



HAL
open science

Systematisation d'interprétation des données de séquençage à haut débit : application au diagnostic des maladies neuromusculaires

Svetlana Gorokhova

► **To cite this version:**

Svetlana Gorokhova. Systematisation d'interprétation des données de séquençage à haut débit : application au diagnostic des maladies neuromusculaires. Sciences du Vivant [q-bio]. 2020. dumas-02950897

HAL Id: dumas-02950897

<https://dumas.ccsd.cnrs.fr/dumas-02950897v1>

Submitted on 28 Sep 2020

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.

**Systematisation d'interprétation des données de séquençage à haut débit :
application au diagnostic des maladies neuromusculaires**

T H E S E A R T I C L E

Présentée et publiquement soutenue devant

LA FACULTÉ DES SCIENCES MEDICALES ET PARAMEDICALES

DE MARSEILLE

Le lundi 6 avril 2020

Par Madame Svetlana GOROKHOVA épouse DEVRED

Née le 31 mai 1976 à KHARKOV (Ukraine)

Pour obtenir le grade de Docteur en Médecine

D.E.S. de GENETIQUE MEDICALE

Membres du Jury de la Thèse :

Monsieur le Professeur LEVY Nicolas

Président

Monsieur le Professeur KRAHN Martin

Directeur

Madame le Professeur PHILIP Nicole

Assesseur

Monsieur le Professeur ATTARIAN Shahram

Assesseur

**Systematisation d'interprétation des données de séquençage à haut débit :
application au diagnostic des maladies neuromusculaires**

T H E S E A R T I C L E

Présentée et publiquement soutenue devant

LA FACULTÉ DES SCIENCES MEDICALES ET PARAMEDICALES

DE MARSEILLE

Le lundi 6 avril 2020

Par Madame Svetlana GOROKHOVA épouse DEVRED

Née le 31 mai 1976 à KHARKOV (Ukraine)

Pour obtenir le grade de Docteur en Médecine

D.E.S. de GENETIQUE MEDICALE

Membres du Jury de la Thèse :

Monsieur le Professeur LEVY Nicolas

Président

Monsieur le Professeur KRAHN Martin

Directeur

Madame le Professeur PHILIP Nicole

Assesseur

Monsieur le Professeur ATTARIAN Shahram

Assesseur

AIX-MARSEILLE UNIVERSITE

Président : Yvon BERLAND

FACULTE DES SCIENCES MEDICALES ET PARAMEDICALES

Doyen Georges LEONETTI

Vice-doyen aux Affaires Générales et aux Sciences Médicales : Patrick DESSI
Vice-doyen aux Sciences Paramédicales : Philippe BERBIS

Direction d'école :

- **Ecole de Médecine** : Jean-Michel VITON
- **Ecoles de Maïeutique** : Carole ZAKARIAN
- **Ecoles des Sciences de la Réadaptation** : Philippe SAUVAGEON
- **Ecoles des Sciences Infirmières** : Sébastien COLSON

Assesseurs :

- aux Etudes : Kathia CHAUMOITRE
- à la Recherche : Jean-Louis MEGE
- aux Prospectives Hospitalo-Universitaires : Frédéric COLLART
- aux Enseignements Hospitaliers : Patrick VILLANI
- à l'Unité Mixte de Formation Continue en Santé : Fabrice BARLESI
- pour le Secteur Nord : Stéphane BERDAH
- aux centres hospitaliers non universitaires : Jean-Noël ARGENSON

Chargés de mission :

- 1^{er} cycle : Jean-Marc DURAND et Marc BARTHET
- 2^{ème} cycle : Marie-Aleth RICHARD
- 3^{ème} cycle DES/DESC : Pierre-Edouard FOURNIER
- Licences-Masters-Doctorat : Pascal ADALIAN
- DU-DIU : Véronique VITTON
- Stages Hospitaliers : Franck THUNY
- Sciences Humaines et Sociales : Pierre LE COZ
- Préparation à l'ECN : Aurélie DAUMAS
- Démographie Médicale et Filiation : Roland SAMBUC
- Relations Internationales : Philippe PAROLA
- Etudiants : Arthur ESQUER

Chef des services généraux : Déborah ROCCHICCIOLI

Chefs de service :

- Communication : Laetitia DELOUIS
- Examens : Caroline MOUTTET
- Intérieur : Joëlle FAVREGA
- Maintenance : Philippe KOCK
- Scolarité : Christine GAUTHIER

DOYENS HONORAIRES

M. Yvon BERLAND
M. André ALI CHERIF
M. Jean-François PELLISSIER

Mis à jour 01/01/2019

PROFESSEURS HONORAI RES

| | | | |
|----|------------------------|----|---------------------------|
| MM | AGOSTINI Serge | MM | FAVRE Roger |
| | ALDIGHERI René | | FIECHI Marius |
| | ALESSANDRINI Pierre | | FARNARIER Georges |
| | ALLIEZ Bernard | | FIGARELLA Jacques |
| | AQUARON Robert | | FONTES Michel |
| | ARGEME Maxime | | FRANCOIS Georges |
| | ASSADOURIAN Robert | | FUENTES Pierre |
| | AUFFRAY Jean-Pierre | | GABRIEL Bernard |
| | AUTILLO-TOUATI Amapola | | GALINIER Louis |
| | AZORIN Jean-Michel | | GALLAIS Hervé |
| | BAILLE Yves | | GAMERRE Marc |
| | BARDOT Jacques | | GARCIN Michel |
| | BARDOT André | | GARNIER Jean-Marc |
| | BERARD Pierre | | GAUTHIER André |
| | BERGOIN Maurice | | GERARD Raymond |
| | BERNARD Dominique | | GEROLAMI-SANTANDREA André |
| | BERNARD Jean-Louis | | GIUDICELLI Roger |
| | BERNARD Pierre-Marie | | GIUDICELLI Sébastien |
| | BERTRAND Edmond | | GOUDARD Alain |
| | BISSET Jean-Pierre | | GOUIN François |
| | BLANC Bernard | | GRILLO Jean-Marie |
| | BLANC Jean-Louis | | GRISOLI François |
| | BOLLINI Gérard | | GROULIER Pierre |
| | BONGRAND Pierre | | HADIDA/SAYAG Jacqueline |
| | BONNEAU Henri | | HASSOUN Jacques |
| | BONNOIT Jean | | HEIM Marc |
| | BORY Michel | | HOUEL Jean |
| | BOTTA Alain | | HUGUET Jean-François |
| | BOURGEADE Augustin | | JAQUET Philippe |
| | BOUVENOT Gilles | | JAMMES Yves |
| | BOUYALA Jean-Marie | | JOUVE Paulette |
| | BREMOND Georges | | JUHAN Claude |
| | BRICOT René | | JUIN Pierre |
| | BRUNET Christian | | KAPHAN Gérard |
| | BUREAU Henri | | KASBARIAN Michel |
| | CAMBOULIVES Jean | | KLEISBAUER Jean-Pierre |
| | CANNONI Maurice | | LACHARD Jean |
| | CARTOUZOU Guy | | LAFFARGUE Pierre |
| | CAU Pierre | | LAUGIER René |
| | CHABOT Jean-Michel | | LE TREUT Yves |
| | CHAMLIAN Albert | | LEVY Samuel |
| | CHARREL Michel | | LOUCHET Edmond |
| | CHAUVEL Patrick | | LOUIS René |
| | CHOUX Maurice | | LUCIANI Jean-Marie |
| | CIANFARANI François | | MAGALON Guy |
| | CLEMENT Robert | | MAGNAN Jacques |
| | COMBALBERT André | | MALLAN- MANCINI Josette |
| | CONTE-DEVOLX Bernard | | MALMEJAC Claude |
| | CORRIOL Jacques | | MARANINCHI Dominique |
| | COULANGE Christian | | MARTIN Claude |
| | DALMAS Henri | | MATTEI Jean François |
| | DE MICO Philippe | | MERCIER Claude |
| | DESSEIN Alain | | METGE Paul |
| | DELARQUE Alain | | MICHOTEY Georges |
| | DEVIN Robert | | MILLET Yves |
| | DEVRED Philippe | | MIRANDA François |
| | DJIANE Pierre | | MONFORT Gérard |
| | DONNET Vincent | | MONGES André |
| | DUCASSOU Jacques | | MONGIN Maurice |
| | DUFOUR Michel | | MONTIES Jean-Raoul |
| | DUMON Henri | | NAZARIAN Serge |
| | ENJALBERT Alain | | NICOLI René |

MM NOIRCLERC Michel
OLMER Michel
OREHEK Jean
PAPY Jean-Jacques
PAULIN Raymond
PELOUX Yves
PENAUD Antony
PENE Pierre
PIANA Lucien
PICAUD Robert
PIGNOL Fernand
POGGI Louis
POITOUT Dominique
PONCET Michel
POUGET Jean
PRIVAT Yvan
QUILICHINI Francis
RANQUE Jacques
RANQUE Philippe
RICHAUD Christian
RIDINGS Bernard
ROCHAT Hervé
ROHNER Jean-Jacques
ROUX Hubert
ROUX Michel
RUFO Marcel
SAHEL José
SALAMON Georges
SALDUCCI Jacques
SAN MARCO Jean-Louis
SANKALE Marc
SARACCO Jacques
SASTRE Bernard
SCHIANO Alain
SCOTTO Jean-Claude
SEBAHOUN Gérard
SERMENT Gérard
SERRATRICE Georges
SOULAYROL René
STAHL André
TAMALET Jacques
TARANGER-CHARPIN Colette
THOMASSIN Jean-Marc
UNAL Daniel
VAGUE Philippe
VAGUE/JUHAN Irène
VANUXEM Paul
VERVLOET Daniel
VIALETTES Bernard
WEILLER Pierre-Jean

PROFESSEURS HONORIS CAUSA

- 1967
 MM. les Professeurs DADI (Italie)
 CID DOS SANTOS (Portugal)
- 1974
 MM. les Professeurs MAC ILWAIN (Grande-Bretagne)
 T.A. LAMBO (Suisse)
- 1975
 MM. les Professeurs O. SWENSON (U.S.A.)
 Lord J.WALTON of DETCHANT (Grande-Bretagne)
- 1976
 MM. les Professeurs P. FRANCHIMONT (Belgique)
 Z.J. BOWERS (U.S.A.)
- 1977
 MM. les Professeurs C. GAJDUSEK-Prix Nobel (U.S.A.)
 C.GIBBS (U.S.A.)
 J. DACIE (Grande-Bretagne)
- 1978
 M. le Président F. HOUPHOUET-BOIGNY (Côte d'Ivoire)
- 1980
 MM. les Professeurs A. MARGULIS (U.S.A.)
 R.D. ADAMS (U.S.A.)
- 1981
 MM. les Professeurs H. RAPPAPORT (U.S.A.)
 M. SCHOU (Danemark)
 M. AMENT (U.S.A.)
 Sir A. HUXLEY (Grande-Bretagne)
 S. REFSUM (Norvège)
- 1982
 M. le Professeur W.H. HENDREN (U.S.A.)
- 1985
 MM. les Professeurs S. MASSRY (U.S.A.)
 KLINSMANN (R.D.A.)
- 1986
 MM. les Professeurs E. MIHICH (U.S.A.)
 T. MUNSAT (U.S.A.)
 LIANA BOLIS (Suisse)
 L.P. ROWLAND (U.S.A.)
- 1987
 M. le Professeur P.J. DYCK (U.S.A.)
- 1988
 MM. les Professeurs R. BERGUER (U.S.A.)
 W.K. ENGEL (U.S.A.)
 V. ASKANAS (U.S.A.)
 J. WEHSTER KIRKLIN (U.S.A.)
 A. DAVIGNON (Canada)
 A. BETTARELLO (Brésil)
- 1989
 M. le Professeur P. MUSTACCHI (U.S.A.)

1990
MM. les Professeurs J.G. MC LEOD (Australie)
J. PORTER (U.S.A.)

1991
MM. les Professeurs J. Edward MC DADE (U.S.A.)
W. BURGDORFER (U.S.A.)

1992
MM. les Professeurs H.G. SCHWARZACHER (Autriche)
D. CARSON (U.S.A.)
T. YAMAMURO (Japon)

1994
MM. les Professeurs G. KARPATI (Canada)
W.J. KOLFF (U.S.A.)

1995
MM. les Professeurs D. WALKER (U.S.A.)
M. MULLER (Suisse)
V. BONOMINI (Italie)

1997
MM. les Professeurs C. DINARELLO (U.S.A.)
D. STULBERG (U.S.A.)
A. MEIKLE DAVISON (Grande-Bretagne)
P.I. BRANEMARK (Suède)

1998
MM. les Professeurs O. JARDETSKY (U.S.A.)

1999
MM. les Professeurs J. BOTELLA LLUSIA (Espagne)
D. COLLEN (Belgique)
S. DIMAURO (U. S. A.)

2000
MM. les Professeurs D. SPIEGEL (U. S. A.)
C. R. CONTI (U.S.A.)

2001
MM. les Professeurs P-B. BENNET (U. S. A.)
G. HUGUES (Grande Bretagne)
J-J. O'CONNOR (Grande Bretagne)

2002
MM. les Professeurs M. ABEDI (Canada)
K. DAI (Chine)

2003
M. le Professeur T. MARRIE (Canada)
Sir G.K. RADDI (Grande Bretagne)

2004
M. le Professeur M. DAKE (U.S.A.)

2005
M. le Professeur L. CAVALLI-SFORZA (U.S.A.)

2006
M. le Professeur A. R. CASTANEDA (U.S.A.)

2007
M. le Professeur S. KAUFMANN (Allemagne)

PROFESSEURS EMERITE

| | | |
|-------------------|----------------------------|------------|
| 2008 | | |
| M. le Professeur | LEVY Samuel | 31/08/2011 |
| Mme le Professeur | JUHAN-VAGUE Irène | 31/08/2011 |
| M. le Professeur | PONCET Michel | 31/08/2011 |
| M. le Professeur | KASBARIAN Michel | 31/08/2011 |
| M. le Professeur | ROBERTOUX Pierre | 31/08/2011 |
| 2009 | | |
| M. le Professeur | DJIANE Pierre | 31/08/2011 |
| M. le Professeur | VERVLOET Daniel | 31/08/2012 |
| 2010 | | |
| M. le Professeur | MAGNAN Jacques | 31/12/2014 |
| 2011 | | |
| M. le Professeur | DI MARINO Vincent | 31/08/2015 |
| M. le Professeur | MARTIN Pierre | 31/08/2015 |
| M. le Professeur | METRAS Dominique | 31/08/2015 |
| 2012 | | |
| M. le Professeur | AUBANIAC Jean-Manuel | 31/08/2015 |
| M. le Professeur | BOUVENOT Gilles | 31/08/2015 |
| M. le Professeur | CAMBOULIVES Jean | 31/08/2015 |
| M. le Professeur | FAVRE Roger | 31/08/2015 |
| M. le Professeur | MATTEI Jean-François | 31/08/2015 |
| M. le Professeur | OLIVER Charles | 31/08/2015 |
| M. le Professeur | VERVLOET Daniel | 31/08/2015 |
| 2013 | | |
| M. le Professeur | BRANCHEREAU Alain | 31/08/2016 |
| M. le Professeur | CARAYON Pierre | 31/08/2016 |
| M. le Professeur | COZZONE Patrick | 31/08/2016 |
| M. le Professeur | DELMONT Jean | 31/08/2016 |
| M. le Professeur | HENRY Jean-François | 31/08/2016 |
| M. le Professeur | LE GUICHAOUA Marie-Roberte | 31/08/2016 |
| M. le Professeur | RUFO Marcel | 31/08/2016 |
| M. le Professeur | SEBAHOUN Gérard | 31/08/2016 |
| 2014 | | |
| M. le Professeur | FUENTES Pierre | 31/08/2017 |
| M. le Professeur | GAMERRE Marc | 31/08/2017 |
| M. le Professeur | MAGALON Guy | 31/08/2017 |
| M. le Professeur | PERAGUT Jean-Claude | 31/08/2017 |
| M. le Professeur | WEILLER Pierre-Jean | 31/08/2017 |
| 2015 | | |
| M. le Professeur | COULANGE Christian | 31/08/2018 |
| M. le Professeur | COURAND François | 31/08/2018 |
| M. le Professeur | FAVRE Roger | 31/08/2016 |
| M. le Professeur | MATTEI Jean-François | 31/08/2016 |
| M. le Professeur | OLIVER Charles | 31/08/2016 |
| M. le Professeur | VERVLOET Daniel | 31/08/2016 |

| | | |
|------------------|-------------------|------------|
| 2016 | | |
| M. le Professeur | BONGRAND Pierre | 31/08/2019 |
| M. le Professeur | BOUVENOT Gilles | 31/08/2017 |
| M. le Professeur | BRUNET Christian | 31/08/2019 |
| M. le Professeur | CAU Pierre | 31/08/2019 |
| M. le Professeur | COZZONE Patrick | 31/08/2017 |
| M. le Professeur | FAVRE Roger | 31/08/2017 |
| M. le Professeur | FONTES Michel | 31/08/2019 |
| M. le Professeur | JAMMES Yves | 31/08/2019 |
| M. le Professeur | NAZARIAN Serge | 31/08/2019 |
| M. le Professeur | OLIVER Charles | 31/08/2017 |
| M. le Professeur | POITOUT Dominique | 31/08/2019 |
| M. le Professeur | SEBAHOUN Gérard | 31/08/2017 |
| M. le Professeur | VIALETTES Bernard | 31/08/2019 |

| | | |
|------------------|---------------------|------------|
| 2017 | | |
| M. le Professeur | ALESSANDRINI Pierre | 31/08/2020 |
| M. le Professeur | BOUVENOT Gilles | 31/08/2018 |
| M. le Professeur | CHAUVEL Patrick | 31/08/2020 |
| M. le Professeur | COZZONE Pierre | 31/08/2018 |
| M. le Professeur | DELMONT Jean | 31/08/2018 |
| M. le Professeur | FAVRE Roger | 31/08/2018 |
| M. le Professeur | OLIVER Charles | 31/08/2018 |
| M. le Professeur | SEBBAHOUN Gérard | 31/08/2018 |

| | | |
|------------------|----------------------|------------|
| 2018 | | |
| M. le Professeur | MARANINCHI Dominique | 31/08/2021 |
| M. le Professeur | BOUVENOT Gilles | 31/08/2019 |
| M. le Professeur | COZZONE Pierre | 31/08/2019 |
| M. le Professeur | DELMONT Jean | 31/08/2019 |
| M. le Professeur | FAVRE Roger | 31/08/2019 |
| M. le Professeur | OLIVER Charles | 31/08/2019 |

PROFESSEURS DES UNIVERSITÉS-PRATICIENS HOSPITALIERS

| | | |
|--|---------------------------------------|---------------------------------------|
| AGOSTINI FERRANDES Aubert | CHINOT Olivier | GRIMAUD Jean-Charles |
| ALBANESE Jacques | CHOSSEGROS Cyrille | GROB Jean-Jacques |
| ALIMI Yves | <i>CLAVERIE Jean-Michel Surnombre</i> | GUEDJ Eric |
| AMABILE Philippe | COLLART Frédéric | GUIEU Régis |
| AMBROSI Pierre | COSTELLO Régis | GUIS Sandrine |
| ANDRE Nicolas | COURBIERE Blandine | GUYE Maxime |
| ARGENSON Jean-Noël | COWEN Didier | GUYOT Laurent |
| ASTOUL Philippe | CRAVELLO Ludovic | GUYS Jean-Michel |
| ATTARIAN Shahram | CUISSET Thomas | HABIB Gilbert |
| AUDOUIN Bertrand | CURVALE Georges | HARDWIGSEN Jean |
| AUQUIER Pascal | DA FONSECA David | HARLE Jean-Robert |
| AVIERINOS Jean-François | DAHAN-ALCARAZ Laetitia | <i>HOFFART Louis Disponibilité</i> |
| AZULAY Jean-Philippe | DANIEL Laurent | HOUVENAEGHEL Gilles |
| BAILLY Daniel | DARMON Patrice | JACQUIER Alexis |
| BARLESI Fabrice | D'ERCOLE Claude | JOURDE-CHICHE Noémie |
| BARLIER-SETTI Anne | D'JOURNO Xavier | JOUVE Jean-Luc |
| BARTHET Marc | DEHARO Jean-Claude | KAPLANSKI Gilles |
| BARTOLI Christophe | DELAPORTE Emmanuel | KARSENTY Gilles |
| BARTOLI Jean-Michel | DELPERO Jean-Robert | KERBAUL François |
| BARTOLI Michel | DENIS Danièle | KRAHN Martin |
| BARTOLOMEI Fabrice | DISDIER Patrick | LAFFORGUE Pierre |
| BASTIDE Cyrille | DODDOLI Christophe | LAGIER Jean-Christophe |
| BENSOUSSAN Laurent | DRANCOURT Michel | LAMBAUDIE Eric |
| BERBIS Philippe | DUBUS Jean-Christophe | LANCON Christophe |
| BERDAH Stéphane | DUFFAUD Florence | LA SCOLA Bernard |
| <i>BERLAND Yvon Surnombre</i> | DUFOUR Henry | LAUNAY Franck |
| BERNARD Jean-Paul | DURAND Jean-Marc | LAVIEILLE Jean-Pierre |
| BEROUD Christophe | DUSSOL Bertrand | LE CORROLLER Thomas |
| BERTUCCI François | EUSEBIO Alexandre | LECHEVALLIER Eric |
| BLAISE Didier | FAKHRY Nicolas | LEGRE Régis |
| BLIN Olivier | <i>FAUGERE Gérard Surnombre</i> | LEHUCHER-MICHEL Marie-Pascale |
| BLONDEL Benjamin | FELICIAN Olivier | LEONE Marc |
| BONIN/GUILLAUME Sylvie | FENOLLAR Florence | LEONETTI Georges |
| BONELLO Laurent | FIGARELLA/BRANGER Dominique | LEPIDI Hubert |
| BONNET Jean-Louis | FLECHER Xavier | LEVY Nicolas |
| <i>BOTTA/FRI DLUND Danielle Surnom</i> | FOURNIER Pierre-Edouard | MACE Loïc |
| BOUBLI Léon | <i>FRANCESCHI Yves Surnombre</i> | MAGNAN Pierre-Edouard |
| BOUFI Mourad | FRANCESCHI Frédéric | <i>MATONTI Frédéric Disponibilité</i> |
| BOYER Laurent | FUENTES Stéphane | MEGE Jean-Louis |
| BREGEON Fabienne | GABERT Jean | MERROT Thierry |
| BRETELLE Florence | GABORIT Bénédicte | METZLER/GUILLEMAIN Catherine |
| BROUQUI Philippe | GAINNIER Marc | MEYER/DUTOUR Anne |
| BRUDER Nicolas | GARCIA Stéphane | MICCALEF/ROLL Joëlle |
| BRUE Thierry | GARIBOLDI Vlad | MICHEL Fabrice |
| BRUNET Philippe | GAUDART Jean | MICHEL Gérard |
| BURTEY Stéphane | GAUDY-MARQUESTE Caroline | MICHEL Justin |
| CARCOPINO-TUSOLI Xavier | GENTILE Stéphanie | MICHELET Pierre |
| CASANOVA Dominique | GERBEAUX Patrick | MILH Mathieu |
| CASTINETTI Frédéric | GEROLAMI/SANTANDREA René | MOAL Valérie |
| CECCALDI Mathieu | GILBERT/ALESSI Marie-Christine | MONCLA Anne |
| CHAGNAUD Christophe | GIORGI Roch | MORANGE Pierre-Emmanuel |
| CHAMBOST Hervé | GIOVANNI Antoine | MOULIN Guy |
| CHAMPSAUR Pierre | GIRARD Nadine | MOUTARDIER Vincent |
| CHANEZ Pascal | GIRAUD/CHABROL Brigitte | <i>MUNDLER Olivier Surnombre</i> |
| CHARAFFE-JAUFFRET Emmanuelle | GONCALVES Anthony | NAUDIN Jean |
| CHARREL Rémi | GORINCOUR Guillaume | NICOLAS DE LAMBALLERIE Xavier |
| <i>CHARPIN Denis Surnombre</i> | GRANEL/REY Brigitte | NICOLLAS Richard |
| CHAUMOITRE Kathia | GRANVAL Philippe | OLIVE Daniel |
| CHIARONI Jacques | GREILLIER Laurent | OUAFIK L'Houcine |

| | | |
|---|--------------------------------|----------------------------|
| PAGANELLI Franck | ROCHE Pierre-Hugues | THOMAS Pascal |
| PANUEL Michel | ROCH Antoine | THUNY Franck |
| PAPAZIAN Laurent | ROCHWERGER Richard | TREBUCHON-DA FONSECA Agnès |
| PAROLA Philippe | ROLL Patrice | TRIGLIA Jean-Michel |
| <i>PARFATTE Sébastien Disponibilité</i> | ROSSI Dominique | TROPIANO Patrick |
| PELLISSIER-ALICOT Anne-Laure | ROSSI Pascal | TSIMARATOS Michel |
| PELLETIER Jean | ROUDIER Jean | TURRINI Olivier |
| PERRIN Jeanne | SALAS Sébastien | VALERO René |
| PETIT Philippe | <i>SAMBUC Roland Surnombre</i> | VAROQUAUX Arthur Damien |
| PHAM Thao | SARLES Jacques | VELLY Lionel |
| PIERCECCHI/MARTI Marie-Dominique | SARLES/PHILIP Nicole | VEY Norbert |
| PIQUET Philippe | SARLON-BARTOLI Gabrielle | VIDAL Vincent |
| PIRRO Nicolas | SCAVARDA Didier | VIENS Patrice |
| POINSO François | SCHLEINITZ Nicolas | VILLANI Patrick |
| RACCAH Denis | SEBAG Frédéric | VITON Jean-Michel |
| RANQUE Stéphane | SEITZ Jean-François | VITTON Véronique |
| RAOULT Didier | SIELEZNEFF Igor | VIEHWEGER Heide Elke |
| REGIS Jean | SIMON Nicolas | VIVIER Eric |
| REYNAUD/GAUBERT Martine | STEIN Andréas | XERRI Luc |
| REYNAUD Rachel | TAIEB David | |
| RICHARD/LALLEMAND Marie-Aleth | THIRION Xavier | |

PROFESSEUR DES UNIVERSITES

ADALIAN Pascal
 AGHABABIAN Valérie
 BELIN Pascal
 CHABANNON Christian
 CHABRIERE Eric
 FERON François
 LE COZ Pierre
 LEVASSEUR Anthony
 RANJEVA Jean-Philippe
 SOBOL Hagay

PROFESSEUR CERTIFIE

BRANDENBURGER Chantal

PRAG

TANTI-HARDOUIN Nicolas

PROFESSEUR ASSOCIE DE MEDECINE GENERALE A MI-TEMPS

ADNOT Sébastien
 FILIPPI Simon

MAI TRE DE CONFERENCES DES UNI VERSI TES - PRATI CI EN HOSPI TALI ER

| | | |
|---|------------------------------------|-------------------------|
| <i>ACHARD Vincent (disponibilité)</i> | EBBO Mikael | NGUYEN PHONG Karine |
| AHERFI Sarah | FABRE Alexandre | NINOVE Laetitia |
| <i>ANGELAKIS Emmanouil (dispo oct 2018)</i> | FAURE Alice | NOUGAIREDE Antoine |
| <i>ATLAN Catherine (disponibilité)</i> | FOLETTI Jean- Marc | OLLIVIER Matthieu |
| BARTHELEMY Pierre | FOUILLOUX Virginie | OVAERT Caroline |
| BEGE Thierry | FROMNOT Julien | PAULMYER/LACROIX Odile |
| BELIARD Sophie | GASTALDI Marguerite | PESENTI Sébastien |
| BERBIS Julie | GELSI/BOYER Véronique | RESSEGUIER Noémie |
| BERGE-LEFRANC Jean-Louis | GIUSIANO Bernard | REY Marc |
| BERTRAND Baptiste | GIUSIANO COURCAMBECK Sophie | ROBERT Philippe |
| BEYER-BERJOT Laura | GONZALEZ Jean-Michel | SABATIER Renaud |
| BIRNBAUM David | GOURIET Frédérique | SARI-MINODIER Irène |
| BONINI Francesca | GRAILLON Thomas | SAVEANU Alexandru |
| BOUCRAUT Joseph | GRISOLI Dominique | SECQ Véronique |
| BOULAMERY Audrey | GUERIN Carole | SUCHON Pierre |
| BOULLU/CIOCCA Sandrine | GUENOUN MEYSSIGNAC Daphné | TABOURET Emeline |
| BUFFAT Christophe | GUIDON Catherine | TOGA Caroline |
| CAMILLERI Serge | HAUTIER/KRAHN Aurélie | TOGA Isabelle |
| CARRON Romain | HRAIECH Sami | TOMASINI Pascale |
| CASSAGNE Carole | KASPI-PEZZOLI Elise | TOSELLO Barthélémy |
| CHAUDET Hervé | L'OLLIVIER Coralie | TROUSSE Delphine |
| CHRETIEN Anne-Sophie | LABIT-BOUVIER Corinne | TUCHTAN-TORRENTS Lucile |
| COZE Carole | LAFAGE/POCHITALOFF-HUVALE Marina | VELY Frédéric |
| CUNY Thomas | <i>LAGIER Aude (disponibilité)</i> | VION-DURY Jean |
| <i>DADOUN Frédéric (disponibilité)</i> | LAGOUANELLE/SIMEONI Marie-Claude | ZATTARA/CANNONI Hélène |
| DALES Jean-Philippe | LEVY/MOZZICONACCI Annie | |
| DAUMAS Aurélie | LOOSVELD Marie | |
| DEGEORGES/VITTE Joëlle | MANCINI Julien | |
| DELLIAUX Stéphane | MARY Charles | |
| DESPLAT/JEGO Sophie | MASCAUX Céline | |
| DEVILLIER Raynier | MAUES DE PAULA André | |
| DUBOURG Grégory | MILLION Matthieu | |
| DUFOUR Jean-Charles | MOTTOLA GHIGO Giovanna | |

