SLC30A10, DLX3, AIMP1, AP4S1, SLC27A4, AVPR1A, GJB2, SOX8, ATXN8OS, ICAM1, MTHFS, ZNF674, BRSK2, CD19, SCARB2, SLC20A2, CRBN, STK4, SLC16A3, DLX5, ADK, CELSR1, TNNT3K, CHRND, THBS2, NEXN, CA2, FOXF1, TPO, LTF, TT2, TAS1R1, ETS1, FGD1, AGRP, HSP90B1, SMOC1, TMEM138, GSTA3, HOXB6, USP26, CD81, MYH7, TCIRG1, SH2B3, CYP2W1, TMCO1, PARL, BICC1, TRRAP, RYR2, GRN, TRIM32, HOXA4, PTS, ADCY6, SLC36A2, DDC, NXF3, NPR2, PROM1, PRLR, ZFYVE26, GDF5, SNX3, CHIA, CATSPER3, TBX20, INMT, ADSL, FTSJ1, DST, BAP1, NPSR1, GLRX5, MSH2, LNX2, NPPC, ZEB2, CYLD, HIFIAN, ABCG5, BUB1B, MIR16-1, USF1, JMJD1C, SLC46A1, MTMR2, CHRM2, PRRC2A, NRIH4, RRM1, COX7A1, SALL4, PDE11A, CADM1, NOTCH4, ZCCHC8, CD5, ABCA4, OPRM1, MAT1A, CLYBL, CYP26B1, CALM3, APCDD1, RGS9, PCCA, CNKSR1, B4GALT7, AQP5, TSSK4, CEP41, SDHC, MPDZ, ATF6, GNB5, EYS, FMO2, IL16, ITIH1, MSTN, VCL, RNASEH2A, SERPINE1, CCDC88C, HFE, SMN1, BRWD3, VCAM1, EDARADD, PTAFR, PORCN, ROBO2, ALG3, CR2, GNA14, LPAR6, ZMYND11, MYH8, CDH23, PEX10, TTN, SLC26A5, RGS6, GALE, SRGAP3, NCALD, LMTK3, AKAP13, FLG, SLITRK1, ALS2CL, HHIP, RAD51, RAD51D, ADH1B, MEST, FERMT1, DYRK1A, HTRA2, SARDH, RANGRF, HBS1L, WISP3, CRK, KIAA0319, RNASET2, PKM, NEDD9, RPS7, JAK3, HSPA1B, FCER2, CXCL5, SPRR3, GSTO1, PITPNM3, ENPP1, SCNN1A, IDUA, KCNE2, P2RX4, CFB, UBIAD1, CDC42BPB, FCN3, SLC34A2, NPTN, NPY2R, SART3, CAPN10, FOXP1, CCR5, NDUFS5, SLC5A1, EPHB6, SYT14, PDLIM3, SCN8A, UNC93B1, RGMA, TJP2, KRT10, LILRA3, CCT5, RAB27A, CSF2RA, PEX19, PAX3, EYA4, RALGAP1, YAP1, EDN2, PSM7, ITPKC, ST8SIA2, RAD21, PPP1R3C, LMX1B, LRP4, SLC5A2, DLGAP2, MIF, PFN1, ORC6, IRF8, POMT2, CHRNA9, ITGB3, SPATA13, CEP152, CFHR3, BMPR1B, PDE8B, SF3B4, MAP7D3, LTC4S, SMARCA2, TUBGCP6, TLL1, ADAMTS10, COL10A1, BLOC1S6, NLGN4Y, BCL11A, F12, KDM5D, ACTA2, FANCE, FZD1, MKLI, ANKRD1, FREM1, SAA2, DDR1, PTGS1, ALDOB, LAMA4, DCC, FBXO18, TMEM185A, SYNPO, GUCA1A, UBESX, CLIC2, KLF10, ASCL1, LTA, TTPA, IFNA10, MIR499A, MGEA5, UBR7, PVT1, CYP46A1, BCAT1, PARD3B, SLC2A9, CHN1, RAG2, HGSNAT, PMM2, HMGAI, KRT85, SLC14A1, CA1, RANBP2, DCP1B, LIG3, SERPINI2, IRAK4, KCN115, EC11, DHODH, TULP3, ATP2A2, MLPH, ATIC, CXCL16, ATOH7, NRG1, NPL, CAST, TGFBI, SLC4A11, SNAI2, DGCR2, MARVELD2, CISH, JPH3, SPINT2, CYP2J2, CASK, HSPG2, APOL3, FLNA, BAX, CP, LRPAP1, SLC17A8, G6PC3, CYP24A1, SMARCAL1, ANO6, NPAS3, PRMT7, AHSPL, USH1C, DNAH5, CTC1, GHSR, ATP8B1, CLN3, TRPC3, SEMA4C, FMN2, TUBB1, SCN5A, FANCB, KAT6B, FBXO10, PDGFC, TP63, PEPD, ANXA11, VHL, CSF2RB, LMBR1, ZPBP, RPS6KA3, FGF23, NLRP14, PNPLA2, PCNT, BCKDHB, GALNTL5, PAH, PTCH2, ERCC2, VPS13B, DCLRE1C, HBG1, PSMB9, CYP2R1, HCLS1, CSF2, GHRH, IL11RA, RPS6KB1, GDF3, MMP9, POLR3B, RXFP2, HBA1, CTSC, AQP7, ADIPOQ, MPST, NDUFS6, ABCB7, SLC25A3, COX15, DGUOK, COQ9, NDUFS4, ELAC2, POLG, PDPI, COX10, L2HGDH, ACO2, OPA3, ETFA, ISCU, FOXRED1, NDUFAF1, UQCRCQ, TUFM, AUH, NDUFV2, FARS2, TK2, NDUFB3, FH, SLC25A4, D2HGDH, TMEM126A, TAZ, MTO1, PDHX, PANK2, NDUFA12, BOLA3, HARS2, MARS2, TSFM, NDUFS3, NFU1, TYMP, DARS2, AIFM1, POLG2, ACACA, YARS2, NDUFAF6, COA5, PUS1, OXCT1, EARS2, NDUFS2, UQCRCB, AMPD1, AGK, LRPPRC, COX4I2, SUCLG1, COQ6, MRPL3, COX6B1, NUBPL, HIBCH, SURF1, PC, SLC25A19, TIMM44, RRM2B, NDUFA9, SLC25A12, DLAT, LARS2, ATP5E, PDHB, ATPAF2, GARS, PNKD, NDUFB9, PDHA1, NDUFS1, NDUFAF4, ALDH3A2, PCK2, NDUFA2, FXN, OPA1, CI2orf65, PDSS1, RARS2, BCS1L, APTX, FASTKD2, SDHAF1, GCDH, TIMM8A, TTC19, LIAS, TMEM70, AASS, NDUFV1, XPNPEP3, NDUFA1, MRPS16, MPV17, GFER, SCO2, ETFB, COQ4, DNMI1, SCO1, NDUFAF3, COX14, SARS2, NDUFA11, AFG3L2, NDUFS8, IDH2, SLC19A3, MTFMT, CHAT, GFM1, MRPS22, GAMT, TWNK, MTPAP, AARS2, SACS, ACAD9, ETFDH, NDUFAF2, ETHE1, TPK1, ABAT, NDUFAF5, DLD, TACO1, PDSS2, NDUFA10, SUCLA2, NDUFS7, COQ2, TRMU, DNAJC19