**MAI TRES DE CONFERENCES DES UNI VERSI TES
(mono-appartenants)**

| | | |
|------------------------------|------------------------------|----------------|
| ABU ZAINEH Mohammad | DEGIOANNI/SALLE Anna | RUEL Jérôme |
| BARBACARU/PERLES T. A. | DESNUES Benoît | THOLLON Lionel |
| BERLAND/BENHAIM Caroline | MARANINCHI Marie | THIRION Sylvie |
| BOUCAULT/GARROUSTE Françoise | MERHEJ/CHAUVEAU Vicky | VERNA Emeline |
| BOYER Sylvie | MINVIELLE/DEVICTOR Bénédicte | |
| COLSON Sébastien | POGGI Marjorie | |

MAI TRE DE CONFERENCES DES UNI VERSI TES DE MEDECI NE GENERALE

CASANOVA Ludovic
GENTILE Gaëtan

MAI TRES DE CONFERENCES ASSOCI ES DE MEDECI NE GENERALE à MI - TEMPS

BARGIER Jacques
BONNET Pierre-André
CALVET-MONTREDON Céline
GUIDA Pierre
JANCZEWSKI Aurélie

MAI TRE DE CONFERENCES ASSOCI E à MI - TEMPS

MATHIEU Marion
REVIS Joana

PROFESSEURS DES UNIVERSITÉS et MAÎTRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS
 PROFESSEURS ASSOCIÉS, MAÎTRES DE CONFÉRENCES DES UNIVERSITÉS mono-appartenants

| ANATOMIE 4201 | ANTHROPOLOGIE 20 |
|---|--|
| <p>CHAMPSAUR Pierre (PU-PH) LE CORROLLER Thomas (PU-PH) PIRRO Nicolas (PU-PH)</p> <p>GUENOUN-MEYSSIGNAC Daphné (MCU-PH) <i>LAGIER Aude (MCU-PH) disponibilité</i></p> <p>THOLLON Lionel (MCF) (60ème section)</p> | <p>ADALIAN Pascal (PR)</p> <p>DEGIOANNI/SALLE Anna (MCF) VERNA Emeline (MCF)</p> |
| ANATOMIE ET CYTOLOGIE PATHOLOGIQUES 4203 | BACTÉRIOLOGIE-VIROLOGIE - HYGIÈNE HOSPITALIÈRE 4501 |
| <p>CHARAFE/JAUFFRET Emmanuelle (PU-PH) DANIEL Laurent (PU-PH) FIGARELLA/BRANGER Dominique (PU-PH) GARCIA Stéphane (PU-PH) XERRI Luc (PU-PH)</p> <p>DALES Jean-Philippe (MCU-PH) GIUSIANO COURCAMBECK Sophie (MCU-PH) LABIT/BOUVIER Corinne (MCU-PH) MAUES DE PAULA André (MCU-PH) SECQ Véronique (MCU-PH)</p> | <p>CHARREL Rémi (PU-PH) DRANCOURT Michel (PU-PH) FENOLLAR Florence (PU-PH) FOURNIER Pierre-Edouard (PU-PH) NICOLAS DE LAMBALLERIE Xavier (PU-PH) LA SCOLA Bernard (PU-PH) RAOULT Didier (PU-PH)</p> <p>AHERFI Sarah (MCU-PH) <i>ANGELAKIS Emmanouil (MCU-PH) disponibilité octobre 2018</i> DUBOURG Grégory (MCU-PH) GOURIET Frédérique (MCU-PH) NOUGAIREDE Antoine (MCU-PH) NINOVE Laetitia (MCU-PH)</p> <p>CHABRIERE Eric (PR) (64ème section) LEVASSEUR Anthony (PR) (64ème section) DESNUES Benoit (MCF) (65ème section) MERHEJ/CHAUVEAU Vicky (MCF) (87ème section)</p> |
| ANESTHÉSIOLOGIE ET RÉANIMATION CHIRURGICALE : MÉDECINE URGENCE 4801 | BIOPHYSIQUE ET BIOLOGIE MOLECULAIRE 4401 |
| <p>ALBANESE Jacques (PU-PH) BRUDER Nicolas (PU-PH) LEONE Marc (PU-PH) MICHEL Fabrice (PU-PH) VELLY Lionel (PU-PH)</p> <p>GUIDON Catherine (MCU-PH)</p> | <p>BARLIER/SETTI Anne (PU-PH) GABERT Jean (PU-PH) GUIEU Régis (PU-PH) OUAFIK L'Houcine (PU-PH)</p> <p>BUFFAT Christophe (MCU-PH) FROMONOT Julien (MCU-PH) MOTTOLA GHIGO Giovanna (MCU-PH) SAVEANU Alexandru (MCU-PH)</p> |
| ANGLAIS 11 | BIOLOGIE CELLULAIRE 4403 |
| <p>BRANDENBURGER Chantal (PRCE)</p> | <p>ROLL Patrice (PU-PH)</p> <p>GASTALDI Marguerite (MCU-PH) KASPI-PEZZOLI Elise (MCU-PH) LEVY-MOZZICONNACCI Annie (MCU-PH)</p> |
| BIOLOGIE ET MÉDECINE DU DÉVELOPPEMENT ET DE LA REPRODUCTION - GYNÉCOLOGIE MÉDICALE 5405 | |
| <p>METZLER/GUILLEMAIN Catherine (PU-PH) PERRIN Jeanne (PU-PH)</p> | |
| BIOPHYSIQUE ET MÉDECINE NUCLEAIRE 4301 | CARDIOLOGIE 5102 |
| <p>GUEDJ Eric (PU-PH) GUYE Maxime (PU-PH) <i>MUNDLER Olivier (PU-PH) Surnombre</i> TAIEB David (PU-PH)</p> <p>BELIN Pascal (PR) (69ème section) RANJEVA Jean-Philippe (PR) (69ème section)</p> <p>CAMILLERI Serge (MCU-PH) VION-DURY Jean (MCU-PH)</p> <p>BARBACARU/PERLES Téodora Adriana (MCF) (69ème section)</p> | <p>AVIERINOS Jean-François (PU-PH) BONELLO Laurent (PU-PH) BONNET Jean-Louis (PU-PH) CUISSET Thomas (PU-PH) DEHARO Jean-Claude (PU-PH) FRANCESCHI Frédéric (PU-PH) HABIB Gilbert (PU-PH) PAGANELLI Franck (PU-PH) THUNY Franck (PU-PH)</p> |
| BIOSTATISTIQUES, INFORMATIQUE MÉDICALE ET TECHNOLOGIES DE COMMUNICATION 4604 | CHIRURGIE DIGESTIVE 5202 |
| <p><i>CLAVERIE Jean-Michel (PU-PH) Surnombre</i> GAUDART Jean (PU-PH) GIORGI Roch (PU-PH)</p> <p>CHAUDET Hervé (MCU-PH) DUFOUR Jean-Charles (MCU-PH) GIUSIANO Bernard (MCU-PH) MANCINI Julien (MCU-PH)</p> <p>ABU ZAINEH Mohammad (MCF) (5ème section) BOYER Sylvie (MCF) (5ème section)</p> | <p>BERDAH Stéphane (PU-PH) HARDWIGSEN Jean (PU-PH) SIELEZNEFF Igor (PU-PH)</p> <p>BEYER-BERJOT Laura (MCU-PH)</p> |
| | CHIRURGIE GÉNÉRALE 5302 |
| | <p>DELPERO Jean-Robert (PU-PH) MOUTARDIER Vincent (PU-PH) SEBAG Frédéric (PU-PH) TURRINI Olivier (PU-PH)</p> <p>BEGE Thierry (MCU-PH) BIRNBAUM David (MCU-PH)</p> |

| | |
|--|---|
| CHI RURGIE ORTHOPEDIQUE ET TRAUMATOLOGIQUE 5002 | GUERIN Carole (MCU-PH) |
| ARGENSON Jean-Noël (PU-PH) BLONDEL Benjamin (PU-PH) CURVALE Georges (PU-PH) FLECHER Xavier (PU-PH) <i>PARFATTE Sébastien (PU-PH) Disponibilité</i> ROCHWERGER Richard (PU-PH) TROPIANO Patrick (PU-PH) | CHI RURGIE INFANTILE 5402 |
| OLLIVIER Matthieu (MCU-PH) | GUY Jean-Michel (PU-PH) JOUVE Jean-Luc (PU-PH) LAUNAY Franck (PU-PH) MERROT Thierry (PU-PH) VIEHWEGER Heide Elke (PU-PH) FAURE Alice (MCU-PH) PESENTI Sébastien (MCU-PH) |
| CANCEROLOGIE - RADIOTHERAPIE 4702 | CHI RURGIE MAXILLO-FACIALE ET STOMATOLOGIE 5503 |
| BERTUCCI François (PU-PH) CHINOT Olivier (PU-PH) COWEN Didier (PU-PH) DUFFAUD Florence (PU-PH) GONCALVES Anthony (PU-PH) HOUVENAEGHÉ Gilles (PU-PH) LAMBAUDIE Eric (PU-PH) SALAS Sébastien (PU-PH) VIENS Patrice (PU-PH) | CHOSSEGROS Cyrille (PU-PH) GUYOT Laurent (PU-PH) |
| SABATIER Renaud (MCU-PH) TABOURET Emeline (MCU-PH) | FOLETTI Jean-Marc (MCU-PH) |
| CHI RURGIE THORACIQUE ET CARDIOVASCULAIRE 5103 | CHI RURGIE PLASTIQUE, RECONSTRUCTRICE ET ESTHETIQUE : BRŪOLOGIE 5004 |
| COLLART Frédéric (PU-PH) D'JOURNO Xavier (PU-PH) DODDOLI Christophe (PU-PH) GARIBOLDI Vlad (PU-PH) MACE Loïc (PU-PH) THOMAS Pascal (PU-PH) | CASANOVA Dominique (PU-PH) LEGRE Régis (PU-PH) |
| FOUILLOUX Virginie (MCU-PH) GRISOLI Dominique (MCU-PH) TROUSSE Delphine (MCU-PH) | BERTRAND Baptiste (MCU-PH) HAUTIER/KRAHN Aurélie (MCU-PH) |
| CHI RURGIE VASCULAIRE - MEDECINE VASCULAIRE 5104 | GASTROENTEROLOGIE - HEPATOLOGIE - ADDICTOLOGIE 5201 |
| ALIMI Yves (PU-PH) AMABILE Philippe (PU-PH) BARTOLI Michel (PU-PH) BOUFI Mourad (PU-PH) MAGNAN Pierre-Edouard (PU-PH) PIQUET Philippe (PU-PH) SARLON-BARTOLI Gabrielle (PU-PH) | BARTHET Marc (PU-PH) BERNARD Jean-Paul (PU-PH) <i>BOTTA-FRIDLUND Danielle (PU-PH) Surnombre</i> DAHAN-ALCARAZ Laetitia (PU-PH) GEROLAMI-SANTANDREA René (PU-PH) GRANDVAL Philippe (PU-PH) GRIMAUD Jean-Charles (PU-PH) SEITZ Jean-François (PU-PH) VITTON Véronique (PU-PH) |
| HISTOLOGIE, EMBRYOLOGIE ET CYTOGENETIQUE 4202 | GONZALEZ Jean-Michel (MCU-PH) |
| LEPIDI Hubert (PU-PH) <i>ACHARD Vincent (MCU-PH) disponibilité</i> PAULMYER/LACROIX Odile (MCU-PH) | GENETIQUE 4704 |
| DERMATOLOGIE - VENEREOLOGIE 5003 | BEROUD Christophe (PU-PH) KRAHN Martin (PU-PH) LEVY Nicolas (PU-PH) MONCLA Anne (PU-PH) SARLES/PHILIP Nicole (PU-PH) |
| BERBIS Philippe (PU-PH) GAUDY/MARQUESTE Caroline (PU-PH) GROB Jean-Jacques (PU-PH) RICHARD/LALLEMAND Marie-Aleth (PU-PH) | NGYUEN Karine (MCU-PH) TOGA Caroline (MCU-PH) ZATTARA/CANNONI Hélène (MCU-PH) |
| DUSI | |
| COLSON Sébastien (MCF) | |
| ENDOCRINOLOGIE, DIABETE ET MALADIES METABOLIQUES : GYNECOLOGIE MEDICALE 5404 | |
| BRUE Thierry (PU-PH) CASTINETTI Frédéric (PU-PH) CUNY Thomas (MCU-PH) | GYNECOLOGIE-OBSTETRIQUE : GYNECOLOGIE MEDICALE 5403 |
| EPIDEMIOLOGIE, ECONOMIE DE LA SANTE ET PREVENTION 4601 | AGOSTINI Aubert (PU-PH) BOUBLI Léon (PU-PH) BRETTELLE Florence (PU-PH) CARCOPINO-TUSOLI Xavier (PU-PH) COURBIERE Blandine (PU-PH) CRAVELLO Ludovic (PU-PH) D'ERCOLE Claude (PU-PH) |
| AUQUIER Pascal (PU-PH) BOYER Laurent (PU-PH) GENTILE Stéphanie (PU-PH) <i>SAMBUC Roland (PU-PH) Surnombre</i> THIRION Xavier (PU-PH) | |
| BERBIS Julie (MCU-PH) LAGOUANELLE/SIMEONI Marie-Claude (MCU-PH) RESSEGUIER Noémie (MCU-PH) | |
| MINVIELLE/DEVICTOR Bénédicte (MCF)(06ème section) TANTI-HARDOUIN Nicolas (PRAG) | |

| | |
|---|---|
| <p style="text-align: center;">IMMUNOLOGIE 4703</p> <p>KAPLANSKI Gilles (PU-PH) MEGE Jean-Louis (PU-PH) OLIVE Daniel (PU-PH) VIVIER Eric (PU-PH)</p> <p>FERON François (PR) (69ème section)</p> <p>BOUCRAUT Joseph (MCU-PH) CHRETIEN Anne-Sophie (MCU PH) DEGEORGES/VITTE Joëlle (MCU-PH) DESPLAT/JEGO Sophie (MCU-PH) ROBERT Philippe (MCU-PH) VELY Frédéric (MCU-PH)</p> <p>BOUCAULT/GARROUSTE Françoise (MCF) 65ème section)</p> | <p style="text-align: center;">HEMATOLOGIE : TRANSFUSION 4701</p> <p>BLAISE Didier (PU-PH) COSTELLO Régis (PU-PH) CHIARONI Jacques (PU-PH) GILBERT/ALESSI Marie-Christine (PU-PH) MORANGE Pierre-Emmanuel (PU-PH) VEY Norbert (PU-PH)</p> <p>DEVILLIER Raynier (MCU PH) GELSI/BOYER Véronique (MCU-PH) LAFAGE/POCHITALOFF-HUVALE Marina (MCU-PH) LOOSVELD Marie (MCU-PH) SUCHON Pierre (MCU-PH)</p> <p>POGGI Marjorie (MCF) (64ème section)</p> |
| <p style="text-align: center;">MALADIES INFECTIEUSES : MALADIES TROPICALES 4503</p> <p>BROUQUI Philippe (PU-PH) LAGIER Jean-Christophe (PU-PH) PAROLA Philippe (PU-PH) STEIN Andréas (PU-PH)</p> <p>MILLION Matthieu (MCU-PH)</p> | <p style="text-align: center;">MEDECINE LEGALE ET DROIT DE LA SANTE 4603</p> <p>BARTOLI Christophe (PU-PH) LEONETTI Georges (PU-PH) PELLISSIER-ALICOT Anne-Laure (PU-PH) PIERCECCHI-MARTI Marie-Dominique (PU-PH)</p> <p>TUCHTAN-TORRENTS Lucile (MCU-PH)</p> <p>BERLAND/BENHAIM Caroline (MCF) (1ère section)</p> |
| <p style="text-align: center;">MEDECINE D'URGENCE 4805</p> <p>KERBAUL François (PU-PH) MICHELET Pierre (PU-PH)</p> | <p style="text-align: center;">MEDECINE PHYSIQUE ET DE READAPTATION 4905</p> <p>BENSOUSSAN Laurent (PU-PH) VITON Jean-Michel (PU-PH)</p> |
| <p style="text-align: center;">MEDECINE INTERNE : GERIATRIE ET BIOLOGIE DU VIELLISSEMENT : MEDECINE GENERALE : ADDICTOLOGIE 5301</p> <p>BONIN/GUILLAUME Sylvie (PU-PH) DISDIER Patrick (PU-PH) DURAND Jean-Marc (PU-PH) <i>FRANCES Yves (PU-PH) Surnombre</i> GRANEL/REY Brigitte (PU-PH) HARLE Jean-Robert (PU-PH) ROSSI Pascal (PU-PH) SCHLEINITZ Nicolas (PU-PH)</p> <p>EBBO Mikael (MCU-PH)</p> <p>GENTILE Gaëtan (MCF Méd. Gén. Temps plein)</p> <p>ADNOT Sébastien (PR associé Méd. Gén. à mi-temps) FILIPPI Simon (PR associé Méd. Gén. à mi-temps)</p> <p>BARGIER Jacques (MCF associé Méd. Gén. à mi-temps) BONNET Pierre-André (MCF associé Méd. Gén. à mi-temps) CALVET-MONTREDON Céline (MCF associé Méd. Gén. à temps plein) GUIDA Pierre (MCF associé Méd. Gén. à mi-temps) JANCZEWSKI Aurélie (MCF associé Méd. Gén. à mi-temps)</p> | <p style="text-align: center;">MEDECINE ET SANTE AU TRAVAIL 4602</p> <p>LEHUCHER/MICHEL Marie-Pascale (PU-PH)</p> <p>BERGE-LEFRANC Jean-Louis (MCU-PH) SARI/MINODIER Irène (MCU-PH)</p> <p style="text-align: center;">NEPHROLOGIE 5203</p> <p><i>BERLAND Yvon (PU-PH) Surnombre</i> BRUNET Philippe (PU-PH) BURTEY Stéphanne (PU-PH) DUSSOL Bertrand (PU-PH) JOURDE CHICHE Noémie (PU PH) MOAL Valérie (PU-PH)</p> |
| <p style="text-align: center;">NUTRITION 4404</p> <p>DARMON Patrice (PU-PH) RACCAH Denis (PU-PH) VALERO René (PU-PH)</p> <p><i>ATLAN Catherine (MCU-PH) disponibilité</i> BELIARD Sophie (MCU-PH)</p> <p>MARANINCHI Marie (MCF) (66ème section)</p> | <p style="text-align: center;">NEUROCHIRURGIE 4902</p> <p>DUFOUR Henry (PU-PH) FUENTES Stéphane (PU-PH) REGIS Jean (PU-PH) ROCHE Pierre-Hugues (PU-PH) SCAVARDA Didier (PU-PH)</p> <p>CARRON Romain (MCU PH) GRAILLON Thomas (MCU PH)</p> |
| <p style="text-align: center;">ONCOLOGIE 65 (BIOLOGIE CELLULAIRE)</p> <p>CHABANNON Christian (PR) (66ème section) SOBOL Hagay (PR) (65ème section)</p> | <p style="text-align: center;">NEUROLOGIE 4901</p> <p>ATTARIAN Sharham (PU PH) AUDOIN Bertrand (PU-PH) AZULAY Jean-Philippe (PU-PH) CECCALDI Mathieu (PU-PH) EUSEBIO Alexandre (PU-PH) FELICIAN Olivier (PU-PH) PELLETIER Jean (PU-PH)</p> |

| | |
|--|--|
| OPHTALMOLOGIE 5502 | PEDOPSYCHIATRIE - ADDICTOLOGIE 4904 |
| DENIS Danièle (PU-PH) HOFFART Louis (PU-PH) <i>Disponibilité</i> MATONTI Frédéric (PU-PH) <i>Disponibilité</i> | DA FONSECA David (PU-PH) POINSO François (PU-PH) |
| OTO-RHINO-LARYNGOLOGIE 5501 | PHARMACOLOGIE FONDAMENTALE - PHARMACOLOGIE CLINIQUE - ADDICTOLOGIE 4803 |
| DESSI Patrick (PU-PH) FAKHRY Nicolas (PU-PH) GIOVANNI Antoine (PU-PH) LAVIEILLE Jean-Pierre (PU-PH) MICHEL Justin (PU-PH) NICOLLAS Richard (PU-PH) TRIGLIA Jean-Michel (PU-PH) DEVEZE Arnaud (MCU-PH) <i>Disponibilité</i> REVIS Joana (MAST) (Orthophonie) (7ème Section) | BLIN Olivier (PU-PH) FAUGERE Gérard (PU-PH) <i>Surnombre</i> MICALLEF/ROLL Joëlle (PU-PH) SIMON Nicolas (PU-PH) BOULAMERY Audrey (MCU-PH) |
| PARASITOLOGIE ET MYCOLOGIE 4502 | PHILOSOPHIE 17 |
| RANQUE Stéphane (PU-PH) CASSAGNE Carole (MCU-PH) L'OLLIVIER Coralie (MCU-PH) MARY Charles (MCU-PH) TOGA Isabelle (MCU-PH) | LE COZ Pierre (PR) (17ème section) MATHIEU Marion (MAST) |
| PEDIATRIE 5401 | PHYSIOLOGIE 4402 |
| ANDRE Nicolas (PU-PH) CHAMBOST Hervé (PU-PH) DUBUS Jean-Christophe (PU-PH) GIRAUD/CHABROL Brigitte (PU-PH) MICHEL Gérard (PU-PH) MILH Mathieu (PU-PH) REYNAUD Rachel (PU-PH) SARLES Jacques (PU-PH) TSIMARATOS Michel (PU-PH) COZE Carole (MCU-PH) FABRE Alexandre (MCU-PH) OVAERT Caroline (MCU-PH) TOSELLO Barthélémy (MCU-PH) | BARTOLOMEI Fabrice (PU-PH) BREGON Fabienne (PU-PH) GABORIT Bénédicte (PU-PH) MEYER/DUTOUR Anne (PU-PH) TREBUCHON/DA FONSECA Agnès (PU-PH) BARTHELEMY Pierre (MCU-PH) BONINI Francesca (MCU-PH) BOULLU/CIOCCA Sandrine (MCU-PH) DADOUN Frédéric (MCU-PH) <i>(disponibilité)</i> DELLIAUX Stéphane (MCU-PH) REY Marc (MCU-PH) RUEL Jérôme (MCF) (69ème section) THIRION Sylvie (MCF) (66ème section) |
| PSYCHIATRIE D'ADULTES : ADDICTOLOGIE 4903 | |
| BAILLY Daniel (PU-PH) LANCON Christophe (PU-PH) NAUDIN Jean (PU-PH) | |
| PSYCHOLOGIE - PSYCHOLOGIE CLINIQUE - PSYCHOLOGIE SOCIALE 16 | |
| AGHABABIAN Valérie (PR) | |
| RADIOLOGIE ET IMAGERIE MEDICALE 4302 | PNEUMOLOGIE - ADDICTOLOGIE 5101 |
| BARTOLI Jean-Michel (PU-PH) CHAGNAUD Christophe (PU-PH) CHAUMOTRE Kathia (PU-PH) GIRARD Nadine (PU-PH) GORINCOUR Guillaume (PU-PH) JACQUIER Alexis (PU-PH) MOULIN Guy (PU-PH) PANUEL Michel (PU-PH) PETIT Philippe (PU-PH) VAROQUAUX Arthur Damien (PU-PH) VIDAL Vincent (PU-PH) | ASTOUL Philippe (PU-PH) BARLESI Fabrice (PU-PH) CHANEZ Pascal (PU-PH) CHARPIN Denis (PU-PH) <i>Surnombre</i> GREILLIER Laurent (PU-PH) REYNAUD/GAUBERT Martine (PU-PH) MASCAUX Céline (MCU-PH) TOMASINI Pascale (MCU-PH) |
| REANIMATION MEDICALE - MEDECINE D'URGENCE 4802 | THERAPEUTIQUE - MEDECINE D'URGENCE - ADDICTOLOGIE 4804 |
| GAINNIER Marc (PU-PH) GERBEAUX Patrick (PU-PH) PAPAZIAN Laurent (PU-PH) ROCH Antoine (PU-PH) HRAIECH Sami (MCU-PH) | AMBROSI Pierre (PU-PH) VILLANI Patrick (PU-PH) DAUMAS Aurélie (MCU-PH) |
| RHUMATOLOGIE 5001 | UROLOGIE 5204 |
| GUIS Sandrine (PU-PH) LAFFORGUE Pierre (PU-PH) PHAM Thao (PU-PH) ROUDIER Jean (PU-PH) | BASTIDE Cyrille (PU-PH) KARSENTY Gilles (PU-PH) LECHEVALLIER Eric (PU-PH) ROSSI Dominique (PU-PH) |

Remerciements

A Monsieur le Professeur Nicolas LEVY, Président de jury :

Merci pour votre bienveillance et votre accueil dans le service de Génétique Médicale. Le soutien que vous m'avez apporté pendant toutes ces années m'a été d'une grande aide.

A Monsieur le Professeur Martin KRAHN, Directeur de thèse :

Merci pour ton soutien et ta disponibilité. Bientôt sept ans que nous travaillons ensemble et que j'ai plaisir à partager ton optimisme et ton enthousiasme.

A Madame le Professeur Nicole PHILIP, membre du jury :

Je suis très honorée que vous ayez accepté de siéger dans mon jury. Votre passion pour la génétique et la rigueur qui l'accompagne sont des inspirations pour moi.

A Monsieur le Professeur Shahram ATTARIAN, membre du jury :

Je suis très honorée que vous ayez accepté de siéger dans mon jury. Nos échanges pendant les RCP m'ont beaucoup appris sur les maladies neuromusculaires et ont été très enrichissants.

Je remercie toutes les équipes du Service de Génétique à l'Hôpital de la Timone et de Biologie Moléculaire à l'Hôpital de la Conception dans lesquels j'ai effectué mes stages. Merci aux Docteurs et Professeurs de m'avoir transmis leur savoir. Je remercie aussi les techniciens, ingénieurs, conseillères en génétique, secrétaires, infirmiers, cadres de santé ainsi que mes chers co-internes et assistants pour leur soutien et pour d'avoir rendu mes stages aussi agréables.

Je remercie Dr. Marc Bartoli et l'équipe Myologie Translationnelle pour les discussions scientifiques toujours très animées et stimulantes. Une partie des travaux présentés ici a été effectuée par les membres de l'équipe.

Enfin, je remercie ma famille et mes amis pour leur soutien absolu sans lequel ce travail n'aurait pas été possible

Table des matières

| | |
|---|--------|
| Articles | - 20 - |
| Article 1 : A new tool CovReport generates easy-to-understand sequencing coverage summary for diagnostic reports. | - 20 - |
| Article 2 : Recommandations françaises pour l'interprétation de variations nucléotidiques générées par les analyses en séquençage haut-débit | - 27 - |
| Article 3 : Systematic functional assessment of impact on splicing of deep exonic missense variants in CAPN3..... | - 28 - |
| Introduction | - 43 - |
| Séquençage à haut débit et son implication dans le diagnostic des maladies neuromusculaires rares | - 43 - |
| Interprétation des données issues d'un test diagnostique par Séquençage à Haut Débit..... | - 46 - |
| Critères de qualité indispensables pour l'interprétation des résultats d'un séquençage : couverture des séquences cibles et profondeur de lecture du séquençage..... | - 46 - |
| Recommandations pour l'interprétation des variants identifiés par un test diagnostique par séquençage à haut débit | - 50 - |
| Importance de tests fonctionnels pour l'interprétation des variants | - 52 - |
| Structuration du diagnostic génétique des maladies neuromusculaires par Séquençage à Haut Débit en France | - 53 - |
| Matériel et méthode | - 55 - |
| Mise en place de CovReport – l'outil de visualisation de la couverture..... | - 55 - |
| Mise à jour des recommandations françaises pour l'interprétation des variants..... | - 56 - |
| Sélection et analyse des variants faux-sens dans le gène <i>CAPN3</i> avec l'impact potentiel sur l'épissage..... | - 57 - |
| Résultats | - 58 - |
| Utilisation de CovReport pour visualiser la couverture d'un test diagnostique par séquençage à haut débit..... | - 58 - |
| Mise à jour des recommandations françaises pour l'interprétation des variants..... | - 60 - |
| Etudes fonctionnelles des variants faux-sens dans le gène <i>CAPN3</i> | - 63 - |
| Discussion | - 66 - |
| Améliorations apportées par utilisation de CovReport dans le cadre du diagnostic par séquençage à haut débit | - 67 - |
| Les défis de l'interprétation de variants identifiés par séquençage à haut débit et l'importance des RCP (Réunions de Concertation Pluridisciplinaire) | - 68 - |
| La double face des variants faux-sens exoniques profonds | - 68 - |
| Les défis de la prédiction bio-informatique de l'effet d'un variant sur l'épissage | - 69 - |
| Les tests fonctionnels de l'effet sur épissage dans le cadre du diagnostic des maladies neuromusculaires..... | - 70 - |
| Nécessité d'un diagnostic rapide pour bénéficier des traitements potentiels..... | - 70 - |
| Conclusion | - 72 - |
| Bibliographie | - 74 - |
| Annexes | - 79 - |

Articles



Article 1 : A new tool CovReport generates easy-to-understand sequencing coverage summary for diagnostic reports.

Publication de pré-print sur bioRxiv, Juin 14, 2019. doi:10.1101/671511

Article usage: June 2019 to March 2020

| Show by month | Abstract | Full-text HTML | PDF |
|---------------|----------|----------------|-----|
| Total | 1,710 | 99 | 412 |



 Tweeted by 42
 9 readers on Mendeley

Une version révisée de l'article a été resoumise dans **Scientific Reports** suite une revue favorable.

| | |
|-----------------------------------|--|
| Manuscript # | SREP-19-42001A |
| Current Revision # | 1 |
| Other Version | SREP-19-42001-T |
| Submission Date | 23rd February 20 |
| Current Stage | Quality Check Started |
| Title | A new tool CovReport generates easy-to-understand sequencing coverage summary for diagnostic reports. |
| Manuscript Type | Original Research |
| Collection | N/A |
| Corresponding Author | Dr. Svetlana Gorokhova (svetlana.gorokhova@univ-amu.fr) (Aix Marseille University) |
| Contributing Authors | Mr. Mark Gorokhov , Dr. Mathieu Cerino , Dr. Jérémie Mortreux , Dr. Florence Riccardi , Prof. Nicolas Lévy , Marc Bartoli , Dr. Martin Krahn |
| Authorship | Yes |
| Abstract | In order to properly interpret the results of a diagnostic gene panel sequencing test, gene coverage needs to be taken into consideration. If coverage is too low, an additional re-sequencing test is needed to make sure that a pathogenic variant is not missed. To facilitate the interpretation of coverage data, we designed CovReport, a novel easy-to-use visualization tool. CovReport generates a concise coverage summary that allows one-glance assessment of the sequencing test performance. Both gene-level and exon-level coverage can be immediately appreciated and taken into consideration for further medical decisions. CovReport does not require complex installation and can thus be easily implemented in any diagnostic laboratory setting. A user-friendly interface generates a graphic summary of coverage that can be directly included in the diagnostic report. In addition to a stand-alone version, we also provide a command line version of CovReport that can be integrated into any bioinformatics pipeline. This flexible tool is now part of routine sequencing analysis at the Department of Medical Genetics at La Timone Hospital (Marseille, France). CovReport is available at http://jdotsoft.com/CovReport.php . It is implemented in Java and supported on Windows, Mac OS X and Linux. |
| Techniques | Life sciences techniques [Genomic analysis]; Life sciences techniques, Genomic analysis [High-throughput screening assays]; Life sciences techniques [High throughput screening]; Life sciences techniques, High throughput screening [High-throughput screening assays]; |
| Subject Terms | Biological sciences/Computational biology and bioinformatics/Software Biological sciences/Genetics/Clinical genetics Biological sciences/Computational biology and bioinformatics/Genome informatics |
| Competing Interests Policy | There is NO Competing Interest. |
| Applicable Funding Source | No Applicable Funding |

De : scientificreports@nature.com <scientificreports@nature.com>

Envoyé : vendredi 13 mars 2020 11:54

À : GOROKHOVA Svetlana

Objet : Receipt of revised manuscript SREP-19-42001B

Dear Dr Gorokhova,

Thank you for submitting your revised manuscript entitled "A new tool CovReport generates easy-to-

understand sequencing coverage summary for diagnostic reports."

Your revision has been assigned a tracking number of SREP-19-42001B. Please quote this tracking number in the subject line of any correspondence.

You may check the status of your manuscript by selecting the "Check manuscript status" link at the following URL:

<https://mts-srep.nature.com/cgi-bin/main.plex?el=A5CG6CKYK7C7KGoe6F2A9ftd7Fiv2bySiFbWDQDb2csZgZ>

You will be contacted shortly with a decision.

Best regards,

Manuscript Administration

Scientific Reports

4 Crinan Street

London N1 9XW

E-mail: scientificreports@nature.com

A new tool CovReport generates easy-to-understand sequencing coverage summary for diagnostic reports.

Mark Gorokhov¹, Mathieu Cerino^{2,3}, Jérémie Mortreux^{2,3,4}, Florence Riccardi^{2,3,4}, Nicolas Lévy^{2,3}, Marc Bartoli², Martin Krahn^{2,3}, Svetlana Gorokhova^{2,3*}

1. JDotSoft, Germantown, MD, USA

2. Aix Marseille University, INSERM, MMG, U1251, Marseille, France

3. Department of Medical Genetics, La Timone Hospital, AP-HM, Marseille, France

4. These authors contributed equally

* corresponding author

Abstract

In order to properly interpret the results of a diagnostic gene panel sequencing test, gene coverage needs to be taken into consideration. If coverage is too low, an additional re-sequencing test is needed to make sure that a pathogenic variant is not missed. To facilitate the interpretation of coverage data, we designed CovReport, a novel easy-to-use visualization tool. CovReport generates a concise coverage summary that allows one-glance assessment of the sequencing test performance. Both gene-level and exon-level coverage can be immediately appreciated and taken into consideration for further medical decisions. CovReport does not require complex installation and can thus be easily implemented in any diagnostic laboratory setting. A user-friendly interface generates a graphic summary of coverage that can be directly included in the diagnostic report. In addition to a stand-alone version, we also provide a command line version of CovReport that can be integrated into any bioinformatics pipeline. This flexible tool is now part of routine sequencing analysis at the Department of Medical Genetics at La Timone Hospital (Marseille, France).

CovReport is available at <http://jdotsoft.com/CovReport.php>. It is implemented in Java and supported on Windows, Mac OS X and Linux.

Background

Since the arrival of short-read sequencing technologies, significant efforts have been spent to improve the quality of the obtained data in order to satisfy the requirements of molecular diagnostics. However, many regions of the human genome remain difficult to analyze using standard short-read sequencing approaches. These "dark" genome regions contain a number of genes responsible for human diseases¹. For example, several neuromuscular disease-causing genes, such as *NEBULIN* (*NEB*) and *SELENON* (*SEPN1*), overlap these difficult to sequence regions. Disease-causing variants in gene regions with suboptimal sequence coverage can be overlooked. Thus, when short-read sequencing is

performed in a diagnostic setting, identifying poorly-covered regions is critical for the interpretation of a diagnostic sequencing result. If coverage is not sufficient even for a small region of a highly suspected candidate gene, an additional re-sequencing test is needed to make sure that the pathogenic variant is not missed. To facilitate the coverage data interpretation by test prescribers, we designed a novel easy-to-use visualization tool, CovReport. The concise coverage data summary generated by CovReport allows one-glance assessment of the sequencing test performance.

Methods

Implementation

CovReport is implemented as a standalone Java application. It can run on any platform where Java Runtime (JRE) is installed (Windows, Mac OS and Linux). The supported Java version is 8. The application is using open source external libraries (JARs) which are embedded into the main executable JAR. The following external dependencies are used:

- Apache PdfBox (<https://pdfbox.apache.org>),
- Apache CLI (<http://commons.apache.org/proper/commons-cli>),
- JarClassLoader (<http://www.jdotsoft.com/JarClassLoader.php>).

The application could be downloaded from <http://www.jdotsoft.com/CovReport.php> as a compressed CovReport.zip file. After extraction into the local drive, the folder contains the following:

- `msg` – folder with internalization message files used for the generated PDF file
- `CovReport.jar` – Java executable archive
- `run.cmd` – Windows command to start the application
- `runFromCommandLine.cmd` – Windows command line helper

After CovReport first execution the following items are created:

- `pdf-results` – folder with PDF files generated by the application
- `CovReport.config` – file with configuration persistence data; this file could be manually updated and reused to replace default for command line execution

The application can be started with user interface (UI) or from command line. The command line options are:

- i, --input <arg> input file
- n, --name <arg> patient name
- c, --config <arg> config file (optional)

The input file containing per-exon coverage information is a tab delimited CSV file. All entries in the file can be quoted per CSV file standard or can be without quotes (there should be no special characters or in the input file). The following columns are expected: RefSeqName, GeneSymbol, Exon, Size, Mean Depth, SD Depth, Coverage 1x, Coverage 5x, Coverage 10x, Coverage 20x, Coverage 30x. An example of an input file is provided in the Supplementary data (also available at http://www.jdotsoft.com/CovReport/44genes-stat_coverage_exon.txt). The input file can be easily generated by the Coverage Module of VarAFT tool² (<https://varaft.eu/>) that uses BEDTools^{3,4} to calculate exon-level coverage. However, results of any other coverage calculating pipeline can be converted into the appropriate format and used as CovReport input. Moreover, the command line version of CovReport allows integrating this visualization tool as a step in an automated diagnostic sequence analysis pipeline.

The application user interface is in English, but the generated PDF file is in the current locale language. English and French are supported in the current distribution, with the possibility to add other languages by the user. The text included in the output pdf report can also be customized, making it possible to adopt CovReport in any diagnostic setting.

Detailed instructions for downloading and running CovReport are described at <http://www.jdotsoft.com/CovReport.php>

Exome sequencing of NA12878

DNA for reference sample NA12878 was obtained from the Coriell Institute for Medical Research Repository (Coriell Institute, Camden, NJ, USA). Whole Exome Sequencing (WES) was performed by the Genomics and Bioinformatics Platform (GBiM) from the U 1251/Marseille Medical Genetics facility, using the NimbleGen SeqCap EZ MedExome kit (total design size 47 Mb) according to the manufacturer's protocol (Roche Sequencing Solutions, Madison, USA). The SeqCap EZ MedExome kit targets the entire human exome with enhanced coverage of exons from medically relevant genes in Mendelian diseases. Enriched fragment libraries were sequenced on the Illumina NextSeq 500 platform (Illumina, San Diego, CA, USA) using a 150 bp paired-end sequencing protocol. Raw data were mapped to the built of the human genome (hg19) by using BWA 0.7.5.

Results

CovReport allows one-glance overview of exon coverage for a diagnostic sequencing test by generating a concise easy-to-understand report in the PDF format. Intuitive interface and absence of complex installation steps makes CovReport easy to apply by users without special bioinformatics training. CovReport can also be launched by command line allowing integration into any bioinformatics pipeline. Several features make CovReport especially useful in a diagnostic setting. First, the application runs on a local computer allowing complete data security. Second, the patient's name from the previous analysis is automatically reset upon entering the new coverage file in order to avoid identity errors. Third, the format, content and the language of the report can be easily customized by the laboratory, adapting to any diagnostic setting. Fourth, information about additional Sanger re-sequencing of suboptimally covered regions can be integrated into the coverage report of the initial short-read sequencing allowing easy tracing of sequencing experiments.

Exon coverage for genes on the panel is visualized by drawing exons shaded according to the level of coverage: 100% covered exons are white, 90-100% exons are light gray, <90% covered exons are dark gray, non-covered exons are black. Several visualization options allow adapting the graphical presentation to the user's needs. Genes with 100% coverage can be listed at the top of the report without drawing the exons (*Skip white genes* option, default). Schematic gene/exon structures will be drawn for the remaining genes, shading the lower-coverage exons. If *Merge white exons* option is selected, CovReport will fuse exons covered at 100% for more compact representation, which is useful for genes with numerous exons. Unchecking this option will produce the report with each individual exon drawn. Similarly, exons with the same shading can be merged for more compact representation (*Merge non-white exons* option). Average per-gene coverage can be shown next to each gene (*Show gene weighted coverage* option), which is also used for statistics of the total gene panel coverage (*Show statistics* option). RefSeq transcript information can be included in the report (*Show gene transcripts* option), recommended if transcripts differ in exon number leading to differences in coverage between isoforms. The default 20x depth of coverage in the report can be changed to 1x, 5x, 10x or 30x. Finally, additional comments can be included in the report.

Use case

CovReport has been used for all diagnostic gene panel sequencing tests at the Laboratory of Molecular Genetics at the Timone Hospital (Marseille, France) since June 2018. The reports generated are directly annexed to the diagnostic results of various sequencing panels. Moreover, the coverage reports produced by CovReport are also routinely used for technical validation of the test since gene coverage statistics for the panel are integrated in the output.

To demonstrate the performance of CovReport, we used exome sequencing data for NA12878 to visualize the coverage of 40 genes on the extended Limb-Girdle Muscular Dystrophy (LGMD) gene list defined by the French National Network for Rare Neuromuscular Diseases⁵. Coverage module of VarAFT⁵ was used to obtain exon coverage for these genes. The default options were used to generate the pdf report using CovReport with *Show gene transcript* and *Show statistics* options activated. As seen from Figure 1, coverage was visualized for 43 transcripts corresponding to 40 genes. Most of the genes were covered at 100%. The advantage of using *Show gene transcript* option is clear from this example, since *TTN (NM_133379)* is covered at 100%, while *TTN (NM_001267550)* has several exons with lower coverage. Indeed, *NM_133379* isoform is much shorter, as exons 50-219 of *NM_001267550* are not included in this transcript. Six exons had no regions with above 20x threshold coverage (black exons). Several other exons had suboptimal coverage in regions corresponding to more than 10% of their length (dark gray). If similar coverage of these genes is obtained after a diagnostic sequencing for a patient affected with LGMD, the prescribing clinician will take into account the presentation of the disease in order to decide if the phenotype of the patient could be potentially explained by pathogenic variants in the exons with low coverage. If that is the case, additional resequencing of these regions will be necessary.

Discussion

Importance of coverage data is well established in genetic diagnostics. Indeed, most sequencing reports provide some information about coverage of target regions. However, these data are usually presented as an overall gene panel coverage or as an average-per-gene coverage. To our knowledge, no exon-level coverage data are provided with diagnostic reports. However, the distribution of pathogenic variants in a given gene is not homogeneous, as the disease-causing amino acid changes are often clustered in functionally important regions of the protein. Thus, information about coverage of the regions where disease-causing variants are concentrated is much more informative than an overall average gene coverage⁶. CovReport highlights individual exons with suboptimal coverage, facilitating the interpretation of the sequencing data.

The ultimate goal of diagnostic sequencing is to obtain an above-threshold coverage for the entire genomic area targeted by a given gene panel. This is feasible for small panels with genes that can be robustly sequenced. However, this goal is much harder to achieve for larger gene panels containing difficult to sequence genes. This is the case for the neuromuscular disorder field, since 33 out of 203 genes on the consensus myopathy gene lists⁵ contain “dark” regions of the genome that are not easily accessible using standard short-read sequencing approaches¹. Disease-causing variants in these regions can therefore be overlooked leading to a false negative molecular diagnostic result. CovReport effectively highlights and draws attention to the exons with suboptimal coverage, allowing the prescriber to evaluate the need for a complementary re-sequencing test for these areas.

Since the arrival of high-throughput sequencing technology, several tools have been developed for calculating sequencing coverage, such as BEDTools *genomcov*^{3,4}, *Sambamba*⁷ and *Mosdepth*⁸. Diagnostic laboratories use one or combination of these tools as part of their standard sequencing test analysis pipelines. However, the optimal way to visualize and evaluate the obtained coverage data is not universally defined. Several tools have been designed to evaluate the coverage of target regions after short-read sequencing in a diagnostic setting^{2,9,10}. While these tools are useful to monitor quality control of runs and samples during technical validation of sequencing tests, unlike CovReport, they do not generate exon-level coverage visualization that can be directly annexed to the diagnostic report. To our knowledge, CovReport is the first tool that allows comprehensive and visual evaluation of coverage results for a sequencing test, allowing molecular geneticists or prescribing physicians to make important diagnostic decisions, such as pursuing further explorations of the poorly covered genomic regions.

In conclusion, CovReport generates a one-glance graphical overview of coverage for individual exons on a gene panel, facilitating interpretation of the sequencing test. CovReport is flexible and easy-to-use, making it easy to adopt in any diagnostic setting.

Acknowledgements

We would like to thank the initial users of CovReport at the Laboratory of Molecular Genetics at the Timone Hospital, Marseille (Caroline Lacoste, Véronique Blanc, Christophe Pecheux and Amandine Boyer) as well as Henri Pégeot at the Laboratory of Molecular Genetics, Montpellier, France for the valuable comments and suggestions for improvement. We would also like to thank the Genomics&Bioinformatics Platform at MMG for providing the exome sequencing data for NA12878.

References

1. Ebbert, M. T. W. *et al.* Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* **20**, 97 (2019).
2. Desvignes, J.-P. *et al.* VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res.* **46**, W545–W553 (2018).
3. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* vol. 26 841–842 (2010).
4. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics* vol. 47 11.12.1–11.12.34 (2014).
5. Krahn, M. *et al.* A National French consensus on gene lists for the diagnosis of myopathies using next-generation sequencing. *Eur. J. Hum. Genet.* **27**, 349–352 (2019).
6. Gorokhova, S. *et al.* Comparing targeted exome and whole exome approaches for genetic diagnosis of neuromuscular disorders. *Appl Transl Genom* **7**, 26–31 (2015).
7. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* vol. 31 2032–2034 (2015).
8. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
9. Doig, K. D. *et al.* PathOS: a decision support system for reporting high throughput sequencing of cancers in clinical diagnostic laboratories. *Genome Med.* **9**, 38 (2017).
10. Münz, M. *et al.* CoverView: a sequence quality evaluation tool for next generation sequencing data. *Wellcome Open Res* **3**, 36 (2018).

Availability: CovReport is available at <http://jdotsoft.com/CovReport.php>.

Competing interests: The authors declare that they have no competing financial or non-financial interests.

Authors contributions: M.G. wrote the software, participated in the designing of the tool and in writing of the manuscript. M.C., J.M., F.R., M.B. and M.K. participated in the testing of the tool and commented on the manuscript. S.G. conceived the tool, participated in the designing of the tool and wrote the manuscript.

Appendix 1. Coverage of target sequences

Patient name: NA12878 MedExome

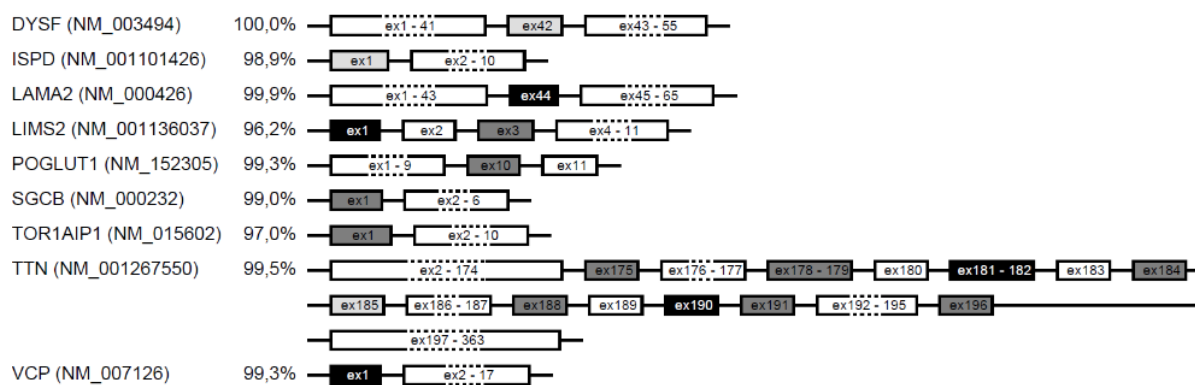
The following genes are 100% covered (with a depth of at least 20x):

ANO5 (NM_213599), BVES (NM_147147), CAPN3 (NM_000070), CAV3 (NM_033337), COL6A1 (NM_001848), COL6A2 (NM_001849), COL6A3 (NM_004369), DAG1 (NM_001165928), DES (NM_001927), DNAJB6 (NM_058246), DPM3 (NM_153741), FHL1 (NM_001159702), FKRP (NM_024301), FKTN (NM_001079802), GAA (NM_000152), GAA (NM_001079803), GMPPB (NM_013334), GNE (NM_001128227), HNRNPDL (NM_001207000), HNRNPDL (NM_031372), LMNA (NM_170707), MYOT (NM_006790), PLEC (NM_000445), POMGNT1 (NM_001243766), POMT1 (NM_007171), POMT2 (NM_013382), SGCA (NM_000023), SGCD (NM_000337), SGCG (NM_000231), TCAP (NM_003673), TNPO3 (NM_012470), TRAPPC11 (NM_021942), TRIM32 (NM_012210), TTN (NM_133379)

Certain parts of the following genes are not 100% covered (see details below)

Legend:

- exons covered at 100% (with a depth of at least 20x)
- exons covered at 90-100% (with a depth of at least 20x)
- exons covered at <90% (with a depth of at least 20x)
- exons not covered (with a depth of at least 20x)



Analysis statistics:

- number of genes: 40
- number of transcripts: 43
- transcripts covered at 100% (with a depth of at least 20x): 79,1%
- transcripts covered at 90-100% (with a depth of at least 20x): 20,9%

Figure1. Example of a coverage report provided by CovReport.

An example report was generated by CovReport in order to visualize the sequencing coverage of 40 genes on LGMD (Limb Girdle Muscular Dystrophies) panel using exome sequencing data for the standard DNA sample NA12878. The name of the patient is listed as specified in the CovReport interface. The date of the report is generated automatically (if option “Show report date” is activated as in this example). In the top part of the report, the genes covered at 100% are listed (if option “Skip white genes” is activated). Next, genes with sub-optimal coverage are shown with the exon shading according to the coverage level. Percentage of gene-level coverage is shown for each sub-optimally covered gene (if option “Show gene weighted coverage” is activated). Finally, the coverage statistics for the genes analyzed is shown (if option “Show statistics” is activated).

Supplementary data – Example of an input file for CovReport.

Article 2 : Recommandations françaises pour l'interprétation de variations nucléotidiques générées par les analyses en séquençage haut-débit

L'article est actuellement en cours de validation par l'ANPGM. Il sera soumis aux Annales de Biologie Clinique prochainement.

La version 2 des recommandations en cours de validation est jointe en Annexe 2.



Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS

Date de création : 6/3/2020

Date de révision :

Version : 2

Date de 1^{ère} Application : en atteinte

N° document : NGSDIAG_001

Approbation par le board du réseau : en cours

Ce document a été préparé par les membres du Groupe de Travail du Réseau NGS-Diag listés ci-dessous :

- Jean MULLER (Strasbourg ; représentant Réseau NGS-Diag)
- Cécile ROUZIER (Nice; représentant Réseau NGS-Diag)
- Nicolas CHATRON (Lyon ; représentant Réseau ACHro-Puce)
- Boris KREN (La Pitié ; représentant Réseau ACHro-Puce)
- Caroline ROORYCK-THAMBO (Bordeaux ; représentant Réseau ACHro-Puce)
- Martine DOCO FENZY (Reims ; représentant ACLF)
- Sandrine CAPUTO (Paris Curie ; représentant GGC)
- Nicolas DERIVE (Paris Curie ; représentant GGC)
- Claude HOUDAYER (Curie ; représentant GGC)
- Sylvie BANNWARTH (Nice ; ANPGM, mitochondrie)
- Anne BARLIER (Marseille ; ANPGM, FNMR FIRENDO)
- Svetlana GOROKHOVA (Marseille ; ANPGM)
- Claude JARDEL (La Pitié ; ANPGM, mitochondrie)
- Cédric LE MARECHAL (Brest ; ANPGM, FNMR SENSEGENE)
- Gaël NICOLAS (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience)
- Pascale RICHARD (Paris La Pitié ; ANPGM, FNMR CARDIOGEN)

Coordination :

Martin KRAHN (Marseille; ANPGM, FNMR FILNEMUS)

Pascale SAUGIER-VEBER (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience),

Article 3 : Systematic functional assessment of impact on splicing of deep exonic missense variants in CAPN3

Submission Confirmation

Thank you for your submission

Submitted to Human Mutation
Manuscript ID humu-2020-0112
Title Splicing impact of deep exonic missense variants in CAPN3 explored systematically by minigene functional assay
Authors Dionnet, Eugenie
Defour, Aurelia
Da Silva, Nathalie
Salvi, Alexandra
Levy, Nicolas
Krahn, Martin
Bartoli, Marc
Puppo, Francesca
Gorokhova, Svetlana
Date Submitted 15-Mar-2020

Systematic functional assessment of impact on splicing of deep exonic missense variants in CAPN3.

Eugénie Dionnet*¹, Aurélie Defour*¹, Nathalie Da Silva¹, Alexandra Salvi¹, Nicolas Lévy^{1,2}, Martin Krahn^{1,2}, Marc Bartoli¹, Francesca Puppo#^{1,3} and Svetlana Gorokhova#^{1,2}

¹ Aix Marseille Univ, INSERM, MMG, U 1251, Marseille, France.

² Service de génétique Médicale, Hôpital de la Timone, APHM, Marseille, France

³ CNRS, IBDM, UMR7288, Marseille, France

* both authors should be considered as co-first authors

both authors should be considered as co-last authors

Corresponding authors: Francesca Puppo and Svetlana Gorokhova

Corresponding author's postal address: Marseille Medical Genetics, U 1251, Aix Marseille Université, Faculté des Sciences Médicales et Paramédicales Ecole de Médecine, 27 bd Jean Moulin 13385 Marseille, France.

Corresponding author's phone: +33 4 91 32 49 06, fax +33 4 91 80 43 19

Corresponding authors' e-mail addresses:

francesca.puppo@univ-amu.fr; svetlana.gorokhova@univ-amu.fr

ABSTRACT

Improving the accuracy of variant interpretation during diagnostic sequencing is a major goal for genomic medicine. In order to explore an often overlooked splicing effect of missense variants, we developed the functional assay (“minigene”) for the majority of exons of *CAPN3*, the gene responsible for Limb Girdle Muscular Dystrophy (LGMD). By systematically screening 21 missense variants distributed along the gene, we found that eight clinically relevant missense variants located at a certain distance from the exon/intron borders (deep exonic missense variants) disrupted normal splicing of *CAPN3* exons. Several recent machine learning based computational tools failed to predict splicing impact for the majority of these deep exonic missense variants, highlighting the importance of including variants of this type in the training sets during the future algorithm development. Overall, 24 variants in *CAPN3* gene were explored, leading to the change in the ACMG classification of seven of them when results of the “minigene” functional assay were taken into account. Our findings reveal previously unknown splicing impact of several clinically important variants in *CAPN3* and draw attention to the existence of deep exonic variants with a disruptive effect on gene splicing that could be overlooked by the current approaches in clinical genetics.

INTRODUCTION

Calpainopathy, or Limb Girdle Muscular Dystrophy 2A or R1 (LGMD2A, LGMDR1; MIM# 253600), the most frequent form of Limb Girdle Muscular Dystrophy (LGMD) worldwide, is an autosomal recessive disorder characterized by muscle weakness affecting predominantly proximal limb muscles, elevated serum creatine kinase and necrosis/degeneration process observed in muscle biopsy (Fardeau et al. 1996). This disease is caused by pathogenic variants in *CAPN3* (MIM# 114240, 15q15.1, NM_000070.2), a gene coding for 94 kDa protein calpain3 (Calcium-dependant papain-like protease, P20807) that is a muscle-specific member of the calpain family of calcium-dependent enzymes. Calpain3 acts in muscle sarcomere formation and remodeling (Duguez, Bartoli, and Richard 2006). *CAPN3* is the most commonly mutated gene in patients presenting with Limb Girdle Muscular Dystrophy (Duno et al. 2008), making it one of the first genes to be sequenced as part of the diagnostic work up of patients with this type of myopathy (Krahn et al. 2019). More than 400 unique pathogenic or likely pathogenic variants have been reported in *CAPN3* gene so far (359 in LOVD (accessed December 2, 2019) and 555 variants in ClinVar (version November 27, 2019)). However, given the large size of the gene (24 exons, 3316nt), previously unknown variants in *CAPN3* are still being identified in diagnostic laboratories. Protein truncating variants (PTV) in *CAPN3*, such as frame-shift inducing indels, nonsense variants or variants disrupting canonical splicing sites (+/-2 nucleotides from exon-intron junctions), are generally accepted as pathogenic loss-of-function, especially if calpain3 protein was absent in Western Blot performed on muscle biopsy sample. Assigning the clinical significance to a newly identified exonic single nucleotide variant (SNV, either synonymous or missense) is much more difficult. This challenging task is now greatly assisted by ACMG Recommendations for variant classification (Krahn et al. 2019; Richards et al. 2015) that takes into account the frequency of identified variant in the general population, the results of algorithms predicting the impact of the variant on protein function, previously available clinical reports involving this variant as well as other elements. It is important to note that synonymous variants or missense variants with low impact on protein function can still be pathogenic by disrupting splicing (Duguez, Bartoli, and Richard 2006; Kergourlay et al. 2014). There is a great risk of misinterpreting the clinical significance of these variants using standard approaches, which could potentially lead to incorrect diagnosis. Unfortunately, it is not currently known if certain exonic regions are more likely to harbour splicing-affecting variants. Even though several splicing prediction scores are now available (Rowlands, Baralle, and Ellingford 2019), often it is difficult to estimate their validity in the absence of functional analysis.

In order to identify exonic single nucleotide variants with effect on splicing and improve the diagnostics of LGMDR1calpainopathies, we developed systematic functional cell-based assay (Minigene) for the majority of exons in *CAPN3* gene. We then selected a representative set of 21 missense variants

located outside of exon/intron junction sites (deep exonic variants) and tested the effect of these variants on splicing using the developed Minigene assay. We observed that eight out of 21 selected deep exonic missense variants induced abnormal splicing, leading to a change in the classification of clinical significance for half of them. Interestingly, most of the deep exonic variants impacting splicing were not identified by several splice-prediction algorithms tested, highlighting the critical need for robust methods of functional analysis of putative variants disrupting splicing. Thus, in addition to the direct benefit for diagnostics of LGMD2R1 patients, our study draws attention to the existence of deep exonic variants with a strong disruptive effect on gene splicing that could be overlooked by the current approaches in clinical genetics.

MATERIALS AND METHODS

Functional splicing assay for CAPN3 gene

Minigene reporter assay (Gaidrat et al. 2010) was developed for 18 CAPN3 exons. Since this assay is not adapted for U12 type introns, The Intron Annotation and Orthology Database (IAOD)(Gault et al. 2017; Turunen et al. 2013) was used to identify this type of introns, leading to exclusion of exons 19 and 20 from the analysis. As described previously (Kergourlay et al. 2014; Puppo et al. 2015), exons and approximately 150 bp of flanking introns were amplified using the Expand high fidelity PCR system (Roche, Basel, Switzerland). Amplicons were subsequently cloned into the pCAS2 vector. PCR and digestion product purifications were performed using the NucleoSpin Gel and PCR clean Up (Macherey Nagel, Düren, Germany), ligations by Quick ligation kit (Biolabs address), and transformations in 10-beta Electrocompetent E. coli (Biolabs). Variants were introduced into the wild type constructions using the Quick-change II XL site-directed mutagenesis kit (Agilent, Santa Clara, CA, USA). Presence of an abnormally spliced transcript associated with the decrease of the normal transcript was considered as "Impact on splicing". If an abnormally spliced transcript was present but no decrease of the normal transcript was observed for the mutated construct compared to control, the minigene assay conclusion was "Mild impact on splicing".