DEMANDE D'IMPRIMATUR

Date de soutenance : 03/06/2022

<p align="center">DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE</p> <p>Présenté par : ALIX Tom</p> <p><u>Sujet</u> : Prédicteurs de l'utilité clinique du diagnostic moléculaire par séquençage haut débit de l'exome réalisé en première ligne chez des patients présentant un phénotype mendélien : Étude observationnelle sur 603 patients consécutifs</p> <p><u>Jury</u> :</p> <p>Président : Mme. Brigitte LEININGER-MULLER, PU Directeur : M. Abderrahim OUSSALAH, PU-PH Juges : M. Jean-Louis GUEANT, PU-PH M. Jean-Yves JOUZEAU, PU-PH Mme. Elise JEANNESSON, PH</p>	<p align="center">Vu,</p> <p align="center">Nancy, le 31/01/2022</p> <p>Le Président du Jury Directeur de Thèse</p> <p>Mme. Leininger M. Oussalah</p>  
<p align="center">Vu et approuvé,</p> <p align="center">Nancy, le 6.05.2022</p> <p align="center">Doyen de la Faculté de Pharmacie de l'Université de Lorraine,</p>  <p align="center">Pr. Raphaël DUVAL</p>	<p align="center">Vu,</p> <p align="center">Nancy, le 09.05.2022</p> <p align="center">Le Président de l'Université de Lorraine,</p>   <p align="center">Pierre MUTZENHARDT</p> <p align="center">N° d'enregistrement : 12473 C</p>