Cell culture and transfection

C2C12 cells were cultured in DMEM, high glucose, GlutaMAX supplemented with pyruvate, 10% Fetal bovine serum (Lifetechnologies, address) and 1% antibiotic and antimycotic. Transfections were made with Fugene HD (Promega, address) according to the manufacturer's instructions.

Transcriptional study

RNAs were isolated 48 hours after cell transfection, using Trizol/chloroform and DNA Free Removal kit (Invitrogen). RNAs were reverse-transcribed into cDNA and amplified using SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Invitrogen). PCR amplifications were performed using pCAS-KO 1F and pCAS-2R, located in exon A and exon B of pCAS2 (Refs). PCR products were separated by electrophoresis in a 2% agarose gel stained with 0.5 µg/ml of ethidium bromide. A purification step from the agarose gel was performed when several transcripts were present in order to subclone each of them into pGEM®-T Easy Vector System I (Promega) and analyze them separately. Finally, sequencing of each transcript was performed using the Big DyeR Terminator V1-1 Cycle Sequencing Kit (Life Technologies) on ABI Prism 3130xl capillary DNA Sequencer.

Variant selection, annotation and classification.

CAPN3 transcripts ENST00000397163 and NM_000070.2 were used for all analyses. 328 unique missense CAPN3 variants were downloaded from LOVD website for CAPN3 gene (<https://databases.lovd.nl/shared/genes/CAPN3>, accessed on December 2, 2019). 286 unique missense variants in CAPN3 present in ClinVar database were downloaded using Simple ClinVar website (<http://simple-clinvar.broadinstitute.org/>, version of ClinVar database: November 27, 2019, Perez-Palma2019). 447 CAPN3 missense variants found in the general population were downloaded from gnomADv.2.1.1 (Karczewski et al. 2019) using UCSC table browser (Karolchik et al. 2004). Only missense variants (i.e. exonic non-synonymous non-protein truncating single nucleotide substitutions)

were included in analysis. *CAPN3* missense variants from LOVD were assigned the “Clinical Significance” values based on the classification system used in ClinVar (uncertain/conflicting, likely pathogenic, pathogenic/likely pathogenic or pathogenic). We combined missense *CAPN3* variants from LOVD and ClinVar annotated as non benign in both databases, producing a set of 403 variants. Of these, 381 variants were located more than 1 or 2 nucleotides away from exon-intron junction. The clinical significance for LOVD variants were then expressed in the same terms as for ClinVar variants. The summary of the variants from this set are shown in the Supplemental Figure 1. We also included in the analysis missense variants identified in patients from the LGMDR1 cohort from the Department of Medical Genetics at the Timone Hospital (Marseille, France). Two Novel *CAPN3* variants from this cohort, as well as four previously reported variants, were deposited in the LOVD database (Fokkema et al. 2011). A representative set of 21 missense *CAPN3* variants located across 16 exons was selected for functional analysis by minigene (Figure 1, Table 1). These variants were located outside of canonical splice sites (+/- 2nt from exon junctions), were not previously known to disrupt splicing, but were predicted to have an effect on splicing according to Human Splice Finder (HSF version 3.0, <http://www.umd.be/HSF3> (Desmet et al. 2009). Three additional variants with uncertain pathogenicity classification were also tested by minigene approach: two variants located at the canonical splice junction sites c.498G>A, p.(Gln166Gln) and c.1913A>C, p.(Gln638Pro) as well as one synonymous variant c.984C>T, p.(Cys328Cys) (Supplementary table 1). The REVEL (Ioannidis et al 2016) and CADD scores (ref) for all the variants in this study were obtained using Variant Ranker (<http://vsranker.broadinstitute.org/>, Variant Ranker: a web-tool to rank genomic data according to functional significance). SpliceAI scores for the variants were obtained using SpliceAI tool (v1.3, <https://github.com/Illumina/SpliceAI> (Jaganathan et al. 2019). MMSplice (Cheng et al. 2019). Variant impact of calpain3 protein structure was assessed using VarMap web tool (Stephenson et al. 2019). SCAP scores were downloaded from <http://bejerano.stanford.edu/scap/> (Jagadeesh et al. 2019). MutPred Splice annotation was done using <http://www.mutdb.org/mutpredsplice> (Mort et al. 2014). MaxEntScan scores (Yeo and Burge 2004) were obtained using <http://www.umd.be/HSF3>. Variants with MaxEnt threshold score of 3 or a score difference of more than 30% with the wild-type score were considered as splice disrupting. For 13 out of 24 variants tested, the output was “No result found with this matrix”.

The pathogenicity of 24 functionally tested variants was scored before and after the minigene results according to ACMG criteria (Richards et al. 2015) with the following modifications. PS3 score was assigned if calpain3 protein has been previously reported to be absent on the muscle biopsy Western Blot (PS3_moderate if reduced expression) in at least two patients either homozygous for the variant or compound heterozygous *in trans* with a confirmed pathogenic variant (PS3_moderate if absent in one patient, PS3_supporting if decreased in one patient). PS3 score was assigned for variants with a minigene result “Impact on splicing”, PS3_supporting for “Mild impact on splicing”. Individuals with the examined variant classified as VUS were not counted for PS4 score. PP2, PP4 and PP5 criteria were not used, consistent with several other recent updates to the original ACMG criteria (Gelb et al. 2018; Lee et al. 2018; Kelly et al. 2018). PP3 score was assigned for variants with REVEL score above 0.7. PP3 was also assigned if two out of three algorithms (HSF, MaxEntScan, SpliceAI) indicated an impact on splicing. PM1 score was assigned for variants affecting highly conserved positions located in the Calpain catalytic domain (PF00648, amino acids 74-417), in Calpain III domain (PF01067, amino acids 436-579) or in EF-hand domains (amino acids 649-683; 692-718; 722-757; 787-821), as visualised by VarMap web tool. PM3 score (*in trans* with a pathogenic variant) was assigned according to SVI Recommendation for *in trans* Criterion (PM3) - Version 1.0). PP1 score (segregation data) was assigned according to recommendations by the Hearing Loss ClinGen Working group that focused in part on recessive disorders (Oza et al. 2018). The cut-offs for allele frequency criteria were attributed as following: PM2 - 0.02%, BS1 - 0.5%, BA1 - 5%.

Data analysis and visualization was performed with R 3.5.3 using packages dplyr (v.0.8.5, <https://dplyr.tidyverse.org/>) and ggplot2 (3.2.1, <https://ggplot2.tidyverse.org/>). ProteinPaint

(<https://pecan.stjude.cloud/proteinpaint>, (Zhou et al. 2016)) was used to visualize variants on the CAPN3 gene structure in Figure 1.

RESULTS

Functional analysis of deep exonic CAPN3 missense variants using minigene assay.

In order to assess the splicing impact of missense variants located further into exon from the exon-intron junctions (deep exonic missense variants), we used minigene assay to functionally test 21 clinically relevant variants homogeneously distributed along the CAPN3 gene (Figure 1A, Table 1). Out of the 21 variants tested, 8 showed splicing defects in the minigene assay. Of these, five variants induced expression of a shorter transcript lacking one exon (exon skipping), associated with the decrease in the normal transcript (c.551C>G, p.(Thr184Arg); c.551C>T, p.(Thr184Met); c.743T>G, p.(Met248Arg); c.1063C>G, p.(Arg355Gly) and c.1466G>A, p.(Arg489Gln) (Figure 2). In all of these cases, the skipping of the exon was associated with the loss of the reading frame, probably inducing a truncated protein or no protein at all if nonsense mediated decay of the abnormal transcript is induced. Thus, the amount of calpain3 protein present in muscle cells of patients carrying these variants is likely to be lower than normal.

Minigene assay for variant c.1187A>G, p.(Glu396Gly) showed presence of a transcript with an abnormal inclusion of the first 31 nucleotides of intron 9 (Figure 3A). The normal transcript was present at a lower level compared to control. Decrease in the normal transcript associated with predicted frameshift in the abnormally spliced transcript will likely lead to lower expression of normal calpain3 protein. Interestingly, minigene assay did not show any splicing abnormality for another nucleotide substitution at the same position (variant c.1187A>T, p.(Glu396Val)).

Minigene assays for two variants (c.1622G>A, p.(Arg541Gln) and c.2306G>A, p.(Arg769Gln)) showed presence of abnormally spliced transcripts, but the normal transcript expression seemed to be comparable to that of control (Figure 3B, C). We therefore concluded that the impact of these variants on splicing is likely to be milder compared to the variants described above.

Localization of deep exonic variant with effect on splicing within exons.

In order to explore the localization of splice-affecting exonic missense variants, we calculated distances from these variants to 5' and 3' ends of the exons. These variants were located in seven different exons (4, 5, 8, 9, 11, 13 and 22). For exons larger than 100nt, we plotted the splice-affecting exonic variants based on their distance from either 5' or from 3' exon ends (Figure 4A). For two smaller exons (exons 8 and 9), distances from both ends of the exon are shown (Figure 4B). Interestingly, two deep exonic variants with impact on splicing (c.743T>G, p.(Met248Arg) in exon5 and c.1466G>A, p.(Arg489Gln) in exon 11) were located at the exact same distance of 59 nucleotides from the 3' end of the exon.

We then screened all known CAPN3 missense variants located in deep exonic regions for potential effect on splicing. To do so, we constituted a set of clinically relevant missense variants (pathogenic, likely pathogenic and uncertain/conflicting) from LOVD and ClinVar database. We then selected only the missense variants located outside of canonical splice sites and used SpliceAI score (Jaganathan et al. 2019) to estimate the putative effect on splicing for each variant. The obtained variants were plotted along the CAPN3 gene with exon boundaries marked (Figure 1B). Deep exonic missense variants present in the general population (gnomADv2.1.1) were also analysed (Figure 1C). As seen from these results, certain exonic regions contained clusters of variants with higher SpliceAI scores.

Splice-prediction algorithm scores do not correlate with the minigene assay results for deep exonic CAPN3 missense variants

Despite critical need for genetic diagnostics, the currently available splice effect estimation algorithms are still not optimal to correctly predict the impact on splicing, in part due to insufficient number of time-consuming functional validation studies. Our set of CAPN3 variants functionally tested using the well-established minigene assay provide an excellent opportunity to test recent splice prediction

algorithms. All 21 variants tested in this study were predicted to have an impact on splicing according to HSF, yet only eight showed splicing alteration in the minigene assay, demonstrating a significant false positive rate. The splice effect predictions by MaxEntScan (Yeo and Burge 2004) are shown in Figure 5A. Five out of eight variants were correctly predicted as splice-disrupting by MaxEntScan; for the remaining three variants, no results were available with the matrix used. However, three variants were falsely predicted as splice disrupting. SpliceAI scores for the 21 variants are shown in Figure 5B. Only two out of eight variants associated with splicing anomaly on minigene assay had SpliceAI scores above 0.2 (c.1063C>G, p.(Arg355Gly) and c.2306G>A, p.(Arg769Gln)). Of these, c.2306G>A, p.(Arg769Gln) had a mild impact on splicing, since the presence of an abnormally spliced transcript was associated with unchanged level of normal transcript. Predictions of MutPred Splice, a machine learning-based algorithm designed specifically for exonic variants (Mort et al. 2014), was slightly better than SpliceAI, as it also identified c.551C>G, p.(Thr184Arg) variant (Figure 5). Taken together, 6/8 false_negatives for Splice AI and 5/8 for MutPred Splice. Both algorithms had one false positive - c.593A>G, p.(Asn198Ser). Interestingly, this variant was predicted as splice disrupting by all algorithms, yet showed no effect on splicing in minigene assay. Activation of an exonic cryptic donor site was predicted by HSF for this variant. None of the 21 variants with abnormal splicing in minigene assay has an above threshold SCAP score (Jagadeesh et al. 2019). We have also evaluated scores predicting variant's impact on protein function, such as REVEL (Ioannidis et al 2016). As shown in Figure 5D, all of the eight missense variants with impact on splicing were also predicted to be deleterious at the protein level. Indeed, five variants (p.Thr184Met, p.Thr184Arg, p.Met248Arg, p.Arg355Gly and p.Glu396Gly) affected conserved residues of the Calpain-like peptidase C2 domain (PF00648) . Two variants (p.Arg489Gln and p.Arg541Gln) changed conserved amino acids located in the Calpain III domain (PF01067), while p.Arg769Gln affected a conserved amino acid that is expected to participate in protein-protein interaction in a proximity of EF-hand domain (PF13202).

Pathogenic classification for seven out of 24 of variants was modified based on the results of the minigene assay.

Since most of the variants selected for testing by minigene were identified before the ACMG recommendations for assessing variant pathogenicity, we classified each variant according to the ACMG recommendations modified for *CAPN3* gene as described in Materials and Methods. The details of evidence used to classify each variant are available in the Supplementary table 1. The results of the minigene assay, allowing to assign additional PS3 or PS3_sup scores, were then taken into account to calculate "post-minigene ACMG classification" for each variant. For seven out of 24 variants tested by minigene ACMG classification was modified based on the results of this functional assay.

DISCUSSION

Despite recent advances in genomic medicine, the pathogenicity of most variants identified as part of a diagnostic sequencing test is often not well-established. Bioinformatic tools aimed at predicting the identified variant's effect on protein function or on mRNA splicing therefore play an important role in the diagnostic process (Hu et al. 2019). However, these algorithms do not perform equally well for all types of variants, ultimately requiring a functional confirmation of the pathogenicity prediction. Interpretation of variant pathogenicity is often guided by its location - intronic variants as well as exonic variants in the proximity of exon-intron junctions are usually suspected to disrupt mRNA splicing, while missense variants in the coding regions are considered as potentially altering protein function. Several studies have previously drawn attention to the possibility of exonic variants disrupting normal splicing (Savisaar and Hurst 2017; Kergourlay et al. 2014). In this study, we systematically screened 16 exons of *CAPN3* gene for this type of clinically important variants, by using minigene assay to functionally test missense variants located outside of canonical splice regions (deep exonic variants). We detected splicing abnormalities for eight out of 21 missense variants tested. Interestingly, these eight splice-disrupting missense variants were also predicted to have an impact on protein function due to the amino acid change. It is difficult to disentangle the pathogenic effect on

the protein from the effect on mRNA splicing for these variants. It is possible that both mechanisms contribute to the overall pathogenicity *in vivo*. The interplay between variant's effect on protein function and its impact on splicing could potentially have important consequences for cell function and disease expression. For example, the effect of dominant negative pathogenic variants on protein level could be downgraded if their splicing effect is increased, reducing the amount of toxic transcript. Recent studies have shown significant variability in mRNA splicing patterns between individuals and human populations in general (Lu, Jiang, and Xing 2012; Park et al. 2018). Thus, one could imagine that the frequency of a confirmed pathogenic dominant negative variant could be higher in a particular population if this variant has greater impact on splicing in this population. On the contrary, if the main pathogenic effect of a variant comes from its impact on splicing, pathogenic abnormally-spliced transcripts could be expressed at lower levels in certain individuals or populations carrying splice-modifying variants, leading to milder expression of the disease (Cooper et al. 2013). One example of such a variant could be c.551C>T:p.(Thr184Met), one of the variants inducing abnormal splicing in our minigene assay. This missense variant has long been considered as pathogenic, responsible for milder forms of LGMDR1 (Anderson et al. 1998; Richard et al. 1999; Fanin et al. 2004; Barp et al. 2020). This variant had very low frequency in the initially tested subpopulations (0.001% in Europeans, 2 in 129164 alleles in gnomADv2.1.1). However, with more sequencing data now available from different populations, it turns out this variant has 3.3% frequency in African and 0.09% in Latino populations, leading to its reclassification as Benign or Likely benign by many diagnostic laboratories (See Supplementary Table 1 for more details). Given our new evidence of splicing effect for this variant, more studies are needed in order to establish its role in LGMDR1. Another example of a splice-affecting variant with high population frequency (1% in Finnish) is c.1746-20C>G (Nascimbeni et al. 2010). Interestingly, differences in expression of abnormal transcripts was detected between LGMDR1 patients carrying this variant (Nascimbeni et al. 2010).

Given the high clinical importance of accurately detecting splice-disrupting variants, a large number of splice effect prediction algorithms have been developed. Earlier tools, such as MaxEntScan and HSF (Yeo and Burge 2004; Desmet et al. 2009), are still widely used in diagnostics and more recent machine-learning based approaches, such as SpliceAI, S-CAP and MMSplice, are also now gaining significant attention (Rowlands, Baralle, and Ellingford 2019; Jaganathan et al. 2019; Cheng et al. 2019; Jagadeesh et al. 2019). However, these prediction tools were designed and tested using the already known splice variants, most of which are intronic or located at canonical splice sites. In this study, we focused on potential splice-disrupting missense variants located deeper in exons, a group of variants that have been relatively little explored (Savisaar and Hurst 2017). We found the performance of the tools tested suboptimal, as they failed to predict the splicing effect for more than half of the variants showing impact on splicing in the minigene assay. At the same time HSF, used during initial selection of variants, predicted abnormal splicing for 21 variants, while only eight of them affected splicing in the minigene assay. Our results are consistent with a previous study of *BRCA1* and *BRCA2* that found that all variants with splice impact in minigene assay, but not predicted to disrupt splicing by five different algorithms were located in exons (Théry et al. 2011). Taken together, these findings suggest that exonic variants with effect on splicing might be more difficult to predict using currently available algorithms. The splice-disrupting exonic variants identified in this study could contribute to the new dataset of this type of variants in order to develop more efficient splice-predicting tools.

Exonic elements regulating splicing consist of ESEs (Exonic Splice Enhancers) that promote the splicing of the exon and ESSs (Exonic Splice Silencers) that induce exon exclusion from the mature mRNA (Ohno, Takeda, and Masuda 2018). The density of these elements tends to rise towards the exon-intron junctions (Fairbrother et al. 2004; Woolfe, Mullikin, and Elnitski 2010), thus increasing the probability that a variant present in these regions affects splicing. Interestingly, seven out of eight splice-disrupting missense variants identified in this study were located more than 30 nucleotides away from the exon border. Distribution of regulatory splicing information in the deep exonic regions is less clear. Two models have been previously proposed for functional splice elements in exonic regions - rare regions under strong purifying selection and multiple lowly constrained regions (Savisaar and Hurst 2017). Our results are more consistent with the second model, since none of the exonic variants

with impact on splicing described in our study are located in regions of high constraint (CCR > 90 (Havrilla et al. 2019)). Indeed, only a small region of exon 5 of *CAPN3* shows as constrained using this method, most likely due to the fact that the vast majority of *CAPN3* pathogenic variants are responsible for recessive disease.

Previous studies using minigene assays have estimated that about 30% of exonic variants affect splicing (Savisaar and Hurst 2017). Our results are consistent with these findings (eight splice-affecting variants out of 21 tested by minigene), even though all 21 variants tested were predicted to be splice-modifying by HSF, thus overestimating the proportion of splice-affecting exonic variants.

Accessibility and wide use of high throughput sequencing for limb-girdle myopathy diagnosis has led to identification of multiple variants of uncertain pathogenicity, increasing the demand for efficient functional studies. For LGMDR1, presence of calpain3 protein on Western Blot or screening for abnormal transcripts in muscle RNA has proven to be useful to achieve accurate molecular diagnosis. However, both of these approaches require access to patient's muscle biopsy material. A useful *in vitro* alternative is the well-established minigene assay. It has been previously shown that the results of minigene assay are concordant with the analysis of patient RNA (Tournier et al. 2008; Théry et al. 2011; Bonnet et al. 2008). Here, we developed the minigene approach for the majority of *CAPN3* exons and used it to test 24 variants. Abnormal splicing in minigene assay allowed assigning additional pathogenicity evidence points according to ACMG recommendations (Richards et al. 2015) for XX variants, leading to change in ACMG classification in XX of them. These results demonstrate the utility of minigene functional assay for diagnostics of calpainopathy and reinforce the need to incorporate functional studies into diagnostic process on a larger scale. The functional assay described here can be directly used by the diagnostic laboratories to test *CAPN3* variants with suspected splicing effect, allowing to establish long-awaited diagnosis for many LGMDR1 patients.

Acknowledgements We thank patients and their relatives for their participation in this study. We also acknowledge Alexandra Martins and Christiane Duponchel for generously providing the pCAS2 vector.

FIGURES

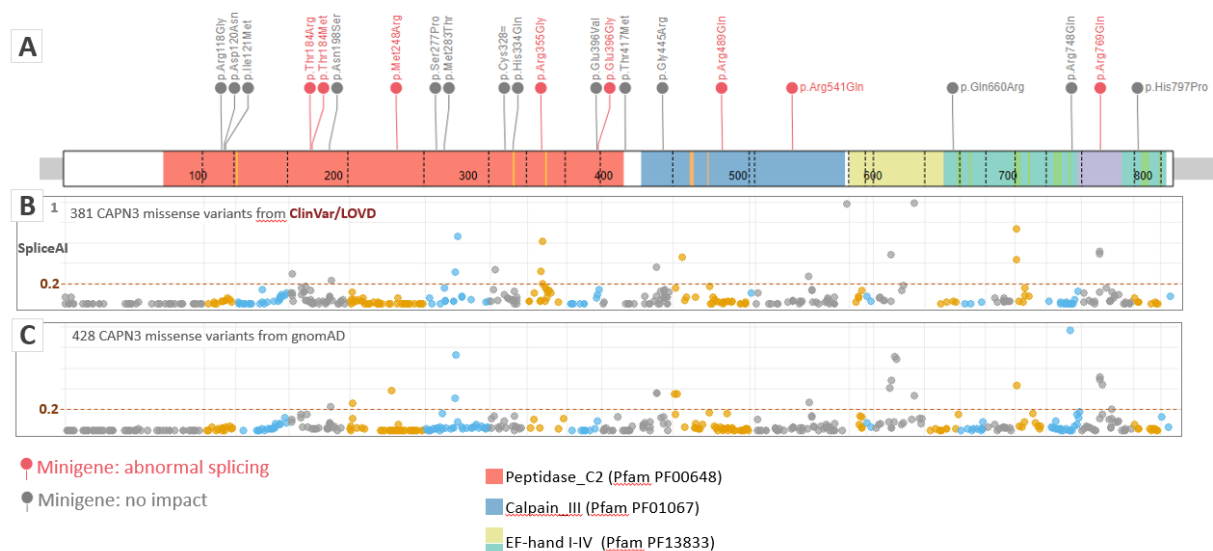


Figure 1: Deep exonic missense variants in *CAPN3* gene. **A.** Twenty one clinically relevant variants tested by minigene assay are shown (variants with effect on splicing are in red, variants without impact are in grey). Exon-intron junctions of the *CAPN3* gene as well as the functional domains of calpain 3 are visualized. **B.** Clinically relevant *CAPN3* variants (pathogenic, likely pathogenic and

uncertain/conflicting) from LOVD and ClinVar are visualized along the *CAPN3* gene with SpliceAI scores on the y axis. **C.** Deep exonic missense variants from gnomADv2.1.1 are visualized along the *CAPN3* gene with SpliceAI scores on the y axis. Variants from adjacent exons are colored differently to visualize the exon junctions. Vertical grids in panels B and C correspond to exon boundaries.

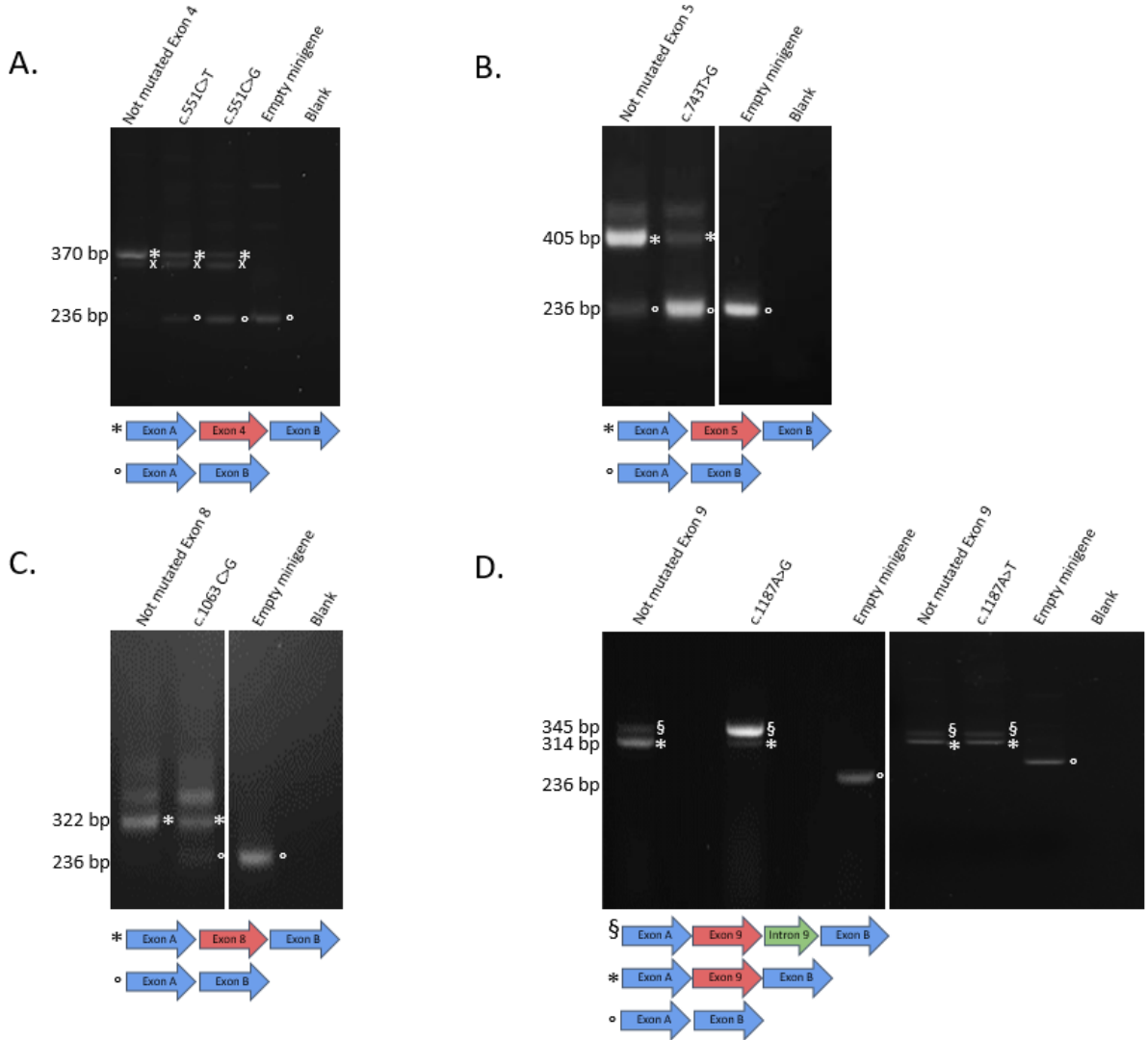


Figure 2: Missense *CAPN3* variants with splicing defects in the minigene assay (exon skipping). RT-PCR gel electrophoresis results from minigene reporter assay, with each band analysed using Sanger sequencing. **A.** Minigene assay for variants c.551 C>G, p.(Thr184Arg) and c.551 C>T, p.(Thr184Met), showing splicing abnormality for both variants. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 4 (same as empty minigene). This abnormally spliced transcript is absent from the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 4. The normal transcript is present at lower levels for the tested variants compared to control. A non-specific band present in both control and mutated constructs is labeled with an "x". **B.** Minigene assay for the variant c.743T>G, p.(Met248Arg), showing splicing abnormality. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 5 (same as empty minigene). A trace amount of the abnormally spliced transcript is present in the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 5. The normal transcript is present at much lower levels for the tested variant compared to control. **C.** Minigene assay for the variant c.1063C>G, p.(Arg355Gly) showing splicing abnormality. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 8 (same as empty minigene). This

abnormally spliced transcript is absent from the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 8. The normal transcript is present at lower levels for the tested variant compared to control. **D.** Minigene assay for the variant c.1466G>A, p.(Arg489Gln) showing splicing abnormality. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 11 (same as empty minigene). This abnormally spliced transcript is absent from the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 11. The normal transcript is present at lower levels for this variant compared to control.

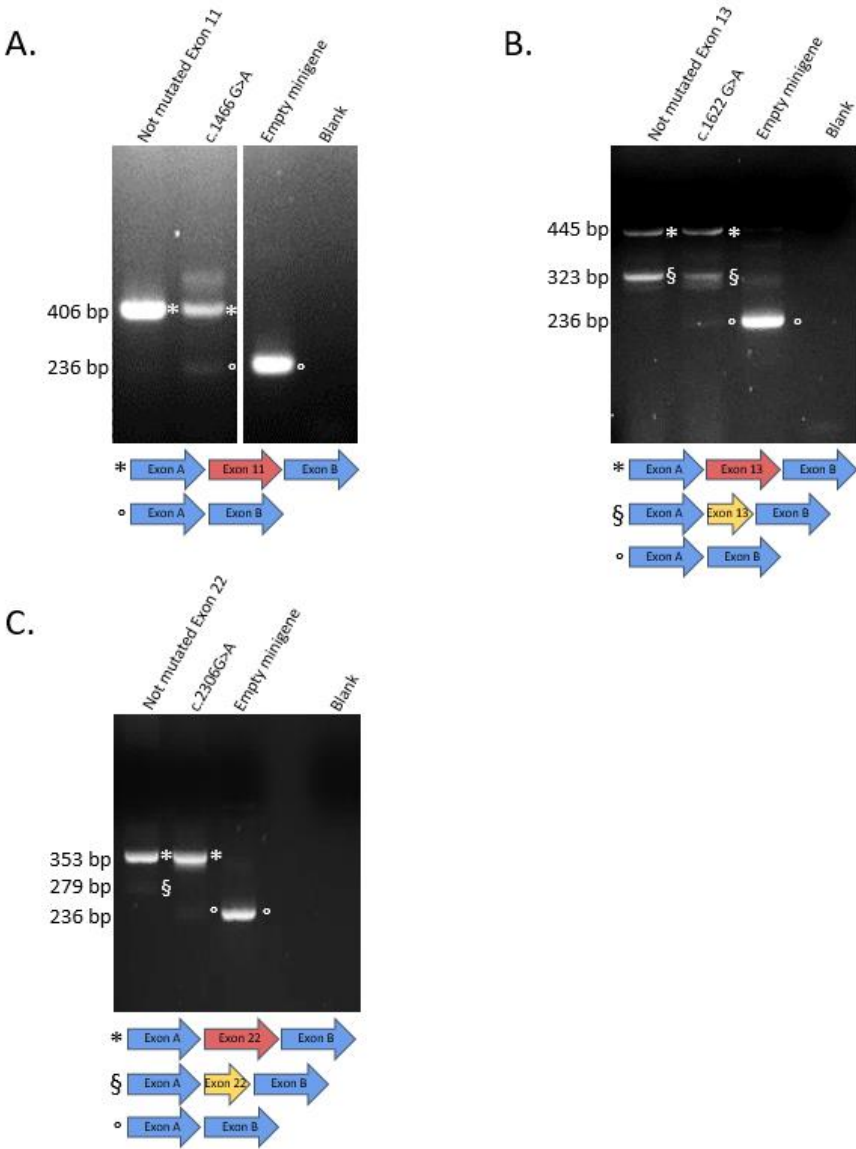


Figure 3: Missense CAPN3 variants with splicing defects in the minigene assay (other types of splicing abnormalities). RT-PCR gel electrophoresis results from minigene reporter assay, with each band analysed using Sanger sequencing. **A.** Minigene assay for variants c.1187A>G, p.(Glu396Gly) and c.1187A>T, p.(Glu396Val) showing splicing abnormality for c.1187A>G but not for c.1187A>T. The higher band (indicated with §) corresponds to an abnormal transcript containing exon 9 and the first 31 nucleotides of intron 9. This transcript is present at high levels for the construct with c.1187A>G variant. Trace amount of this abnormal transcript is also present in other conditions. The normal transcript (indicated with an asterisk, *) is present at much lower levels for the construct with c.1187A>G variant. There is no difference between transcripts observed for the construct with

c.1187A>T variant compared to the control. The empty minigene is indicated with a circle (o). **B.** Minigene assay for the variant c.1622G>A, p.(Arg541Gln) showing splicing pattern different from that of the control construct. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 13 (same as empty minigene). This abnormally spliced transcript is absent from the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 13. The normal transcript seems to be present at similar levels for mutated as well as for control constructs. A third band (indicated with §) corresponds to the exon 13 missing the last 87 nucleotides (no frameshift). This band is present at high levels in the control and at lower levels in the assay with the mutated construct. **C.** Minigene assay for the variant c.2306G>A, p.(Arg769Gln) showing splicing patterns different from that of the control construct. The lower band (indicated with a circle, o) corresponds to the transcript with skipped exon 22 (same as empty minigene). This abnormally spliced transcript is absent from the control lane with the non-mutated construct. The higher band (indicated with an asterisk, *) corresponds to the normal transcript containing exon 22. The normal transcript seems to be present at similar levels for mutated as well as for control constructs. A third band (indicated with §) corresponds to the transcript containing only the first 43 nucleotides of exon 22. This band is present in the control and absent in the assay with the mutated construct.

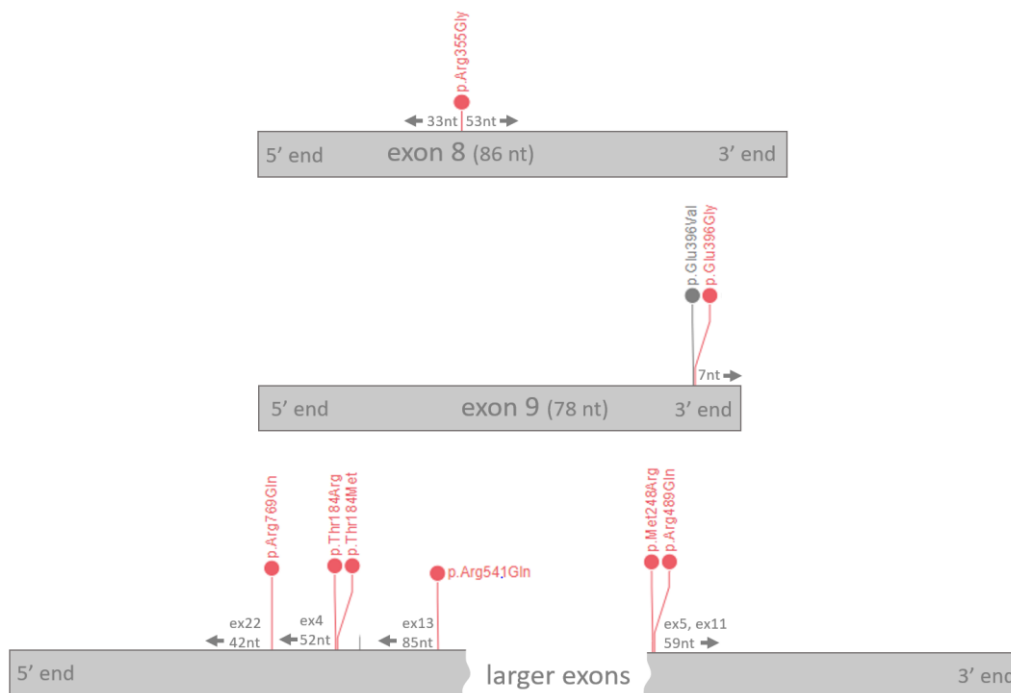


Figure 4: Distance from exonic borders of deep exonic variants with splice impact in minigene assay

Eight missense variants with impact on splicing (red) are visualized along the exons of *CAPN3* gene. Distances from both 5' and 3' ends are shown for smaller exons (< 100bp).

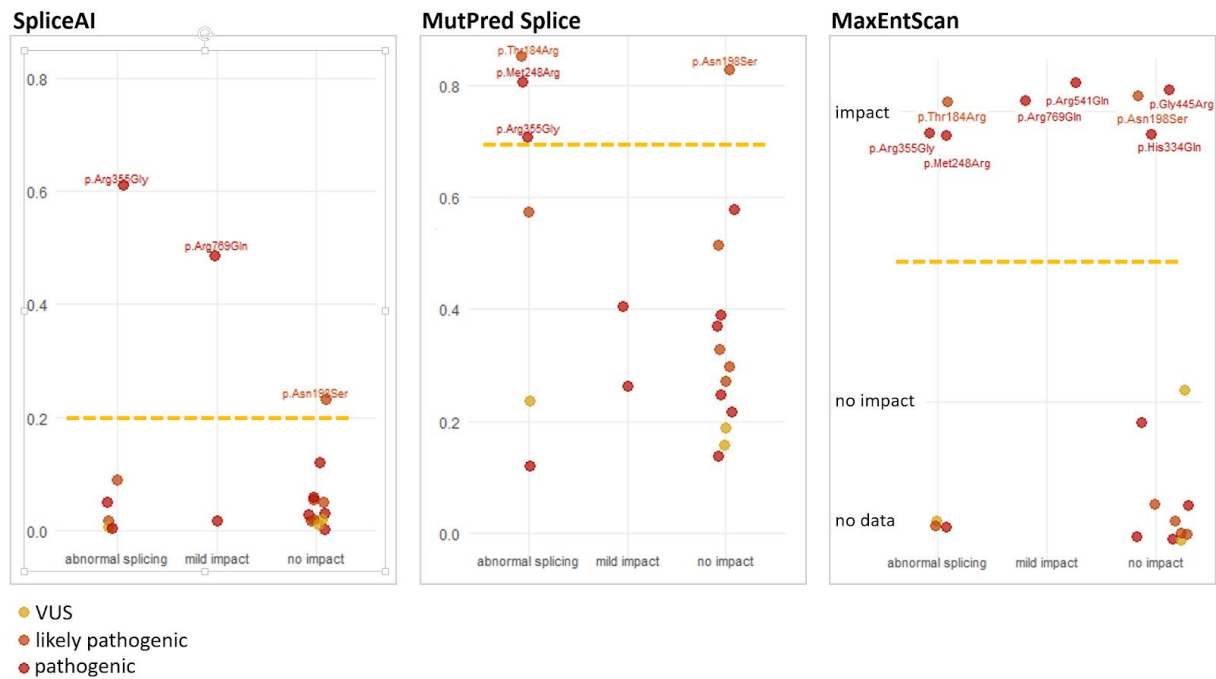


Figure 5: Performance of algorithms predicting impact of splicing for deep exonic variants
 Results of three splice prediction computational tools are shown for 21 *CAPN3* missense variants tested by minigene assay. The variants are separated into three groups depending on the minigene result: “Abnormal splicing”, “Mild impact” or “No impact”. The score cut-off above which a variant is predicted to affect splicing is 0.2 for SpliceAI and 0.7 for MutPred Splice.

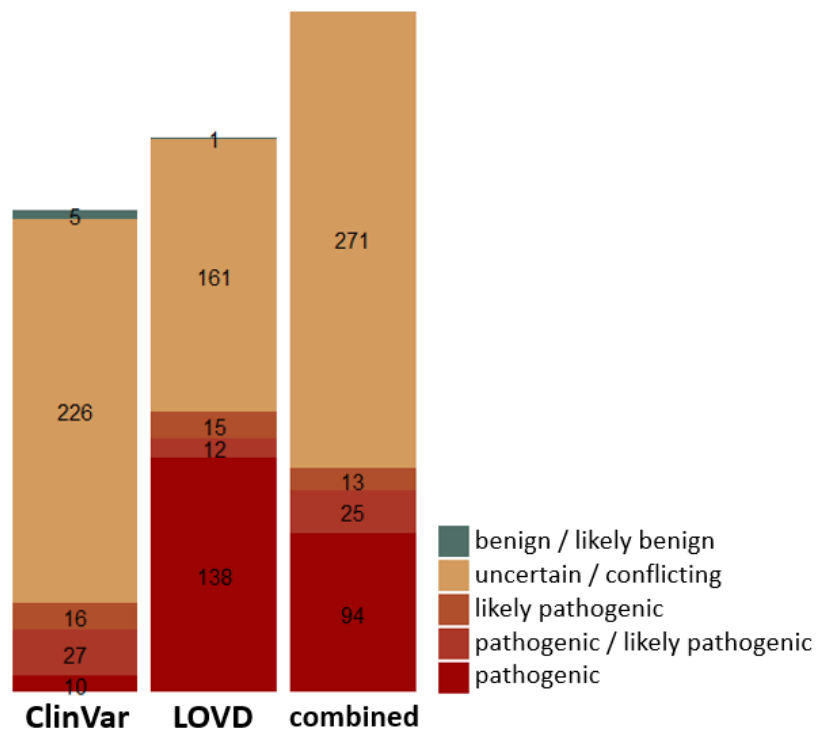


Figure S1: Clinically significant missense variants in *CAPN3* gene
 A set of 403 unique clinically significant missense variants was obtained by combining variants from LOVD and ClinVar annotated as non benign in at least one of these databases (327 variants from LOVD, 284 variants from ClinVar). *CAPN3* missense variants from LOVD were assigned the “Clinical

Significance” values based on the classification system used in ClinVar (uncertain/conflicting, likely pathogenic, pathogenic/likely pathogenic or pathogenic).

Table 1: Description of mutations selected to be tested by minigene assay.

Table 2: Results of splicing effects of mutations described in table 1 using minigene assay.

- Anderson, L. V., K. Davison, J. A. Moss, I. Richard, M. Fardeau, F. M. Tomé, C. Hübner, A. Lasa, J. Colomer, and J. S. Beckmann. 1998. “Characterization of Monoclonal Antibodies to Calpain 3 and Protein Expression in Muscle from Patients with Limb-Girdle Muscular Dystrophy Type 2A.” *The American Journal of Pathology* 153 (4): 1169–79.
- Bahrami-Samani, Emad, and Yi Xing. 2019. “Discovery of Allele-Specific Protein-RNA Interactions in Human Transcriptomes.” *American Journal of Human Genetics* 104 (3): 492–502.
- Barp, Andrea, Pascal Laforet, Luca Bello, Giorgio Tasca, John Vissing, Mauro Monforte, Enzo Ricci, et al. 2020. “European Muscle MRI Study in Limb Girdle Muscular Dystrophy Type R1/2A (LGMDR1/LGMD2A).” *Journal of Neurology* 267 (1): 45–56.
- Bonnet, C., S. Krieger, M. Vezain, A. Rousselin, I. Tournier, A. Martins, P. Berthet, et al. 2008. “Screening BRCA1 and BRCA2 Unclassified Variants for Splicing Mutations Using Reverse Transcription PCR on Patient RNA and an Ex Vivo Assay Based on a Splicing Reporter Minigene.” *Journal of Medical Genetics*. <https://doi.org/10.1136/jmg.2007.056895>.
- Cheng, Jun, Thi Yen Duong Nguyen, Kamil J. Cygan, Muhammed Hasan Çelik, William G. Fairbrother, Žiga Avsec, and Julien Gagneur. 2019. “MMSplice: Modular Modeling Improves the Predictions of Genetic Variant Effects on Splicing.” *Genome Biology* 20 (1): 48.
- Cooper, David N., Michael Krawczak, Constantin Polychronakos, Chris Tyler-Smith, and Hildegard Kehrer-Sawatzki. 2013. “Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease.” *Human Genetics* 132 (10): 1077–1130.
- Desmet, François-Olivier, Dalil Hamroun, Marine Lalande, Gwenaëlle Collod-Bérout, Mireille Claustres, and Christophe Bérout. 2009. “Human Splicing Finder: An Online Bioinformatics Tool to Predict Splicing Signals.” *Nucleic Acids Research* 37 (9): e67.
- Duguez, Stephanie, Marc Bartoli, and Isabelle Richard. 2006. “Calpain 3: A Key Regulator of the Sarcomere?” *FEBS Journal*. <https://doi.org/10.1111/j.1742-4658.2006.05351.x>.
- Duno, Morten, Marie-Louise Sveen, Marianne Schwartz, and John Vissing. 2008. “cDNA Analyses of CAPN3 Enhance Mutation Detection and Reveal a Low Prevalence of LGMD2A Patients in Denmark.” *European Journal of Human Genetics: EJHG* 16 (8): 935–40.
- Fairbrother, William G., Dirk Holste, Christopher B. Burge, and Phillip A. Sharp. 2004. “Single Nucleotide Polymorphism–Based Validation of Exonic Splicing Enhancers.” *PLoS Biology*. <https://doi.org/10.1371/journal.pbio.0020268>.
- Fanin, M., L. Fulizio, A. C. Nascimbeni, M. Spinazzi, G. Piluso, V. M. Ventriglia, G. Ruzza, et al. 2004. “Molecular Diagnosis in LGMD2A: Mutation Analysis or Protein Testing?” *Human Mutation* 24 (1): 52–62.
- Fardeau, M., B. Eymard, C. Mignard, F. M. S. Tomé, I. Richard, and J. S. Beckmann. 1996. “Chromosome 15-Linked Limb-Girdle Muscular Dystrophy: Clinical Phenotypes in Reunion Island and French Metropolitan Communities.” *Neuromuscular Disorders*. [https://doi.org/10.1016/s0960-8966\(96\)00387-2](https://doi.org/10.1016/s0960-8966(96)00387-2).
- Fokkema, Ivo F. A. C., Peter E. M. Taschner, Gerard C. P. Schaafsma, J. Celli, Jeroen F. J. Laros, and Johan T. den Dunnen. 2011. “LOVD v.2.0: The next Generation in Gene Variant Databases.” *Human Mutation* 32 (5): 557–63.
- Gault, Christine M., Federico Martin, Wenbin Mei, Fang Bai, Joseph B. Black, W. Brad Barbazuk, and A. Mark Settles. 2017. “Aberrant Splicing in Maize Reveals a Conserved Role for U12 Splicing in Eukaryotic Multicellular Development.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (11): E2195–2204.
- Gelb, Bruce D., for the ClinGen RASopathy Working Group, Hélène Cavé, Mitchell W. Dillon, Karen W. Gripp, Jennifer A. Lee, Heather Mason-Suares, et al. 2018. “ClinGen’s RASopathy Expert Panel Consensus Methods for Variant Interpretation.” *Genetics in Medicine*. <https://doi.org/10.1038/gim.2018.3>.
- Havrilla, James M., Brent S. Pedersen, Ryan M. Layer, and Aaron R. Quinlan. 2019. “A Map of Constrained Coding Regions in the Human Genome.” *Nature Genetics* 51 (1): 88–95.
- Hu, Zhiqiang, Changhua Yu, Mabel Furutsuki, Gaia Andreoletti, Melissa Ly, Roger Hoskins, Aashish N. Adhikari, and Steven E. Brenner. 2019. “VIPdb, a Genetic Variant Impact Predictor Database.” *Human Mutation* 40 (9): 1202–14.

- Jagadeesh, Karthik A., Joseph M. Paggi, James S. Ye, Peter D. Stenson, David N. Cooper, Jonathan A. Bernstein, and Gill Bejerano. 2019. "S-CAP Extends Pathogenicity Prediction to Genetic Variants That Affect RNA Splicing." *Nature Genetics* 51 (4): 755–63.
- Jaganathan, Kishore, Sofia Kyriazopoulou Panagiotopoulou, Jeremy F. McRae, Siavash Fazel Darbandi, David Knowles, Yang I. Li, Jack A. Kosmicki, et al. 2019. "Predicting Splicing from Primary Sequence with Deep Learning." *Cell* 176 (3): 535–48.e24.
- Karolchik, Donna, Angela S. Hinrichs, Terrence S. Furey, Krishna M. Roskin, Charles W. Sugnet, David Haussler, and W. James Kent. 2004. "The UCSC Table Browser Data Retrieval Tool." *Nucleic Acids Research* 32 (Database issue): D493–96.
- Kelly, Melissa A., Colleen Caleshu, Ana Morales, Jillian Buchan, Zena Wolf, Steven M. Harrison, Stuart Cook, et al. 2018. "Adaptation and Validation of the ACMG/AMP Variant Classification Framework for MYH7-Associated Inherited Cardiomyopathies: Recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel." *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 20 (3): 351–59.
- Kergourlay, Virginie, Ghadi Raï, Gaëlle Blandin, David Salgado, Christophe Bérout, Nicolas Lévy, Martin Krahn, and Marc Bartoli. 2014. "Identification of Splicing Defects Caused by Mutations in the Dysferlin Gene." *Human Mutation* 35 (12): 1532–41.
- Krahn, Martin, Valérie Biancalana, Mathieu Cerino, Aurélien Perrin, Laurence Michel-Calemard, Juliette Nectoux, France Leturcq, et al. 2019. "A National French Consensus on Gene Lists for the Diagnosis of Myopathies Using next-Generation Sequencing." *European Journal of Human Genetics: EJHG* 27 (3): 349–52.
- Lee, Kristy, Kate Krempely, Maegan E. Roberts, Michael J. Anderson, Fatima Carneiro, Elizabeth Chao, Katherine Dixon, et al. 2018. "Specifications of the ACMG/AMP Variant Curation Guidelines for the Analysis of Germline CDH1 Sequence Variants." *Human Mutation* 39 (11): 1553–68.
- Lu, Zhi-Xiang, Peng Jiang, and Yi Xing. 2012. "Genetic Variation of Pre-mRNA Alternative Splicing in Human Populations." *Wiley Interdisciplinary Reviews. RNA* 3 (4): 581–92.
- Mort, Matthew, Timothy Sterne-Weiler, Biao Li, Edward V. Ball, David N. Cooper, Predrag Radivojac, Jeremy R. Sanford, and Sean D. Mooney. 2014. "MutPred Splice: Machine Learning-Based Prediction of Exonic Variants That Disrupt Splicing." *Genome Biology* 15 (1): R19.
- Nissim-Rafinia, Malka, and Batsheva Kerem. 2002. "Splicing Regulation as a Potential Genetic Modifier." *Trends in Genetics: TIG* 18 (3): 123–27.
- Oza, Andrea M., Marina T. DiStefano, Sarah E. Hemphill, Brandon J. Cushman, Andrew R. Grant, Rebecca K. Siegert, Jun Shen, et al. 2018. "Expert Specification of the ACMG/AMP Variant Interpretation Guidelines for Genetic Hearing Loss." *Human Mutation* 39 (11): 1593–1613.
- Park, Eddie, Zhicheng Pan, Zijun Zhang, Lan Lin, and Yi Xing. 2018. "The Expanding Landscape of Alternative Splicing Variation in Human Populations." *American Journal of Human Genetics* 102 (1): 11–26.
- Puppo, Francesca, Eugénie Dionnet, Marie-Cécile Gaillard, Pascaline Gaildrat, Christel Castro, Catherine Vovan, Karine Bertaux, et al. 2015. "Identification of Variants in the 4q35 Gene FAT1 in Patients with a Facioscapulohumeral Dystrophy-like Phenotype." *Human Mutation* 36 (4): 443–53.
- Richard, I., C. Roudaut, A. Saenz, R. Pogue, J. E. Grimbergen, L. V. Anderson, C. Beley, et al. 1999. "Calpainopathy—a Survey of Mutations and Polymorphisms." *American Journal of Human Genetics* 64 (6): 1524–40.
- Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. 2015. "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 17 (5): 405–24.
- Rowlands, Charlie F., Diana Baralle, and Jamie M. Ellingford. 2019. "Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing." *Cells* 8 (12). <https://doi.org/10.3390/cells8121513>.
- Savisaar, Rosina, and Laurence D. Hurst. 2017. "Estimating the Prevalence of Functional Exonic Splice Regulatory Information." *Human Genetics* 136 (9): 1059–78.
- Stephenson, James D., Roman A. Laskowski, Andrew Nightingale, Matthew E. Hurles, and Janet M. Thornton. 2019. "VarMap: A Web Tool for Mapping Genomic Coordinates to Protein Sequence and Structure and Retrieving Protein Structural Annotations." *Bioinformatics* 35 (22): 4854–56.
- Théry, Jean Christophe, Sophie Krieger, Pascaline Gaildrat, Françoise Révillion, Marie-Pierre Buisine, Audrey Killian, Christiane Duponchel, et al. 2011. "Contribution of Bioinformatics Predictions and Functional Splicing Assays to the Interpretation of Unclassified Variants of the BRCA Genes." *European Journal of Human Genetics: EJHG* 19 (10): 1052–58.

- Tournier, Isabelle, Myriam Vezain, Alexandra Martins, Françoise Charbonnier, Stéphanie Baert-Desurmont, Sylviane Olschwang, Qing Wang, et al. 2008. "A Large Fraction of Unclassified Variants of the Mismatch Repair Genes MLH1 and MSH2 Is Associated with Splicing Defects." *Human Mutation* 29 (12): 1412–24.
- Turunen, Janne J., Elina H. Niemelä, Bhupendra Verma, and Mikko J. Frilander. 2013. "The Significant Other: Splicing by the Minor Spliceosome." *Wiley Interdisciplinary Reviews. RNA* 4 (1): 61–76.
- Woolfe, Adam, James C. Mullikin, and Laura Elnitski. 2010. "Genomic Features Defining Exonic Variants That Modulate Splicing." *Genome Biology* 11 (2): R20.
- Yeo, Gene, and Christopher B. Burge. 2004. "Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 11 (2-3): 377–94.
- Zhou, Xin, Michael N. Edmonson, Mark R. Wilkinson, Aman Patel, Gang Wu, Yu Liu, Yongjin Li, et al. 2016. "Exploring Genomic Alteration in Pediatric Cancer Using ProteinPaint." *Nature Genetics* 48 (1): 4–6.

Introduction

Le séquençage à haut débit et son implication dans le diagnostic des maladies neuromusculaires rares

L'arrivée récente du séquençage à haut débit dans la pratique clinique a révolutionné le diagnostic génétique et a lancé l'ère de la médecine génomique. Cette technologie, également appelée séquençage massivement parallèle ou séquençage de prochaine génération (Next Generation Sequencing, NGS), permet une identification objective et simultanée de la variation génétique dans des nombreux gènes chez un patient donné. Développée au début des années 2000, la technologie de séquençage à haut débit a été utilisée pour la première fois dans le cadre diagnostique en 2009 dans une étude preuve de principe pour identifier les variants pathogènes chez quatre patients non-apparentés atteints du syndrome de Freeman–Sheldon, une maladie rare autosomique dominante (Ng et al., 2009).

Le principe de la technologie de séquençage à haut débit repose sur l'incorporation successive de nucléotides complémentaires à la séquence d'intérêt, dans un processus comportant des étapes d'amplification de millions de molécules en parallèle. A chaque ajout d'un nucléotide, il y a une émission d'un signal, qui est ensuite converti en information de séquence. Actuellement il existe deux techniques majeures de séquençage à haut débit : la première, le séquençage avec incorporation de nucléotides modifiés fluorescents d'Illumina, la seconde, le séquençage par mesure de variation de pH lors de l'incorporation de nucléotides de Thermo Fisher (Buermans and den Dunnen, 2014). Les grandes étapes du séquençage à haut débit sont communes à ces deux technologies. La Figure 1 décrit les différentes étapes de ce processus (adapté de (Gorokhova et al., 2015a)). La première étape commence par l'extraction d'ADN (Acide Désoxyribonucléique) à partir d'un échantillon de sang, ou tout autre échantillon d'intérêt, qui sera ensuite fragmenté. Puis vient l'étape d'enrichissement des régions d'intérêt si l'étude porte sur un panel de gènes ou sur l'exome. Il existe différentes méthodes pour enrichir ces régions telles que la capture de séquences avec des sondes spécifiques ou la technique d'amplification. A l'inverse, le séquençage entier du génome amplifie et détermine la séquence de tout l'ADN y compris les parties non codantes telles que les introns et les régions intergéniques. Après cette première étape d'enrichissement, vient l'étape de séquençage. Pendant cette deuxième étape, l'appareil de séquençage génère et détecte l'émission de signaux successifs correspondants aux nucléotides dans chaque fragment d'ADN. A la fin de cette étape, le séquenceur produit un volumineux fichier (type FASTQ) contenant les séquences des millions des fragments d'ADN, appelés « reads » ou « lectures de séquence ».

Les étapes suivantes du séquençage à haut débit sont réalisées *in silico* par un enchaînement de traitements par différents outils bio-informatiques, souvent appelé un « pipeline bio-informatique ». Chacune de ces étapes contient des points de vérification et de contrôle qualité du processus avec un système de scores de qualités attribués. Les reads obtenus lors l'étape de séquençage sont d'abord alignés sur le génome de référence (fichier type BAM). Les différences entre la séquence de référence et la séquence du patient, aussi appelées « variants », sont ensuite identifiées par l'étape de « variant calling » (fichier type VCF). Les variants identifiés sont ensuite annotés avec des informations comme le nom du gène, le type de variant, sa localisation (exonique / intronique), sa fréquence dans les bases de données,

ses prédictions de pathogénicité, etc. (Figure1). La dernière étape, l'interprétation de la signification clinique des variants, est décrite dans la section suivante.

La technologie de séquençage à haut débit est aujourd'hui utilisée selon plusieurs approches en diagnostic génétique. La première approche est une stratégie intitulée « panel de gènes » ou « listes de gènes ». C'est la plus utilisée actuellement en routine diagnostique. Elle consiste à analyser simultanément les séquences d'une dizaine voire d'une centaine de gènes connus pour être impliqués dans une pathologie donnée en ciblant l'analyse sur les régions exoniques et les bornes introniques flanquantes (Gorokhova et al., 2015a; Krahn et al., 2016). La deuxième approche est celle du séquençage de l'exome entier. Elle permet d'analyser en même temps les régions codantes et bornes introniques flanquantes de tous les gènes du génome, ce qui équivaut à environ 1% du génome (~ 20 000 gènes). Cette approche se développe de plus en plus (the NDD Exome Scoping Review Work Group et al., 2019). La troisième approche est celle du séquençage du génome entier, comprenant les séquences codantes et non codantes. La quantité d'informations générée est 100 fois plus importante que celle obtenue pour l'exome entier, soit environ 3 millions de variants par échantillon.

Les approches diagnostiques par séquençage à haut débit sont devenues incontournables dans l'exploration des pathologies présentant une hétérogénéité génétique grâce à la possibilité d'explorer en même temps l'ensemble des gènes impliqués dans une pathologie, augmentant ainsi le rendement diagnostique. Ceci permet ainsi de réduire l'errance diagnostic que l'on trouve dans l'approche classique d'exploration génétique « gène par gène » (Gorokhova et al., 2015a; Krahn and Collège national des enseignants et praticiens de génétique, 2016; Krahn et al., 2016). Ceci concerne notamment le domaine des maladies neuromusculaires où il existe une très grande hétérogénéité génétique et clinique. En effet, cette puissante approche a rapidement gagné sa place parmi les principales méthodes utilisées dans le diagnostic des maladies musculaires (Gorokhova et al., 2015a).

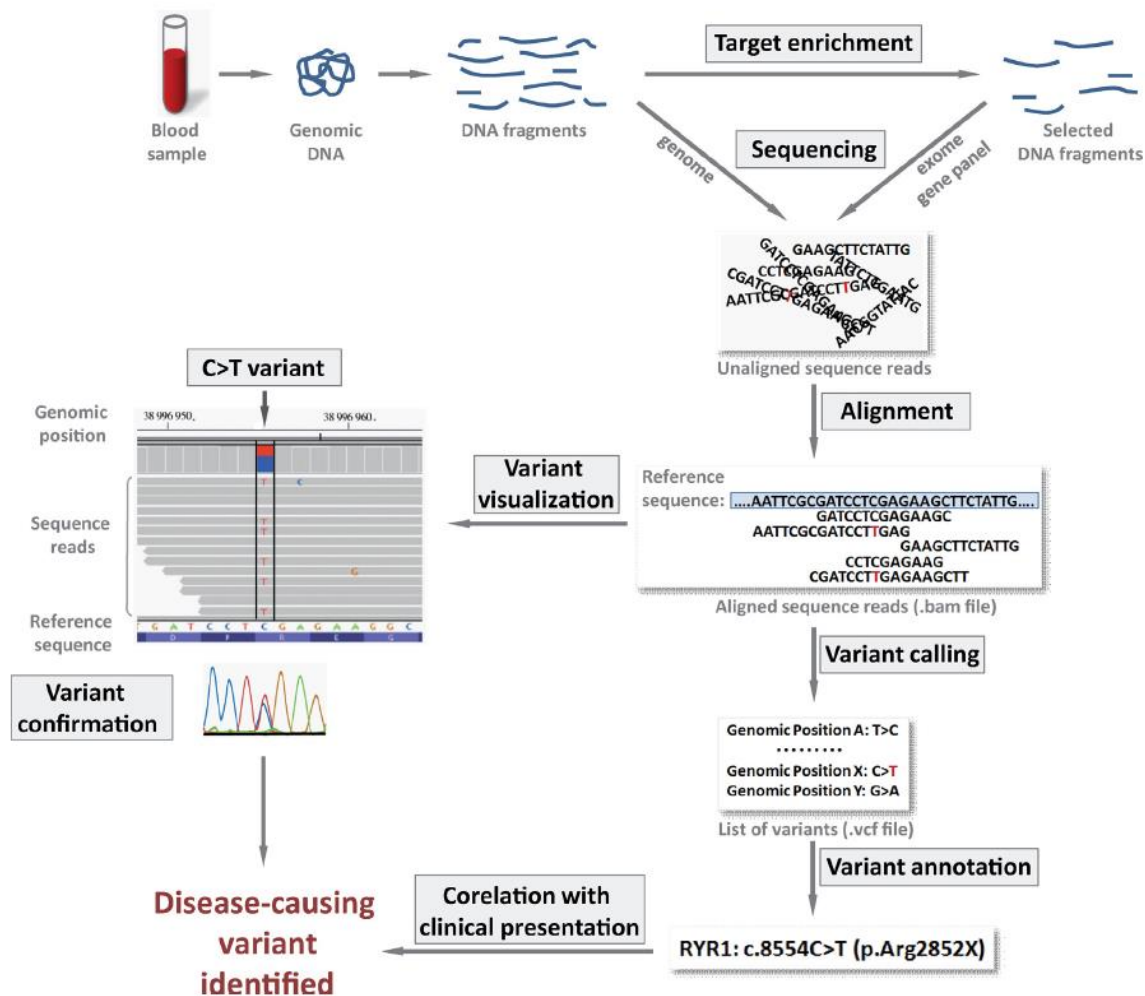


Figure 1 : Les étapes générales du séquençage à haut débit.

- 1) L'ADN génomique est extrait des globules blancs du patient et est fragmenté à l'aide de différentes techniques.
- 2) Tous les fragments génomiques sont séquencés lors du séquençage du génome entier, tandis que seuls les fragments sélectionnés sont séquencés dans des approches ciblées, telles que le séquençage de l'exome ou le séquençage de panel de gènes. Cette étape produit un grand nombre de séquences courtes appelées « reads ».
- 3) Les reads sont ensuite ordonnés en les alignant sur une séquence génomique de référence. Les différences entre la séquence génomique du patient et la séquence de référence sont identifiées à l'aide de programmes « d'appel de variants » (« variant calling »), générant une liste de positions et de modifications de nucléotides, ou de variants.
- 4) L'étape d'annotation correspond à la compilation d'informations disponibles concernant chacun des variants identifiés (type de variant, fréquence du variant dans différentes bases de données, analyse prédictive bioinformatique de pathogénicité, etc.).
- 5) Etape de corrélation entre les données mutationnelles obtenues et le phénotype clinique du patient, afin de déterminer l'implication du variant(s) dans la pathologie présentée par le patient.

Gorokhova S, Biancalana V, Lévy N, Laporte J, Bartoli M, Krahn M. Clinical massively parallel sequencing for the diagnosis of myopathies. Rev Neurol (Paris). juill 2015;171(6-7):558-71.

Interprétation des données issues d'un test diagnostique par Séquençage à Haut Débit

Critères de qualité indispensables pour l'interprétation des résultats d'un séquençage : couverture des séquences cibles et profondeur de lecture du séquençage

Depuis l'arrivée des technologies de séquençage à haut débit, des efforts importants ont été déployés pour améliorer la qualité des données obtenues afin de satisfaire aux exigences du diagnostic moléculaire. La qualité du séquençage est définie par sa profondeur et sa couverture (Figure 2). En diagnostic, le séquençage d'une région nécessite le plus souvent une couverture de 100% à 20x. La technologie actuelle ne permet pas toujours le séquençage parfait de toutes les régions d'intérêt, pouvant ainsi générer des résultats faux négatifs si le variant pathogène est présent mais non détecté. En effet, de nombreuses régions du génome humain restent difficiles à analyser à l'aide des approches standard de séquençage. Ces régions "sombres" du génome contiennent un certain nombre de gènes responsables de pathologies (Ebbert et al., 2019). Par exemple, plusieurs gènes responsables de maladies neuromusculaires, tels que la *NEBULINE (NEB)* et le *SELENON (SEPN1)*, chevauchent ces régions difficiles à séquencer. Les variants pathogènes dans les régions de gènes dont la couverture de séquence est sous-optimale peuvent être négligées. Ainsi, lorsque le séquençage à haut débit est effectué dans un cadre diagnostique, l'identification des régions mal couvertes est essentielle pour l'interprétation d'un résultat. Si la couverture n'est pas suffisante, même pour une petite région d'un gène candidat fortement suspecté, un test de re-séquençage supplémentaire est nécessaire pour s'assurer qu'un variant pathogène n'a pas été ignoré.

Malgré l'importance des données de couverture pour l'interprétation du résultat par séquençage à haut débit, l'information détaillée sur la couverture des exons des gènes étudiés n'est souvent pas disponible. **Pour faciliter l'interprétation des données de couverture par les biologistes et les prescripteurs de tests, nous avons conçu un nouvel outil de visualisation, CovReport. Cet outil génère un résumé concis des données de couverture permettant une visualisation rapide de la couverture de séquençage et donc une identification facile des régions mal séquencées.**

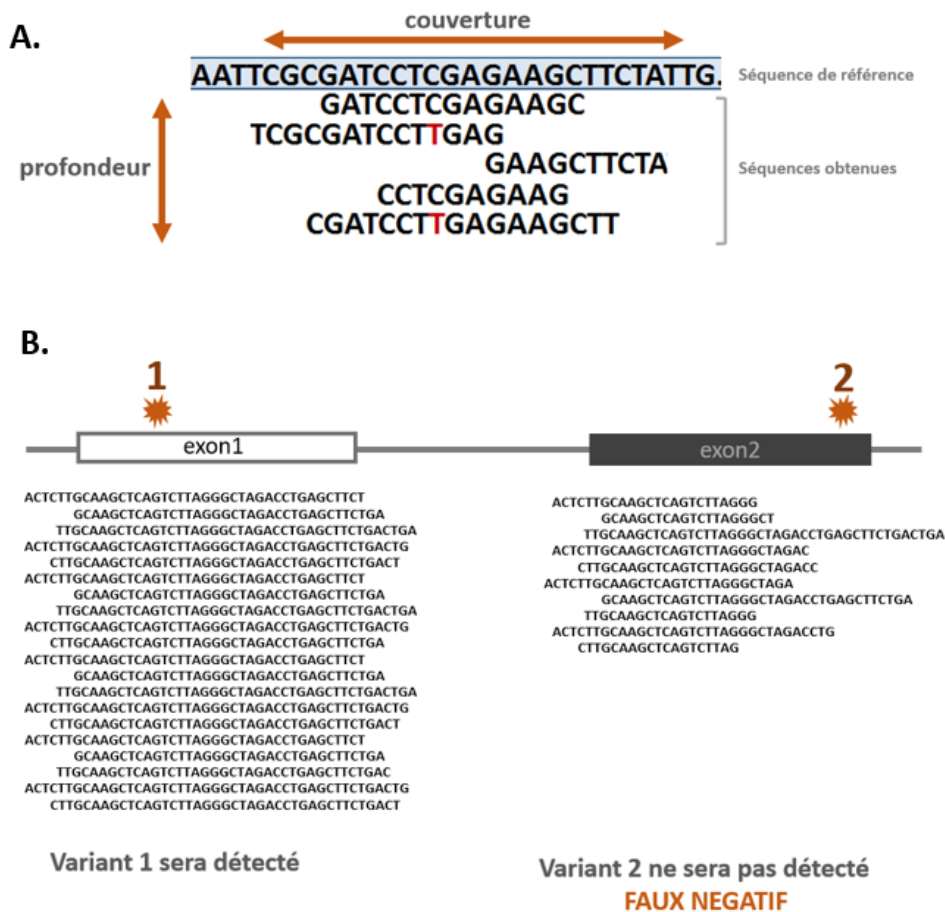


Figure 2 : Les concepts de la profondeur et la couverture d'un test par séquençage à haut débit.

A) La qualité du séquençage est définie par sa profondeur et sa couverture. La profondeur correspond au nombre de reads alignés sur une région ou combien de fois une position a été « lue » par le séquençage à haut débit. La couverture correspond au pourcentage de la région cible séquencée avec une profondeur donnée. Par exemple, une région peut être couverte à 100% avec une profondeur de 1x (chaque nucléotide a été lu au moins une fois) mais avoir une couverture de 50% à 20x (qu'une moitié des positions ont été lues au moins 20 fois). B) Un exemple de résultats de séquençage pour deux exons d'un gène donné. L'exon 1 a une très bonne couverture par les reads obtenus lors du séquençage. Le variant 1 sera détecté sans difficulté par des algorithmes bio-informatiques. La couverture de l'exon 2 n'est pas suffisante. Le variant 2, localisé dans la partie très mal couverte n'a été « lu » par le séquençage que deux fois. Ce variant ne sera pas détecté par les algorithmes bio-informatiques, conduisant à un résultat faux négatif.

L'association gène-maladie pour valider les gènes inclus dans les panels diagnostiques

Selon la base de données OMIM (Online Mendelian Inheritance in Man), un catalogue des gènes humains et des maladies génétiques, il y a actuellement (mars 2020) 4,241 gènes impliqués dans des pathologies génétiques. Toutefois, le niveau de preuve pour ces associations gène-maladie varie considérablement. Certaines des relations gène-maladie ont été évoquées il y a longtemps mais n'ont toujours pas été confirmées ou sont finalement prouvées fausses. Pour mieux définir les gènes et les variants cliniquement pertinents, le Clinical Genome Resource (ClinGen), financé par le NIH, a créé une ressource en libre accès basée sur une évaluation standardisée et transparente des preuves (Rehm et al., 2015). Ce

groupe a développé une méthode qui définit qualitativement la validité clinique des gènes et des maladies en utilisant un système de classification selon le niveau des preuves montrant l'association gène-maladie (Strande et al., 2017a). Cette méthode fournit une approche semi-quantitative standardisée pour évaluer les preuves disponibles et arriver à une classification des associations gène-maladie en six catégories : « Definitive », « Strong », « Moderate », « Limited », « No Reported Evidence » ou « Conflicting Evidence ».

Les panels de gènes développés par des laboratoires de diagnostic moléculaire ont pour objectif à évaluer les causes génétiques potentielles d'une pathologie génétique en réduisant le coût et le temps du test diagnostique. Ces panels sont particulièrement utiles pour le diagnostic de pathologies présentant une hétérogénéité génétique et clinique, tel que les maladies neuromusculaires. En effet, ils permettent de maximiser la couverture des régions cibles et d'éviter les découvertes secondaires, augmentant l'efficacité et la rapidité d'une analyse génétique. Pour un grand nombre de pathologies, les panels de gènes sont préférés à l'exome ou au séquençage du génome et souvent représentent une analyse en première ligne. Jusqu'à très récemment, en raison de l'absence de recommandations claires, les gènes inclus dans des panels ciblant le même groupe de pathologies pouvaient varier d'un laboratoire à un autre avec une possibilité de rendu de résultats très différents pour le même patient. De plus, les panels pouvaient contenir des gènes pour lesquels le lien de causalité avec la pathologie n'était pas établi avec certitude. En effet, depuis la publication récente de la méthode pour déterminer l'implication d'un gène dans une pathologie (Strande et al., 2017a), un certain nombre de gènes initialement utilisés en diagnostic ont été retirés des panels de gènes (DiStefano et al., 2019; Hosseini et al., 2018; Ingles et al., 2019; Seifert et al., 2019). Le séquençage de gènes dont le lien gène-pathologie est limité doit être évité en diagnostic, car l'interprétation biologique des variants identifiés dans ces gènes aboutira en classification « Variant de Signification Incertaine » (VUS), induisant un stress et une anxiété inutiles chez les patients et leurs familles. Les recommandations récentes de l'ACMG (American College of Medical Genetics and Genomics) visent à clarifier quels gènes peuvent être inclus dans les panels diagnostiques (Bean et al., 2019). Selon ces recommandations, les gènes sont définis comme GAD (Gene associated with Mendelian disorder), pour lesquels il existe des preuves suffisantes qu'ils sont responsables de la pathologie, ou GUS (Gene of uncertain significance), les gènes pour lesquels il n'y a pas encore assez de preuves pour établir un lien gène-pathologie (Figure 3). Si l'indication est de confirmer un diagnostic clinique, le panel de gènes utilisé ne doit contenir que des gènes pour lesquels le lien gène-maladie est clairement établi (GAD, Genes associated with Mendelian disorder). Si l'indication est le diagnostic des cas cliniquement complexes, les GUS (Genes of uncertain significance) peuvent être inclus. Dans ce cas, un consentement adapté à une analyse de type exome/génome est requis. De plus, dans le compte rendu, il est nécessaire de clairement séparer les variants identifiés dans les GUS des variants identifiés dans les GAD. Enfin, ces recommandations précisent que la classification maximale attribuée à un variant identifié dans un gène sans lien gène-pathologie établie (GUS) ne peut pas être plus qu'une classification VSI (variant d'une signification incertaine).

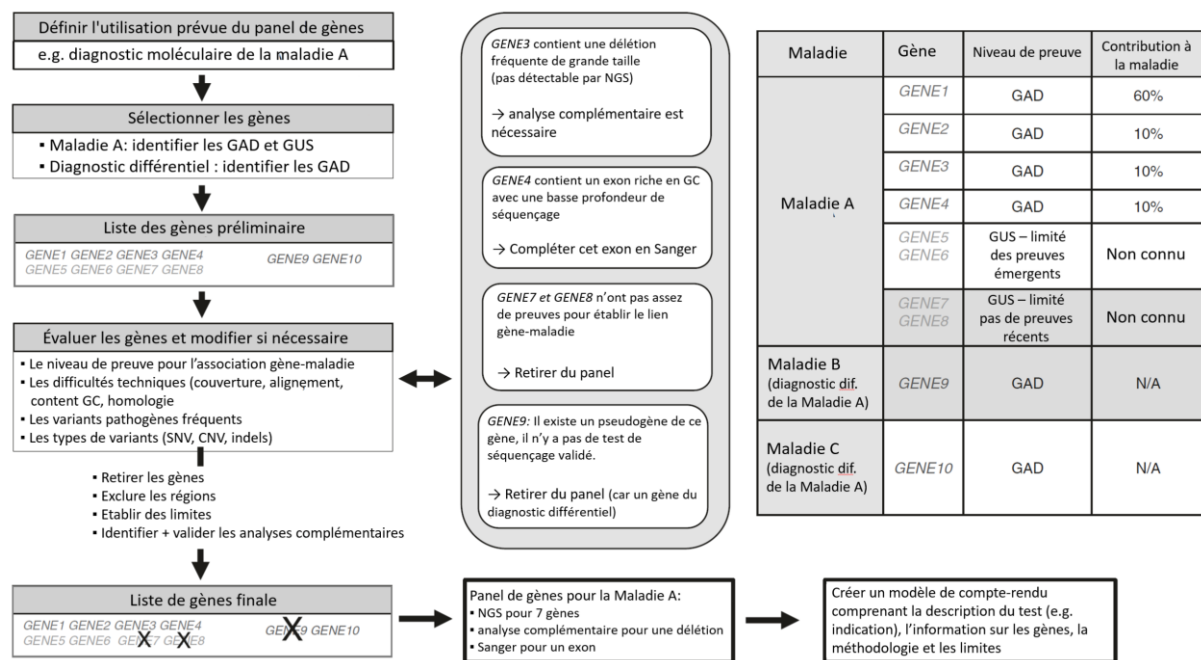


Figure 3. Procédure pour la conception, l'évaluation et la mise en œuvre d'un panel de gènes en diagnostic. (Adapté de Bean et al. 2019)

GAD (Gene associated with Mendelian disorder), gènes pour lesquels il existe des preuves suffisantes qu'ils sont responsables de la pathologie. GUS (Gene of uncertain significance), les gènes pour lesquels il n'y a pas encore assez de preuves pour établir un lien gène-pathologie

Devant l'importance de cibler les efforts diagnostiques sur les gènes avec un fort lien gène-pathologie, plusieurs initiatives et groupes de travaux ont été formés. Un nombre croissant de groupes de travail appelés Gene Curation Expert Panels (GCEPs), coordonnés par ClinGen rassemble des experts internationaux pour avancer sur l'évaluation des associations gène-pathologies (<https://clinicalgenome.org/working-groups/clinical-domain/>). Trois groupes de travail dans le domaine des maladies neuromusculaires viennent d'être formés :

- Charcot-Marie-Tooth Disease Gene Curation Expert Panel
- Congenital Myopathies Gene Curation Expert Panel
- Limb Girdle Muscular Dystrophy Gene Curation Expert Panel (dont nous faisons partie depuis Février 2020).

Sur le même principe, le Royaume-Uni a créé PanelApp, une base de données permettant de classer des corrélations « gène-pathologie » en utilisant comme indicateur un code couleur. Les gènes « vert » sont des gènes avec des preuves suffisantes de l'association gène-maladie. Ces gènes peuvent être utilisés en diagnostic avec confiance. Pour les gènes « orange », il existe des preuves modérées de l'association gène-maladie. Ces gènes ne devraient pas encore être utilisés pour l'interprétation des résultats de séquençage. Pour les gènes « rouges », il n'existe pas des preuves suffisantes de cette association gène-maladie. Ces gènes ne devraient pas être utilisés pour l'interprétation. (Figure 4, (<https://panelapp.genomicsengland.co.uk/>)). Cette base de données permet également à la communauté scientifique et médicale de partager leurs informations et d'échanger.

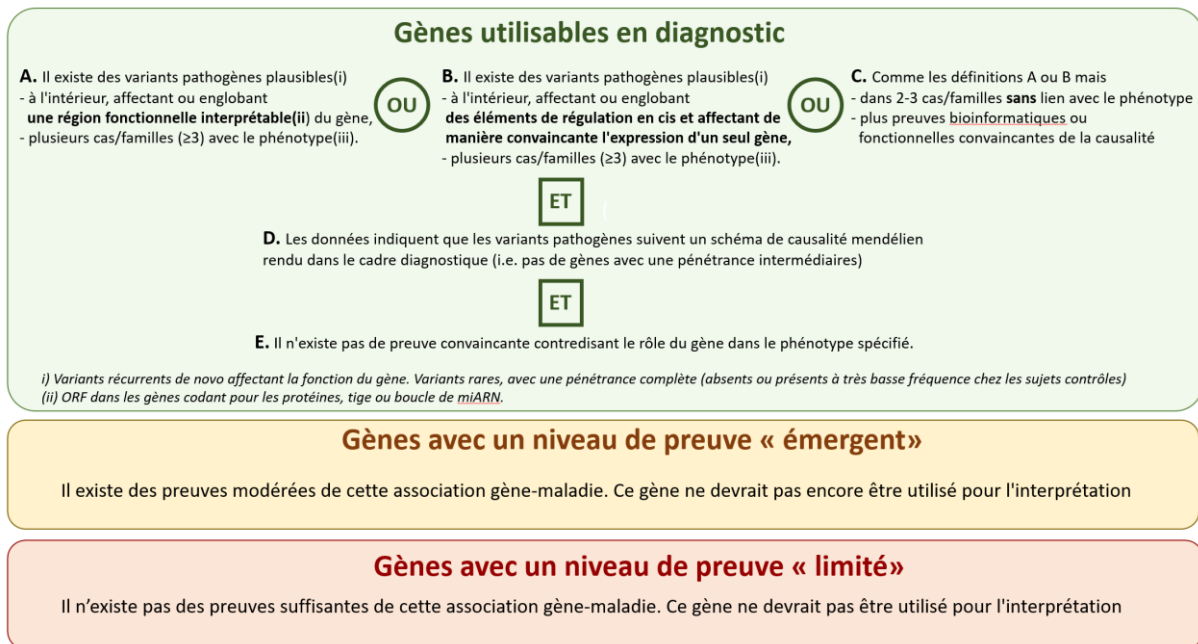


Figure 4. Le système utilisé par PanelApp pour classer les gènes

Les gènes « vert » sont des gènes avec des preuves suffisantes de l'association gène-maladie. Ces gènes peuvent être utilisés en diagnostic avec confiance. Pour les gènes « orange », il existe des preuves modérées de l'association gène-maladie. Ces gènes ne devraient pas encore être utilisés pour l'interprétation des résultats de séquençage. Pour les gènes « rouges », Il n'existe pas des preuves suffisantes de cette association gène-maladie. Ces gènes ne devraient pas être utilisés pour l'interprétation.

(<https://panelapp.genomicsengland.co.uk/>)

En France, à la suite du travail de différents groupes d'experts, des listes de gènes impliqués dans certaines pathologies ont été proposées dans le but d'harmoniser les analyses par séquençage à haut débit au niveau national. Ces travaux ont été menés notamment par des groupes d'experts au sein des Filières Nationales de Santé Maladies Rares (<https://solidarites-sante.gouv.fr/soins-et-maladies/prises-en-charge-specialisees/maladies-rares/article/les-maladies-rares>). C'est le cas pour la thématique des myopathies et la filière FILNEMUS (Filière Nationale des Maladies Rares Neuromusculaires; www.filnemus.fr), qui a publié les listes de gènes définies en fonction des portes d'entrée diagnostiques, avec l'information sur l'association gène-pathologie pour chaque gène. Plus de détails sur la structuration du diagnostic des maladies neuromusculaires en France sont décrits dans la section suivante.

Recommandations pour l'interprétation des variants identifiés par un test diagnostique par séquençage à haut débit

La mise en place du séquençage à haut débit à visée diagnostique dans les laboratoires de génétique a augmenté drastiquement le nombre de variants génétiques identifiés pour chaque analyse effectuée. L'interprétation de la signification clinique de ces variants est devenue un véritable défi pour les laboratoires, surtout pour les variants non décrits précédemment ou pour les variants dans des gènes dont les laboratoires n'avaient pas encore d'expertise spécifique. A titre d'exemple, une analyse d'un panel de 200 gènes de myopathies génère environ 300 variants à interpréter par échantillon de patient, et une analyse d'exome entier et de génome entier respectivement 20 000 et environ 3 millions variants. Pour homogénéiser et standardiser l'interprétation clinique des variants entre les laboratoires diagnostiques, plusieurs sociétés savantes, tel que l'ACMG-AMP (American College of Medical

Genetics & the Association for Molecular Pathology) ou EuroGentest/ESHG (European Society of Human Genetics), ont émis des recommandations pour guider les laboratoires dans l'interprétation (Amendola et al., 2016; Matthijs et al., 2016; Richards et al., 2015). Même si les principes et les mises en garde abordées dans ces documents étaient assez similaires, le système de calcul de pathogénicité des variants proposé par les recommandations d'ACMG-AMP est devenu aujourd'hui le standard mondial pour l'interprétation des analyses par séquençage à haut débit. Selon les propositions par Richards et al., l'interprétation de la signification clinique de chaque variant repose sur un faisceau d'arguments, tel que sa fréquence allélique dans la population générale, présence chez les sujets atteints, données de ségrégation, prédiction de la pathogénicité par des outils bio-informatique ou les données fonctionnelles. Ces arguments ont un poids plus ou moins important dans l'interprétation du variant identifié. Pour les arguments en faveur de la pathogénicité du variant ils peuvent être: « Très fort » (PVS, Pathogenic Very Strong), « Fort » (PS, Pathogenic Strong), « Moyen » (PM, Pathogenic Moderate), « Faible » (PP, Pathogenic Poor). Pour les arguments en faveur du caractère bénin du variant ils sont : « Suffisant » (BA, Benign stand Alone), « Fort » (BS, Benign Strong), « Faible » (BP, Benign Poor). L'interprétation des résultats est sous la responsabilité exclusive du biologiste et consiste à combiner ces arguments pondérés afin d'assigner une des 5 classes suivantes au variant étudié :

- Classe 1 : Variant bénin
- Classe 2 : Variant probablement bénin
- Classe 3 : Variant de signification incertaine
- Classe 4 : Variant probablement pathogène
- Classe 5 : Variant pathogène

Ce système de classification des variants a été incorporé dans les recommandations nationales françaises émises par le Groupe de Travail du Réseau NGS-Diag, en collaboration avec l'ANPGM (Association Nationale des Praticiens de Génétique Moléculaire), l'ACLF (Association des Cytogénéticiens de Langue Française), AchroPuce (Réseau d'Analyse Chromosomique sur Puce à ADN) et le GGC (Groupe Génétique et Cancer), ainsi que plusieurs Filières des Maladies Rares. La première version de ce document a été publiée en juin 2018 avec un objectif de limiter la variabilité inter- et intra-laboratoire dans l'interprétation des variants en génétique moléculaire et d'aboutir à un consensus national. Depuis, ces recommandations ont été adoptées par la grande majorité des laboratoires diagnostiques en France (<http://ffgh.net/index.php/presentation/les-reseaux-partenaires/reseau-ngs-diagnostic/182-reseaux/ngs/354-recommandations-professionnelles-du-reseau-ngs-diag>).

Cependant, les recommandations initiales laissaient une incertitude sur l'attribution de certains critères de classification, entraînant des disparités entre les résultats des différents laboratoires (Amendola et al., 2016). Afin d'homogénéiser l'utilisation des items de la liste d'arguments, il apparaît nécessaire d'apporter certaines précisions comme la valeur seuil de fréquences alléliques ou l'utilisation de scores combinés de prédiction de pathogénicité de différents algorithmes. De plus, l'application des recommandations générales peut être très différente selon le gène étudié. C'est pour cette raison que des initiatives comme ClinGen

(Rehm et al., 2015), financé par le NIH (National Institute of Health) aux Etats-Unis, ou d'autres groupes de travail d'experts, ont publié des mises à jour de la classification ACMG-AMP initiale (Amendola et al., 2016; Richards et al., 2015). Il s'agit de précisions générales sur les critères de classification (Abou Tayoun et al., 2018; Biesecker et al., 2018; Brnich et al., 2019; Ghosh et al., 2018; SVI_Recommendation_for_PM3_Criterion, 2019; SVI_Recommendation_for_PS2_and_PM6_criteria, 2018), et des recommandations gène-spécifiques élaborées par des Groupes d'experts du réseau ClinGen ou autres experts dans le domaine (MYH7, Rasopathies, PAH, Surdités, CDH1, PTEN, MEN1, RUNX1, GAA, TP53)(ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019; ClinGen_TP53_Expert_Panel, 2019; Gelb et al., 2018; Kelly et al., 2018; Lee et al., 2018a; Luo et al., 2019; Mester et al., 2018; Oza et al., 2018; Romanet et al., 2019; Zastrow et al., 2018). Comme les résultats de ces groupes de travail ont été publiés après la parution de la première version des recommandations du réseau NGS-Diag, il est important d'exposer et potentiellement d'incorporer ces mises à jour dans la nouvelle version de ce document, actuellement en cours d'élaboration.

Importance de tests fonctionnels pour l'interprétation des variants

Le système de classification des variants identifiés lors d'une analyse par séquençage à haut débit permet d'attribuer pour chacun d'eux une classe de l'impact clinique en prenant en compte systématiquement tous les arguments disponibles. Cependant, les variants non-identifiés précédemment, ou localisés dans des gènes avec un lien gène-pathologie en cours de validation, seront souvent classés comme « Variant de Signification Incertaine » ou VSI. Un résultat génétique avec un VSI peut induire un stress et anxiété considérable pour le patient et sa famille. Il peut même parfois, à tort, entraîner des explorations et consultations de suivi non-utiles. C'est pour cette raison qu'il y a une nécessité majeure d'établir des tests fonctionnels pour déterminer l'impact pathogène ou non de ces variants sur la fonctionnalité de la protéine ou sur l'épissage d'ARN messenger. La meilleure façon de déterminer si un variant a un effet sur l'épissage, est d'explorer l'ARN provenant d'un tissu atteint du patient. Une apparition d'un transcrit anormal en lien avec le variant est un argument fort pour la pathogénicité de celui-ci. Cependant, très souvent le tissu du patient qui permettent de réaliser ces explorations n'est pas disponible, car d'accès trop invasif (biopsie musculaire) ou non-accessible (cerveau). Une méthode *in vitro* appelée l'étude par « minigène », a été établie pour contourner ce problème (Gaildrat et al., 2010). Dans cette approche, un segment génomique avec la séquence contenant le variant d'intérêt, ainsi que les séquences introniques adjacentes, est intégré dans un vecteur « minigène », constitué des éléments d'épissage artificiels. Suite une transfection transitoire dans une lignée cellulaire, les transcrits générés par les cellules transfectées avec la construction « minigène » contenant le variant, sont testés et comparés aux transcrits dans les cellules contrôle, pour détecter des anomalies d'épissage. Nous avons mis en place cette approche pour la majorité des exons de *CAPN3*, le gène responsable d'une forme majeure des dystrophies musculaires des ceintures. Pour les variants montrant une anomalie d'épissage par le test « minigène », nous avons pu attribuer un argument fort PS3 selon les recommandations ACMG-AMP actuelles, ce que nous a permis de reclasser la majorité de ces variants, démontrant ainsi l'importance de cette approche dans la démarche diagnostique.

Structuration du diagnostic génétique des maladies neuromusculaires par Séquençage à Haut Débit en France

Les maladies neuromusculaires sont des pathologies qui affectent le muscle strié (dystrophies musculaires), la jonction entre le nerf et le muscle (myasthénies), les cellules nerveuses motrices de la moelle épinière ou motoneurones (amyotrophies spinales, sclérose latérale amyotrophique) ou les racines et les nerfs des membres (neuropathies périphériques). Une majorité de ces pathologies est d'origine génétique. La sévérité de l'atteinte musculaire, le type de muscle impacté, l'association avec d'autres symptômes neurologiques ou systémiques sont très variables et dépendent souvent du gène impliqué. Le diagnostic moléculaire est indispensable pour un patient atteint d'une affection neuromusculaire car il permet d'orienter la prise en charge de la maladie, prédire la progression des symptômes et donner un conseil génétique. Un diagnostic précoce a l'avantage de réduire l'errance diagnostique et d'arrêter les examens à visée diagnostique, qui peuvent être invasifs comme par exemple les biopsies musculaires.

En France, la Filière Nationale des Maladies Rares Neuromusculaires (FILNEMUS) a été créée dans le cadre du second plan national maladies rares 2011-2016, afin d'améliorer au niveau national la visibilité des maladies rares neuromusculaires ainsi que de structure de prise en charge et de soins (Centres de Référence Maladies Rares-CRMR et Centres de Compétence Maladies Rares-CCMR). FILNEMUS se situe actuellement à la 2^e place en termes de suivi actif de patients, avec une cohorte de 44000 patients répartis sur 32 centres de référence et 39 centres de compétence en maladies neuromusculaires (Krahn et al., 2019). Avec la généralisation des analyses génétiques par séquençage à haut débit, la filière FILNEMUS a élaboré des recommandations spécifiques concernant les myopathies. Les groupes d'experts de FILNEMUS ont utilisé la procédure similaire à celle publiée par ClinGen en 2017 qui consiste à associer un gène à une pathologie (Krahn et al., 2019; Strande et al., 2017b). Les corrélations « gène-pathologie » ont été établies pour 199 gènes de myopathies, soit au total 300 corrélations « gène-pathologie » différentes car certains gènes présentent plusieurs expressions phénotypiques dans différentes catégories (Krahn et al., 2019).

En plus de ces travaux d'harmonisation, une nouvelle stratégie de diagnostic des myopathies par séquençage à haut débit a été mise en place au niveau national. Pour chaque suspicion de myopathie, l'orientation clinique et/ou histologique permet d'initier l'analyse génétique d'une ou de plusieurs de ces 13 listes de myopathies :

- Dystrophies des ceintures (LGMD) : 40 gènes
- Myopathies congénitales à némaline positive : 12 gènes
- Myopathies congénitales à « némaline négative » : 32 gènes
- Arthrogryposes fœtales et néonatales : 37 gènes
- Dystrophies musculaires congénitales hors alpha-dystroglycanopathies : 17 gènes
- Alpha-dystroglycanopathies : 18 gènes
- Myopathies distales et scopolopéronières : 24 gènes
- Myopathies rétractiles : 28 gènes
- Myopathies myofibrillaires et Myopathies à inclusions : 15 gènes
- Myasthénies congénitales : 32 gènes
- Paralysies périodiques : 3 gènes

- Myotonies non dystrophiques : 7 gènes
- Myopathies métaboliques : 35 gènes

Pour chaque catégorie de myopathie, il existe une liste principale de gènes et une liste exhaustive ou une liste de gènes unique exhaustive (disponibles sur le site de FILNEMUS: <http://www.filnemus.fr>). Dans un premier temps la liste principale est utilisée pour l'analyse par séquençage à haut débit, puisqu'elle est composée des gènes « majeurs » à analyser en priorité. Si les résultats de cette première analyse ne sont pas concluants (absence de variant pathogènes ou probablement pathogènes), la liste exhaustive de gènes sera alors étudiée. Puis, si cette nouvelle liste est toujours négative, l'ensemble des gènes de myopathies pourra être exploré. En l'absence de diagnostic après cette analyse, la réalisation d'un exome ou d'un génome voire même le séquençage de l'ARN messager pourront être envisagés (Krahn et al., 2019) (Figure 4).

Les patients atteints de maladies musculaires non diagnostiqués pourront aussi bientôt bénéficier d'un séquençage du génome dans le cadre du plan « France Médecine Génomique 2025 ». Initié en 2015, ce plan prévoit de placer la France parmi les pays les plus engagés dans la médecine personnalisée de précision et d'intégrer la médecine génomique dans le parcours de soins et la prise en charge des pathologies communes avec pour objectif un séquençage de 235 000 génomes par an. Deux pré-indications ont été proposées par FILNEMUS et acceptées par le PFMG2025, à savoir « Myopathies non étiquetées après panel thématique Filnemus négatif » et « Hypotonies néonatales périphériques suspectes de maladie neuromusculaire ». Cette mesure aidera grandement à diminuer l'errance diagnostique pour les patients atteints de maladies neuromusculaires en France.

Notre Service de Génétique Médicale de Marseille, et l'unité INSERM/AMU 1251/Marseille Medical Genetics (dirigés par le Pr. N. Lévy) sont impliqués depuis de nombreuses années dans la coordination et l'évolution des analyses moléculaires diagnostiques pour les maladies neuromusculaires, notamment par séquençage à haut débit, et les différents travaux réalisés dans le cadre de cette thèse de médecine s'intègrent dans ces démarches.

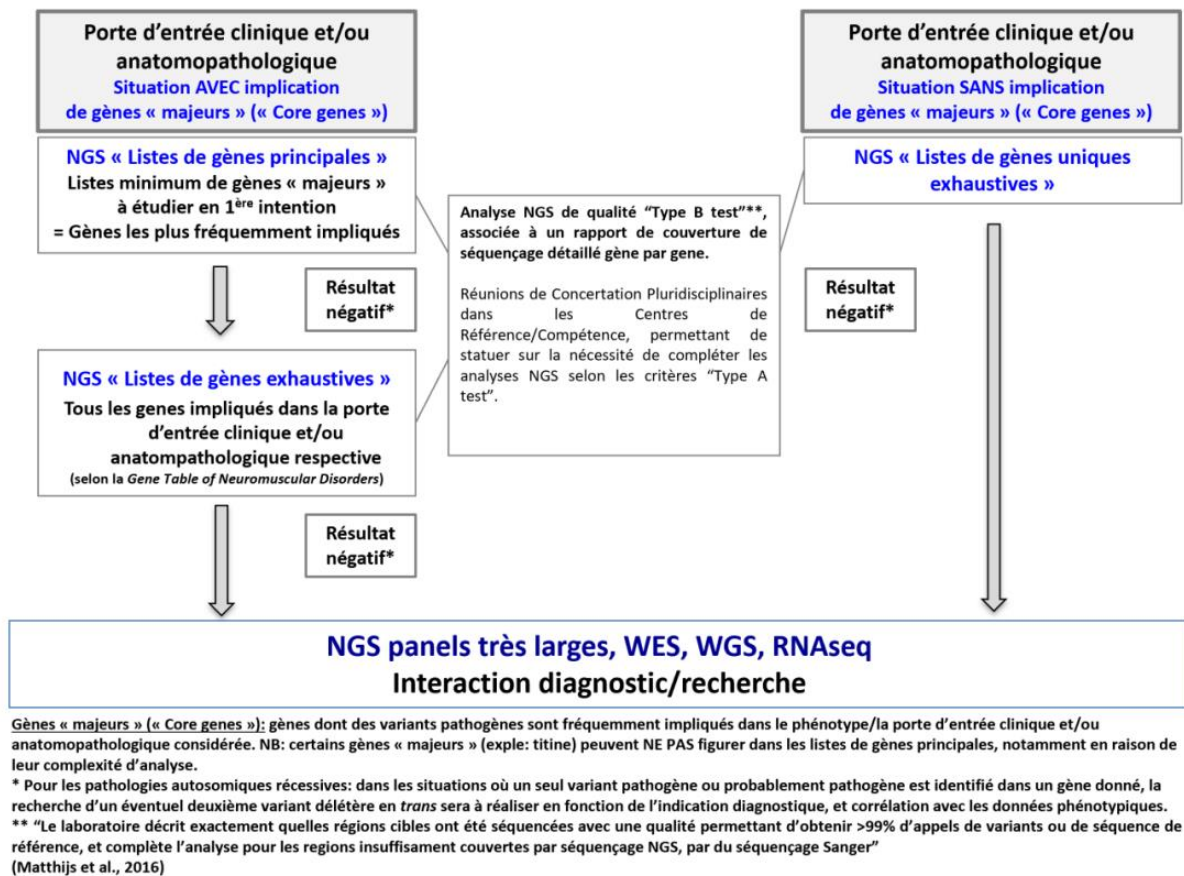


Figure 5. Stratégie consensuel de diagnostic des myopathies par séquençage à haut débit en France. (Krahn et al., 2019)

Matériel et méthode

Dans cette section, une synthèse des sections respectives des trois articles est présentée. L'intégralité du travail effectué dans le cadre de cette thèse de médecine est exposée dans les articles ci-joints.

Mise en place de CovReport – l'outil de visualisation de la couverture de séquence

Design et fonctionnement du logiciel

CovReport a été mis en œuvre sous la forme d'une application Java autonome. Il peut fonctionner sur toute plateforme où le logiciel Java Runtime (JRE) est installé (Windows, Mac OS et Linux). La version Java supportée est la version 8. Des détails techniques sur le logiciel, les instructions d'utilisation ainsi que le lien vers le téléchargement sont disponibles sur <http://www.jdotsoft.com/CovReport.php>. Le fichier d'entrée contenant les informations de couverture par exon est un fichier CSV délimité par des tabulations. Les colonnes suivantes sont attendues : RefSeqName, GeneSymbol, Exon, Size, Mean Depth, SD Depth, Coverage 1x, Coverage 5x, Coverage 10x, Coverage 20x, Coverage 30x. Un exemple de fichier est disponible sur http://www.jdotsoft.com/CovReport/44genes-stat_coverage_exon.txt. Le fichier d'entrée peut être facilement généré par le module de couverture de VarAFT tool, (<https://varaft.eu/>, (Desvignes et al., 2018)) qui utilise BEDTools (Quinlan, 2014; Quinlan and

Hall, 2010) pour calculer la couverture au niveau de l'exon. Toutefois, les résultats de tout autre pipeline de calcul de la couverture peuvent être convertis dans le format approprié et utilisés comme une entrée de CovReport. De plus, la version en ligne de commande de CovReport permet d'intégrer cet outil de visualisation comme une étape dans un pipeline d'analyse de données de séquençage à haut débit. L'interface utilisateur de l'application est en anglais, mais le fichier PDF généré est dans la langue locale actuelle. L'anglais et le français sont supportés dans la distribution actuelle, avec la possibilité d'ajouter d'autres langues par l'utilisateur. Le texte inclus dans le rapport PDF de sortie peut également être personnalisé, ce qui permet d'adopter CovReport dans n'importe quelle procédure de diagnostic.

Séquençage d'un exome pour tester la performance de CovReport

L'ADN de l'échantillon de référence NA12878 a été obtenu auprès du Coriell Institute for Medical Research Repository (Institut Coriell, Camden, NJ, USA). Le séquençage de l'exome entier (WES) a été effectué par la plateforme de génomique et de bioinformatique (GBiM, U 1251/Marseille), à l'aide du kit NimbleGen SeqCap EZ MedExome (taille de séquences cibles 47Mb) selon le protocole du fabricant (Roche Sequencing Solutions, Madison, USA). Le kit SeqCap EZ MedExome cible l'ensemble de l'exome humain avec une couverture accrue des exons des gènes médicalement pertinents dans les maladies mendéliennes. Après une étape d'enrichissement, les bibliothèques ont été séquencées sur la plateforme Illumina NextSeq 500 (Illumina, San Diego, CA, USA) à l'aide d'un protocole de séquençage "paired-end" avec des fragments de 150 bp. Les données brutes ont été mappées au génome humain (hg19) en utilisant BWA 0.7.5.

Mise en place de CovReport pour le diagnostic en routine dans le Laboratoire de Génétique Moléculaire à AP-HM.

CovReport est utilisé en routine dans le Laboratoire de Génétique Moléculaire de l'AP-HM, pour tous les tests diagnostiques par séquençage à haut débit depuis Juin 2018. Pour tous les séquençages de panels de gènes proposés par le laboratoire, le kit de capture SureSelect Custom (Agilent Technologies) ciblant initialement 852 gènes puis 1070 gènes dans sa version actuelle, est utilisé. Le séquençage a été effectué sur la plateforme Ion Proton™, puis depuis 2019 sur Ion GeneStudio S5 Prime™ (Thermo Fisher Scientific). L'analyse bio-informatique de résultats de séquençage est faite par l'Ion Torrent Suite, intégrée dans les plateformes Ion Proton™ et Ion GeneStudio S5 Prime™. L'analyse de la couverture est réalisée à l'aide du module d'analyse de couverture module du logiciel VarAFT (Desvignes et al., 2018). Pour chaque patient, le fichier de couverture issu de VarAFT est ensuite utilisé pour générer le rapport de couverture intitulé « *Annexe 1. Couverture des fichiers cibles* ». Ce rapport sert à valider le séquençage par le responsable technique du test, et est directement intégré dans le dossier du patient pour être joint au compte rendu du résultat. Au total plus de 1 500 rapports diagnostiques ont été générés à l'aide de CovReport depuis Juin 2018.

Mise à jour des recommandations françaises pour l'interprétation des variants

Un travail bibliographique extensif a été effectué pour rechercher les récentes mises à jour des recommandations pour l'interprétation des variants issus des tests par séquençage à haut débit. Les données de PubMed, ainsi que tous les documents disponibles sur le site de ClinGen (<https://www.clinicalgenome.org/>) ont été étudiés. Au total, dix publications décrivant des recommandations spécifiques à un gène ou un groupe de pathologies ont été identifiées :

MYH7 (Kelly et al., 2018), Rasopathie (Gelb et al., 2018), PTEN (Mester et al., 2018), Surdités (Oza et al., 2018), CDH1 (Lee et al., 2018a), MEN1 (Romanet et al., 2019), RUNX1 (Luo et al., 2019), TP53 (ClinGen_TP53_Expert_Panel, 2019), PAH (Zastrow et al., 2018), GAA (ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019). De plus, six nouvelles recommandations et précisions pour les arguments des critères ACMG-AMP ont été étudiées :

- PVS1 – Variant ayant un effet « perte de fonction » (Abou Tayoun et al., 2018)
- PS3/BS3 – arguments attribués selon les données fonctionnelles (Brnich et al., 2019)
- PM3 – argument attribué selon les données de ségrégation (SVI_Recommendation_for_PM3_Criterion, 2019)
- PS2/PM6 – arguments attribués si le variant étudié est *de novo* (SVI_Recommendation_for_PS2_and_PM6_criteria, 2018)
- PP5/BP6 – argument « présence dans les bases de données des variants pathogènes » (Biesecker et al., 2018)
- BA1 – argument attribué aux variants très fréquents dans la population générale, classant automatiquement ces variants en bénin (Ghosh et al., 2018)

Chaque argument a été examiné en détail dans toutes ces études. Au total, tous les 28 arguments des recommandations ACMG-AMP ont été analysés. Les mises à jour ou des précisions ont été proposées pour 11 arguments suivants : PVS1, PS4, PM2/BS1/BA1, PM3, PS2/PM6, PP1, PP5, BP6.

Sélection et analyse des variants faux-sens dans le gène *CAPN3* avec impact potentiel sur l'épissage

Pour explorer l'impact potentiel de variants faux-sens sur l'épissage, nous avons utilisé l'outil HSF (Human Splice Finder, (Desmet et al., 2009)) pour examiner les variants dans le gène *CAPN3* identifiés dans notre laboratoire. Nous avons ensuite sélectionné 24 variants répartis dans tout le gène pour des tests fonctionnels par l'approche « minigène ». L'étude de l'épissage par l'approche « minigène » a été développée pour 18 exons du gène comme décrit précédemment (Gaildrat et al., 2010; Kergourlay et al., 2014). Les arguments de pathogénicité pour les 24 variants sélectionnés ont été étudiés en détail pour classer ces variants selon les recommandations ACMG adaptées à *CAPN3* comme décrit dans l'article ci-joint. Ces classifications ont ensuite été recalculées en prenant en compte les résultats de l'étude fonctionnelle par « minigène ».

Pour estimer la prévalence d'autres variants faux-sens avec un effet potentiel sur l'épissage, nous avons constitué les deux sets de variants suivants : 381 de variants provenant des bases de données ClinVar et LOVD ainsi que 428 variants présents dans la population générale (dans la base de donnée gnomADv2.1.1). Seuls les variants éloignés de plus 2 nucléotides des bords des exons ont été sélectionnés pour ne pas inclure les variants localisés dans les sites d'épissage canoniques. Les variants dans ces deux sets ont ensuite été annotés par les logiciels de prédiction de l'épissage classiques (MaxEntScan (Yeo and Burge 2004)) ainsi que des logiciels plus récents, tels que MutPred Splice (Mort et al. 2014), SpliceAI (Jaganathan et al. 2019), MMSplice (Cheng et al. 2019), SCAP (Jagadeesh et al. 2019). Les 24 variants étudiés par l'étude « minigène » ont également été annotés par ces outils. Les variants faux-sens dans le gène *CAPN3* ont ensuite été visualisés en fonction de leur position dans le gène et le score de prédiction de l'épissage.

Enfin, les 24 variants explorés par l'approche « minigène » ont été divisés en deux groupes – les variants avec un effet sur l'épissage et les variants sans effet. Les résultats de différents scores de prédiction ont été comparés entre ces deux groupes.

Résultats

Dans cette section, une synthèse des sections respectives des trois articles est présentée. L'intégralité du travail effectué dans le cadre de cette thèse de médecine est exposée dans les articles ci-joints.

Utilisation de CovReport pour visualiser la couverture d'un test diagnostique par séquençage à haut débit

CovReport permet d'avoir une vue d'ensemble de la couverture des exons pour un test de séquençage diagnostique en générant un rapport concis et facile à comprendre au format PDF. Grâce à son interface intuitive et à l'absence d'étapes d'installation complexes, CovReport est facile à appliquer par les utilisateurs sans formation spéciale en bio-informatique. CovReport peut également être lancé en ligne de commande, ce qui permet son intégration dans des pipelines bio-informatiques. Plusieurs caractéristiques rendent CovReport particulièrement utile dans un contexte de diagnostic. Tout d'abord, l'application fonctionne sur un ordinateur local, ce qui permet une sécurité totale des données. Deuxièmement, le nom du patient de l'analyse précédente est automatiquement réinitialisé lors de la saisie du nouveau fichier de couverture afin d'éviter les erreurs d'identité. Troisièmement, le format, le contenu et la langue du rapport peuvent être facilement personnalisés par le laboratoire, en s'adaptant au contexte diagnostique. Quatrièmement, des informations sur le séquençage Sanger supplémentaire des régions insuffisamment couvertes peuvent être intégrées dans le rapport de couverture du séquençage à haut débit, ce qui permet de retracer facilement les tests de séquençage effectués pour un patient.

La couverture des exons pour les gènes d'un panel diagnostique est visualisée en dessinant les exons ombrés selon le niveau de couverture : les exons couverts à 100% sont blancs, les exons couverts à 90-100% sont gris clair, les exons couverts à <90% sont gris foncé, les exons non couverts sont noirs. Plusieurs options de visualisation permettent d'adapter la présentation graphique aux besoins de l'utilisateur. Les gènes avec une couverture de 100% peuvent être listés en haut du rapport sans dessiner les exons (option "Skip white genes", par défaut). Des structures schématiques gènes/exons seront dessinées pour les gènes restants, en ombrant les exons de couverture inférieure. Si l'option « Merge white exons » est sélectionnée, CovReport fusionnera les exons couverts à 100% pour une représentation plus compacte, ce qui est utile pour les gènes ayant de nombreux exons. Si cette option est décochée, le rapport sera produit avec chaque exon dessiné individuellement. De même, les exons ayant la même couleur peuvent être fusionnés pour une représentation plus compacte (option « Merge non-white exons »). La couverture moyenne par gène peut être indiquée à côté de chaque gène (option "Show gene weighted coverage"), qui est également utilisée pour les statistiques de la couverture totale du panel de gènes (option "Show statistics"). Des informations sur les transcrits RefSeq peuvent être incluses dans le rapport (option "Show gene transcripts"). Cette option est recommandée si le nombre d'exons est différent entre les transcrits, entraînant des différences de couverture entre les isoformes. La profondeur de la

couverture dans le rapport, par défaut de 20x, peut être modifiée en 1x, 5x, 10x ou 30x. Enfin, des commentaires supplémentaires peuvent être inclus dans le rapport.

Pour démontrer la performance de CovReport, nous avons utilisé les données de séquençage de l'exome pour l'échantillon de l'ADN standard (NA12878) afin de visualiser la couverture de 40 gènes de la liste étendue des gènes de la dystrophie musculaire des ceintures (LGMD) définie par le Filière Nationale des Maladies Rares Neuromusculaires (FILNEMUS). Le rapport de couverture de séquençage en format pdf a été généré à l'aide de CovReport avec les options « Afficher les transcrits des gènes » et « Afficher les statistiques » activées en plus des options par défaut. Comme le montre la Figure 3, la couverture a été visualisée pour 43 transcrits correspondant à 40 gènes. La plupart des gènes sont couverts à 100 %. L'avantage d'utiliser l'option "Show gene transcript" est évident dans cet exemple, puisque TTN (NM_133379) est couvert à 100 %, tandis que TTN (NM_001267550) possède plusieurs exons avec une couverture plus faible. En effet, l'isoforme NM_133379 est beaucoup plus courte, car les exons 50-219 du NM_001267550 ne sont pas inclus dans ce transcrit. Six exons ont une couverture inférieure à 20x (exons noirs). Plusieurs autres exons ont une couverture sous-optimale dans des régions correspondant à plus de 10 % de leur longueur (gris foncé). Ce rapport fournit une information importante pour la prise en charge du patient. En corrélation avec les données mutationnelles disponibles pour le gène, qui peuvent être discutées avec le généticien, le clinicien prescripteur prendra en compte les éléments cliniques disponibles pour le patient par rapport à la maladie suspectée, afin de décider si le phénotype du patient pourrait être potentiellement expliqué par des variants pathogènes dans les exons ayant une faible couverture, Si tel était le cas, un re-séquençage supplémentaire de ces régions serait nécessaire.

CovReport est utilisé pour tous les tests diagnostiques de séquençage de panel de gènes au Laboratoire de génétique moléculaire de l'hôpital de la Timone de l'AP-HM, depuis juin 2018. Les rapports générés sont directement annexés aux résultats diagnostiques des différents panels de séquençage. En outre, les rapports de couverture produits par CovReport sont également utilisés en routine pour la validation technique du test en prenant en compte les statistiques de couverture des gènes du panel intégrées dans les rapports.

Annexe 1. Couverture des séquences cibles

Nom du patient: Patient Macel

Généré le: 1 mars 2020 15:45

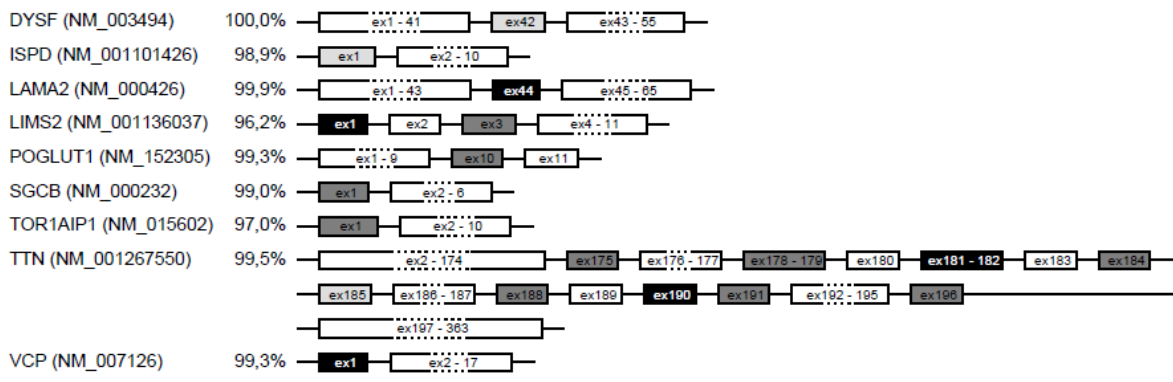
Les gènes suivants sont couverts à 100% (avec une profondeur d'au moins 20x):

ANO5 (NM_213599), BVES (NM_147147), CAPN3 (NM_000070), CAV3 (NM_033337), COL6A1 (NM_001848), COL6A2 (NM_001849), COL6A3 (NM_004369), DAG1 (NM_001165928), DES (NM_001927), DNAJB6 (NM_058246), DPM3 (NM_153741), FHL1 (NM_001159702), FKRP (NM_024301), FKTN (NM_001079802), GAA (NM_000152), GAA (NM_001079803), GMPPB (NM_013334), GNE (NM_001128227), HNRNPDL (NM_001207000), HNRNPDL (NM_031372), LMNA (NM_170707), MYOT (NM_006790), PLEC (NM_000445), POMGNT1 (NM_001243766), POMT1 (NM_007171), POMT2 (NM_013382), SGCA (NM_000023), SGCD (NM_000337), SGCG (NM_000231), TCAP (NM_003673), TNPO3 (NM_012470), TRAPPC11 (NM_021942), TRIM32 (NM_012210), TTN (NM_133379)

Certaines parties des gènes suivants ne sont pas couverts à 100% (voir les détails ci-dessous)

Légende:

- exons couverts à 100% (avec une profondeur d'au moins 20x)
- exons couverts à 90-100% (avec une profondeur d'au moins 20x)
- exons couverts à <90% (avec une profondeur d'au moins 20x)
- exons non couverts (avec une profondeur d'au moins 20x)



Générée à partir de: 2018A782.sorted.dedup.recalib_FILNEMUS-LGMD-exhaustive_hg19.stat_coverage_exon.txt

Statistiques de l'analyse:

- nombre de gènes: 40
- nombre de transcripts: 43
- transcripts couverts à 100% (avec une profondeur d'au moins 20x): 79,1%
- transcripts couverts à 90-100% (avec une profondeur d'au moins 20x): 20,9%

Figure 6 : Un exemple d'un rapport de couverture généré par CovReport.

Un exemple de rapport a été généré par CovReport afin de visualiser la couverture du séquençage de 40 gènes du panel LGMD (dystrophies musculaires des ceintures) en utilisant les données de séquençage d'un exome pour l'échantillon d'ADN standard NA12878. Le nom du patient est affiché comme indiqué dans l'interface de CovReport. La date du rapport est générée automatiquement (si l'option "Afficher la date du rapport" est activée comme dans cet exemple). Dans la partie supérieure du rapport, les gènes couverts à 100% sont listés (si l'option "Exclure les gènes blancs" est activée). Ensuite, les gènes dont la couverture est sous-optimale sont indiqués avec une couleur des exons en fonction du niveau de couverture. Le pourcentage de couverture au niveau du gène est indiqué pour chaque gène couvert de manière sous-optimale (si l'option "Afficher la couverture pondérée par gène" est activée). Enfin, les statistiques de couverture pour les gènes analysés sont affichées (si l'option "Afficher les statistiques" est activée).

Mise à jour des recommandations françaises pour l'interprétation des variants

Les recommandations initiales pour l'interprétation des variants identifiés par séquençage à haut débit proposées par ACMG-AMP (Amendola et al., 2016; Richards et al., 2015) et par le réseau NGS-Diag reposent sur l'évaluation des 28 arguments pour estimer le caractère

pathogène ou bénin d'un variant. Chaque argument a un poids (« faible », « moyen », « fort » ou « très fort ») qui est comptabilisé pour l'attribution d'un classement global de l'impact clinique pour un variant donné. Pour mettre à jour les recommandations françaises datant de juin 2018, nous avons effectué un travail de synthèse bibliographique étudiant dix publications proposant des mises à jour spécifiques pour un gène ou un groupe des pathologies. Six précisions ou recommandations pour des critères individuels issus des groupes de travail ClinGen ont également été examinés. Suite à ce travail, nous avons proposé l'actualisation des recommandations françaises de classification des variants, avec intégration des éléments supplémentaires depuis la publication ACMG princeps. La version actualisée de ce document est en cours de validation par les réseaux NGS-Diag et ANPGM.

Trois arguments des recommandations initiales (PM2, BS1 et BA1) sont basés sur les données de fréquence allélique du variant dans la population générale. Nous avons tout d'abord mis à jour les informations sur les bases de données qui sont devenues disponibles depuis la première version du document NGS_DIAG_001, tel que gnomAD (Karczewski et al., 2019), TopMed (Taliun et al., 2019; TOPMed, 2018) et autres. Des précisions pour ces arguments sont particulièrement nécessaires car de grandes différences sont possibles entre des laboratoires pour attribuer l'argument pathogène PM2 (*absence ou fréquence très faible du variant dans la population générale*) ou l'argument bénin BS1 (*présence avec une fréquence élevée*). En effet, les seuils choisis dépendent largement de la prévalence de la pathologie associée au gène étudié ainsi que sa variabilité allélique. Un outil aidant au calcul de ces seuils a récemment été décrit (Whiffin et al., 2017). Un résumé des seuils déjà calculés proposé dans la nouvelle version du document NGS-Diag (Tableau 1 et Figure 6) sera utile non seulement pour les laboratoires interprétant les variants dans ces gènes, mais également pour l'interprétation des variants dans des gènes avec des caractéristiques similaires.

| | Pathologies Dominantes | | | | | | | | Pathologies Récessives | | |
|----------------|------------------------|------------|----------|-------------|------------------------|------------|----------|---------|------------------------|--------------|--------|
| | MYH7 | Rasopathie | PTEN | Surdités AD | CDH1 | MEN1 | RUNX1 | TP53 | Surdités AR | PAH | GAA |
| PM2 | < 0,004% | 0 | < 0,001% | < 0,002% | < 0,001% (ou < 0,002%) | 0-1 | 0 | | < 0,007% | < 0,02% | < 0,1% |
| PM2_sup | | | | | | | 0 | < 0,07% | | | |
| BS1 | > 0,02% | > 0,025% | > 0,1% | > 0,02% | > 0,1 % | > 0,00784% | > 0,015% | > 0,03% | > 0,3% | > 0,2 - 1,5% | > 0,5% |
| BS1_sup | | | | | | | | > 0,07% | | | |
| BA1 | > 0,1% | > 0,05% | > 1% | > 0,1% | > 0,2% | > 0,76% | > 0,15% | > 0,1% | > 0,5% | > 1,5% | > 1% |

Tableau 1 : Seuils de fréquences alléliques du variant à classer pour attribuer PM2, BS1 ou BA1.

Un résumé des recommandations issues des groupes d'experts pour les différents gènes ou pathologies. Pour calculer ces seuils, les données de prévalence et pénétrance suivantes ont été utilisées : MYH7 1/200 avec une pénétrance de 30% (pour BA1/BS1) et 1/500 avec une pénétrance de 50% (pour PM2) (Kelly et al., 2018). Rasopathie – 1/1000 (pour BA1/BS1) et 1/100 000 (pour PM2) avec une pénétrance de 100% (Gelb et al., 2018). PTEN – 1/250 000 (Mester et al., 2018). Surdités – 1/30 -1/200 prévalence avec une pénétrance de 80-100% (Oza et al., 2018). CDH1 – 1/800 – 1/1250 avec une pénétrance de 30% (pour BA1/BS1) (Lee et al., 2018). MEN1 – 1/30 000 avec une pénétrance de 17% (Romanet et al., 2019). RUNX1 – 1/40 avec une pénétrance de 85% (Luo et al., 2019). PAH – 1/13500 – 1/19000 avec une pénétrance de 80% (Zastrow et al., 2018). TP53 (ClinGen_TP53_Expert_Panel, 2019), GAA (ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019).

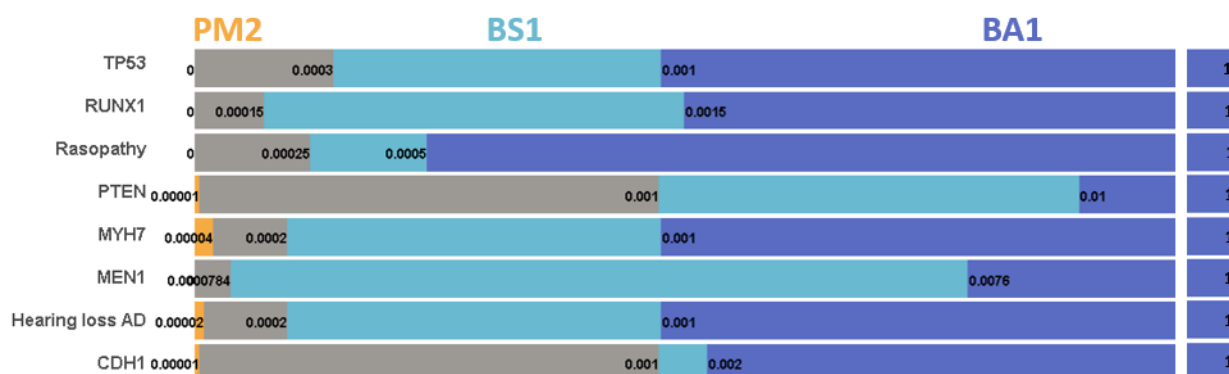


Figure 7 : Un résumé graphique des seuils des fréquences alléliques pour attribuer les arguments PM2/BS1/BA1 selon les recommandations gène-spécifique (pathologies AD)

Une prévalence d'un variant chez les individus atteints significativement supérieure à celle des individus contrôles de la même origine ethnique que les patients permet d'attribuer l'argument PS4 selon les recommandations originales. Cependant, il existe très peu d'études cas-contrôles pour avoir assez de puissance statistique pour faire ce calcul. Les recommandations gène-spécifiques proposent d'attribuer le score PS4 en fonction du nombre de patients porteurs du variant étudié. Huit recommandations gène-spécifiques récentes proposent des seuils pour attribuer l'argument PS4 (Tableau 2). Cette façon d'attribuer PS4 n'est possible que si le variant est classé absent ou extrêmement rare dans la population générale (PM2 attribué). Le résumé de ces recommandations peut aider les biologistes dans l'interprétation des gènes impliqués dans les maladies avec des prévalences similaires.

| | MYH7 | Rasopathie | PTEN | Surdités AD | CDH1 | MEN1 | RUNX1 | TP53 |
|----------------|------|------------|------|-------------|------|---------------------|-------|------|
| PS4_VeryStrong | | | > 16 | | > 16 | | | |
| PS4_Strong | > 15 | > 5 | 4-15 | > 15 | 4-15 | case-control needed | > 4 | > 4 |
| PS4_Moderate | 6-14 | 3-4 | 2-3 | 6-14 | 2-3 | 3 | 2-3 | 2-3 |
| PS4_Supporting | 2-5 | 1-2 | 1 | 2-5 | 1 | 2 | 1 | 1 |

Tableau 2. Seuils de nombre de patients confirmés porteurs de variant pour attribuer l'argument PS4.

Un résumé des recommandations pour huit pathologies ou groupes des pathologies dominantes : MYH7 (Kelly et al., 2018), Rasopathie (Gelb et al., 2018), PTEN (Mester et al., 2018), Surdités (Oza et al., 2018), CDH1 (Lee et al., 2018a), MEN1 (Romanet et al., 2019), RUNX1(Luo et al., 2019), TP53 (ClinGen_TP53_Expert_Panel, 2019). Selon les recommandations pour PTEN et TP53, les évaluations phénotypiques sont également calculées en parallèle, avec nécessité d'avoir plus de patients pour un niveau de PS4 si les phénotypes des patients sont moins évidents.

Pour certains critères, les précisions dans les publications récentes vont faciliter l'interprétation et l'attribution des arguments de classification par le biologiste. C'est le cas pour l'argument PM3 (*dans le cas d'une pathologie à transmission autosomique récessive, variant observé en trans avec un variant pathogène*) et les arguments PS2/PM6 (*variant de novo chez un patient sans antécédent familial, avec confirmation de la paternité et de la maternité (PS2) ou sans confirmation de la paternité et de la maternité si la pénétrance est considérée complète (PM6)*). Les nouvelles recommandations du groupe de travail ClinGen SVI propose le système de calcul de score PM3 (SVI_Recommendation_for_PM3_Criterion, 2019) ainsi que de PS2/PM6 (SVI_Recommendation_for_PS2_and_PM6_criteria, 2018). Le nombre de points précis est donné pour chaque cas en fonction de la pathogénicité du variant *en trans*,

le niveau de preuve que le variant soit vraiment *de novo* ou la spécificité du phénotype. Ainsi la somme de ces points définit le poids des arguments PM3 ou PS2/PM6. De même, un arbre décisionnel très précis a récemment été établi pour l'attribution de l'argument PVS1 (*variant ayant un effet perte de fonction prédit*) (Abou Tayoun et al., 2018) et est proposé d'être incorporé dans les nouvelles recommandations françaises.

Un autre argument pour lequel les précisions ont été rajoutées dans la nouvelle version des recommandations NGS-Diag est PP1 (*variant co-ségrégant avec la pathologie chez plusieurs membres atteints dans la même famille et/ou dans plusieurs familles*). Les tableaux avec des LOD scores pré-calculés ont été rajoutés. Le résumé des dix recommandations gène-spécifiques peut être utile pour aider le biologiste à attribuer le score PP1 (Tableau 2).

| | MYH7, Rasopathy PTEN, CDH1, RUNX1, TP3 | MEN1 | Hearing loss AD | PAH | Hearing loss AR |
|-----------------------|--|---|--|--|-----------------|
| PP1_Strong | > 7 meioses (likelyhood ratio 100, LOD 2,1) | N ≤ 1/32 if 1 family N ≤ 1/16 if 2+ families | 5 affected (likelyhood ratio 1:32, LOD 1,5) | 3 affected + 0 non-affected 2 affected + 3 non-affected | 3 affected |
| PP1_Moderate | 5-6 meioses (likelyhood ratio 30, LOD 1,5) | N ≤ 1/16 if 1 family N ≤ 1/8 if 2+ families | 4 affected (likelyhood ratio 1:16, LOD 1,2) | 2 affected + 0 non-affected | 2 affected |
| PP1_Supporting | 3-4 meioses (likelyhood ratio 10, LOD 0,9) | N ≤ 1/8 if 1 family N ≤ 1/4 if 2+ families | 2 affected (likelyhood ratio 1:4, LOD 0,6) | 1 affected + 3 non-affected | 1 affected |

Tableau 3. Les seuils de nombre de ségrégations nécessaires pour attribuer l'argument PP1.

Un résumé des recommandations issues des panels d'experts pour les gènes ou pathologies suivants : MYH7 (Kelly et al., 2018), Rasopathie (Gelb et al., 2018), PTEN (Mester et al., 2018), Surdités (Oza et al., 2018), CDH1 (Lee et al., 2018a), MEN1 (Romanet et al., 2019), RUNX1 (Luo et al., 2019), TP53 (ClinGen_TP53_Expert_Panel, 2019), PAH (Zastrow et al., 2018), GAA (ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019).

La façon selon laquelle l'argument PP3 (*un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés*) est attribué dépend largement des outils bio-informatiques utilisés pour la prédiction de la pathogénicité du variant. Devant l'évolution rapide du domaine de la bio-informatique, de nombreux outils de prédiction puissants sont devenus disponibles depuis la publication de la première version du document de recommandations émis par le réseau NGS-Diag. Nous avons décrit ces nouveaux outils et rajouté l'information sur les études de comparaison de la performance de ces logiciels en diagnostic (Ghosh et al., 2017).

Enfin, les arguments PP5 et BP6 ont été retirés de la nouvelle recommandation du réseau NGS-Diag en concordance avec celles de ClinGen (Biesecker et al., 2018), ainsi que de la grande majorité des recommandations gènes-spécifiques. En effet, la présence ou l'absence d'un variant dans les bases de données de patients est déjà évaluée pendant l'attribution de PS4. Au total, des précisions ou des mises à jour ont été rajoutées pour onze arguments de la première version des recommandations françaises pour l'interprétation des variants issus du séquençage à haut débit.

Etudes fonctionnelles des variants faux-sens dans le gène CAPN3

Afin d'évaluer l'impact de l'épissage de variants faux-sens situés localisés à distance des jonctions exon-intron (variants faux-sens exoniques « profonds »), nous avons utilisé une étude fonctionnelle de type « minigène » pour tester 21 variants cliniquement pertinents distribués de manière homogène le long du gène CAPN3 (Figure 7). Sur les 21 variants testés,

des défauts d'épissage ont été mis en évidence pour 8 dans le test « minigène ». Parmi ceux-ci, cinq variants ont induit l'expression d'un transcrit plus court avec absence d'un exon (saut d'exon), associée à la diminution du transcrit normal (c.551C>G, p.(Thr184Arg) ; c.551C>T, p.(Thr184Met) ; c.743T>G, p.(Met248Arg) ; c.1063C>G, p.(Arg355Gly) et c.1466G>A, p.(Arg489Gln). Dans tous ces cas, le saut de l'exon a été associé à la perte du cadre de lecture, induisant probablement une protéine tronquée ou l'absence de protéine si le processus NMD (non-sense mediated decay) était induit pour le transcrit anormal. Ainsi, la quantité de la protéine calpaïne3 présente dans les cellules musculaires des patients porteurs de ces variants sera probablement inférieure à la normale.

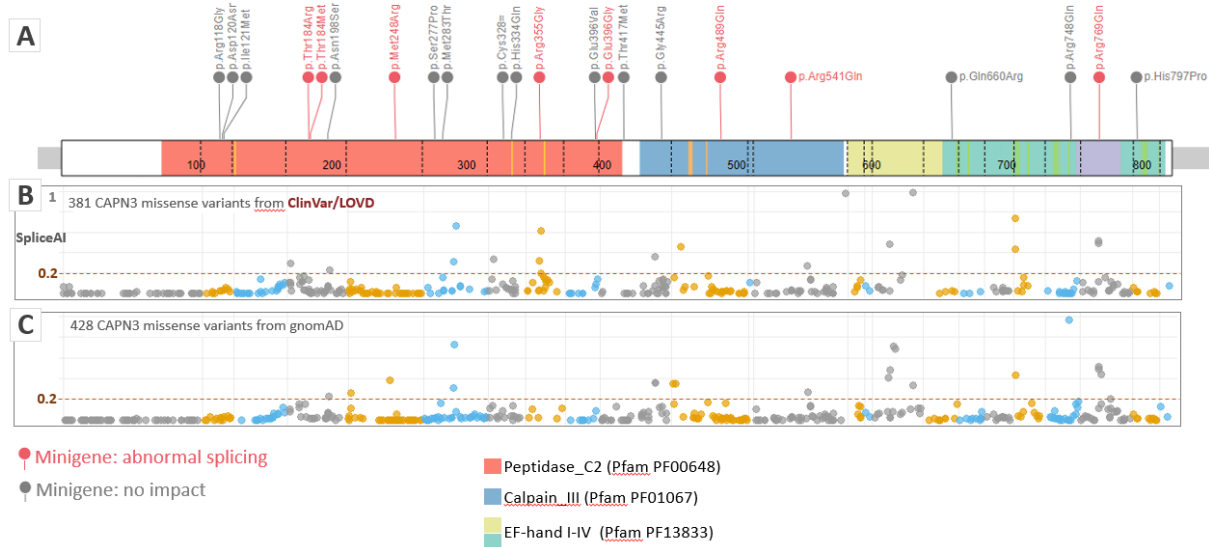


Figure 8 : Variants faux-sens exoniques profonds dans le gène CAPN3.

A). Vingt-et-un variants cliniquement pertinents testés par l'étude « minigène » sont représentés (les variants avec un effet sur l'épissage sont en rouge, les variants sans impact sont en gris). Les jonctions exon-intron du gène CAPN3 ainsi que les domaines fonctionnels de la calpaïne 3 sont visualisés.

B). Les variants faux-sens cliniquement pertinents (pathogènes, probablement pathogènes et incertaines/conflictuelles) de bases de données LOVD et ClinVar sont visualisés le long du gène CAPN3 positionnés verticalement selon leurs scores SpliceAI.

C). Les variants faux-sens exoniques profonds de gnomADv2.1.1 sont visualisés le long du gène CAPN3 positionnés verticalement selon leurs scores SpliceAI. Les variants dans les exons adjacents sont colorés différemment pour visualiser les jonctions entre les exons. Les grilles verticales des panneaux B et C correspondent aux limites des exons. Les variants +/- 2 nucléotides des bords exoniques sont exclus de cet analyse.

Afin d'explorer la localisation des variants faux-sens exoniques affectant l'épissage, nous avons calculé les distances de ces variants aux extrémités 5' et 3' des exons. Ces variants ont été localisés dans sept exons différents (4, 5, 8, 9, 11, 13 et 22). Pour les exons supérieurs à 100nt, nous avons visualisé les variants exoniques affectant l'épissage en fonction de leur distance par rapport aux extrémités 5' ou 3' des exons (figure 8). Il est intéressant de noter que deux variants exoniques profonds ayant un impact sur l'épissage (c.743T>G, p.(Met248Arg) dans l'exon5 et c.1466G>A, p.(Arg489Gln) dans l'exon 11) ont été localisés à la même distance de 59 nucléotides de l'extrémité 3' de l'exon.

Nous avons ensuite exploré tous les variants faux-sens CAPN3 connus situés dans les régions exoniques profondes pour estimer leur effet potentiel sur l'épissage. Pour cela, nous avons téléchargés des variants faux-sens cliniquement pertinents (pathogènes, probablement

pathogènes et incertaines/conflictuelles) à partir des bases de données LOVD et ClinVar. Nous avons ensuite sélectionné uniquement les variants situés en dehors des sites d'épissage canoniques et utilisé le score SpliceAI (Jaganathan et al. 2019) pour estimer l'effet sur l'épissage pour chaque variant. Les variants obtenus ont été visualisés le long du gène *CAPN3* avec les limites des exons marqués (Figure 7B). Les variants faux-sens exoniques profonds présents dans la population générale (gnomADv2.1.1) ont également été analysés (Figure 7C). Comme le montrent ces résultats, certaines régions exoniques contenaient des groupes de variants avec des scores SpliceAI plus élevés.

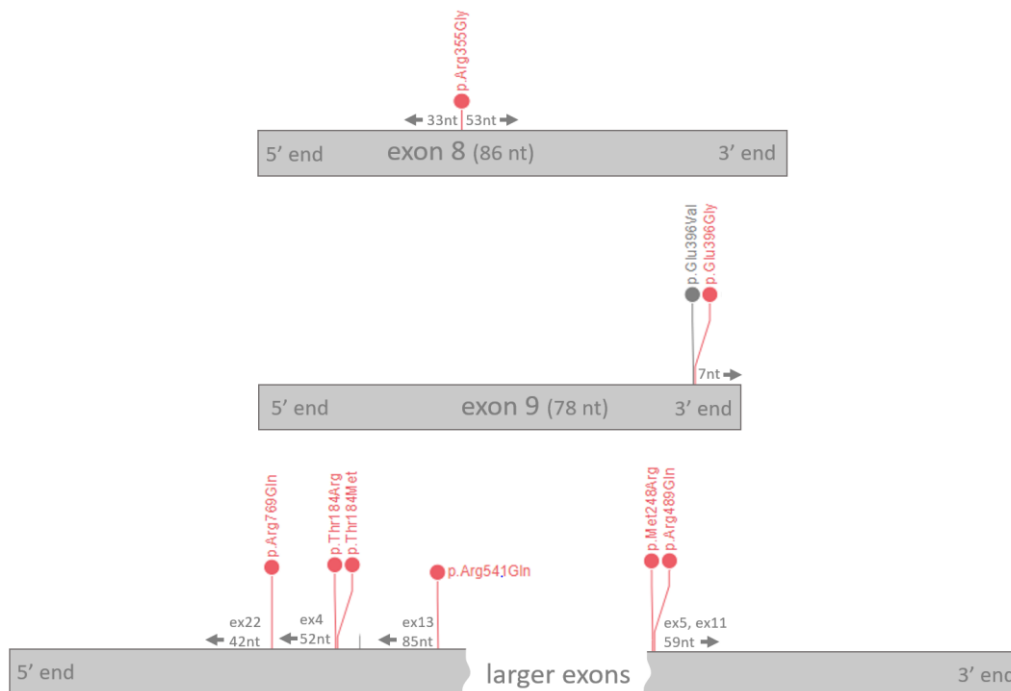


Figure 9 : Localisation de variants exoniques profonds induisant des défauts d'épissage.

Les distances entre les 8 variants faux-sens avec un effet sur l'épissage (rouge) et les bords des exons sont montrés.

Malgré le besoin critique en diagnostic, les algorithmes de prédiction de l'effet d'épissage actuellement disponibles ne sont toujours pas optimaux pour estimer correctement l'impact sur l'épissage, en partie en raison du nombre insuffisant d'études de validation fonctionnelle. Notre série de variants *CAPN3* étudiés fonctionnellement par l'approche « minigène » bien établie offre une excellente occasion de tester les récents algorithmes de prédiction d'épissage. Les 21 variants explorés dans le cadre de cette étude devaient avoir un impact sur l'épissage selon l'outil HSF (Desmet et al., 2009), mais seuls huit ont montré une altération de l'épissage dans le test fonctionnel, démontrant un taux significatif de faux positifs. Dans l'ensemble, la performance de plusieurs outils de prédiction n'était pas optimale pour les variants faux-sens exoniques profonds : 5 sur 8 prédits délétères par MaxEnsScan (Yeo and Burge, 2004), 2 sur 8 par SpliceAI (Jaganathan et al., 2019), 3 sur 8 par MutPred Splice (Mort et al., 2014) et 0/8 pour S-CAP (Jagadeesh et al., 2019) (Figure 9).

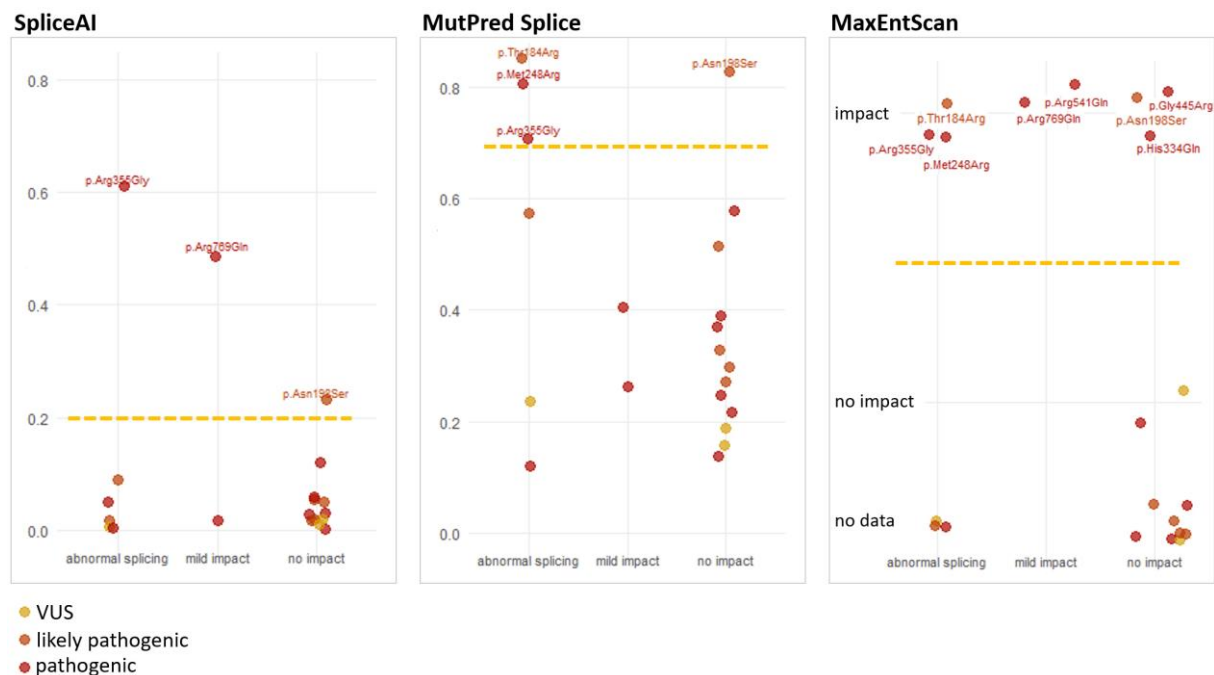


Figure 10 : Performances des algorithmes prédisant l'impact sur l'épissage pour les variants exoniques profonds

Les variants avec un score SpliceAI supérieur à 0,2 sont prédits comme ayant un effet sur l'épissage. Il y a six faux-négatifs et un faux-positif pour l'outil SpliceAI. L'outil MutPred Splice prédit trois sur huit variants avec un effet sur l'épissage (cinq faux négatifs). Les variants avec un score MutPred Splice supérieur à 0,7 sont considérés comme ayant un effet sur l'épissage. L'outil MaxEntScan a trois faux-négatifs et trois faux-positifs.

Comme la plupart des variants sélectionnés pour l'étude par « minigène » ont été identifiés avant les recommandations pour l'interprétation des variants, nous avons classé chaque variant selon les recommandations d'ACMG modifiées pour le gène *CAPN3*. Les résultats des tests fonctionnels permettant d'attribuer des scores PS3 ou PS3_sup supplémentaires, ont ensuite été pris en compte pour calculer la classification « post-minigène » pour chaque variant. Pour 7 des 24 variants testés par « minigène », la classification ACMG a été modifiée sur la base des résultats de cet essai fonctionnel, montrant une grande utilité en diagnostic.

Discussion

Le raccourcissement du délai nécessaire pour obtenir des quantités massives d'informations génétiques pour un patient distingue nettement l'approche par séquençage à haut débit des autres tests diagnostiques en génétique. En effet, l'intégration du séquençage à haut débit dans la pratique clinique a permis d'accélérer considérablement la rapidité des tests de génétique moléculaire, aidant grandement à la prise en charge des patients en terme de démarche diagnostique et de conseil génétique. Cependant, la quantité et la nature de données générées nécessitent des approches innovantes et puissantes pour répondre aux défis de la médecine génomique, notamment afin d'assurer un rendu de données respectant des exigences de qualité compatibles avec une utilisation en diagnostic.

Améliorations apportées par utilisation de CovReport dans le cadre du diagnostic par séquençage à haut débit

L'importance des données de couverture est bien établie dans le diagnostic génétique. En effet, la plupart des rapports de séquençage fournissent des informations sur la couverture des régions cibles. Toutefois, ces données sont généralement présentées comme une couverture globale du panel de gènes ou comme une couverture moyenne par gène. À notre connaissance, aucune donnée de couverture au niveau exonique est fournie avec les rapports de diagnostic. Cependant, la distribution des variants pathogènes dans un gène donné n'est pas homogène, en partie car les modifications des acides aminés responsables de la maladie sont souvent regroupées dans des régions fonctionnellement importantes de la protéine. Ainsi, les informations sur la couverture des régions où sont concentrés les variants pathogènes sont beaucoup plus instructives qu'une couverture moyenne globale des gènes (Gorokhova et al., 2015b). CovReport, le logiciel développé au sein de notre laboratoire, met en évidence les exons individuels dont la couverture est sous-optimale, ce qui facilite l'interprétation des données de séquençage.

L'objectif ultime du séquençage diagnostique est d'obtenir une couverture suffisante pour l'ensemble des régions génomiques ciblées par un panel de gènes donné. Cela est possible pour des petits panels dont les gènes peuvent être séquencés de manière robuste. Toutefois, cet objectif est beaucoup plus difficile à atteindre pour de grands panels de gènes contenant des gènes difficiles à séquencer. C'est le cas pour le domaine des maladies neuromusculaires, puisque 33 des 203 gènes figurant sur les listes consensuelles de gènes de myopathie (Krahn et al., 2019) contiennent des régions "sombres" du génome qui ne sont pas facilement accessibles à l'aide des approches standards de séquençage à haut débit (Ebbert et al., 2019). Les variants pathogènes dans ces régions peuvent donc être non-identifiés, ce qui conduit à un diagnostic moléculaire faussement négatif. CovReport met en évidence et attire l'attention sur les exons dont la couverture est sous-optimale, permettant au prescripteur, en corrélation avec le généticien, d'évaluer la nécessité d'un test de re-séquençage complémentaire pour ces régions.

Depuis l'arrivée de la technologie de séquençage à haut débit, plusieurs outils ont été développés pour calculer la couverture de séquençage, tels que BEDTools genomecov (Quinlan, 2014; Quinlan and Hall, 2010), Sambamba (Tarasov et al., 2015), et Mosdepth (Pedersen and Quinlan, 2018). Toutefois, la manière optimale de visualiser et d'évaluer les données de couverture obtenue n'est pas définie de manière universelle. Plusieurs outils ont été conçus pour évaluer la couverture des régions cibles après un séquençage à haut débit dans un cadre de diagnostic (Doig et al., 2017; Münz et al., 2018). Si ces outils sont utiles pour surveiller le contrôle de la qualité et la validation technique des tests de séquençage, à la différence de CovReport, ils ne permettent pas la visualisation de la couverture au niveau des exons qui peut être directement annexée au rapport de diagnostic. À notre connaissance, CovReport est le premier outil qui permet une évaluation complète et visuelle des résultats de la couverture d'un test de séquençage, ce qui permet aux généticiens moléculaires ou aux médecins prescripteurs de prendre des décisions diagnostiques importantes, comme la poursuite de l'exploration des régions génomiques mal couvertes.

Les défis de l'interprétation de variants identifiés par séquençage à haut débit et l'importance des RCP (Réunions de Concertation Pluridisciplinaire)

Le génome d'un individu comporte en moyenne 4-5 millions de différences avec le génome de référence (1000 Genomes Project Consortium et al., 2015). Pour un panel de 200 gènes des maladies neuromusculaires, le nombre de variants identifiés chez un patient s'élève à 200-400. Si un/des variant(s) pathogène(s) (selon le mode de ségrégation suspecté) est/sont caractérisé(s) de manière cohérente avec les symptômes du patient est présent parmi ces variants, le diagnostic moléculaire peut être posé sans difficulté. Cependant, l'interprétation des résultats de séquençage d'un panel de gènes peut être difficile devant les situations suivantes. Si un variant identifié chez un patient est mal connu ou jamais identifié auparavant, plusieurs éléments de preuve doivent être pris en compte pour conclure à la pathogénicité de ce variant ainsi qu'à son implication dans la présentation clinique du patient. Pour guider le biologiste dans cette démarche, des recommandations nationales (Recommandation NGS-Diag v1) et internationales (Amendola et al., 2016; Richards et al., 2015) pour l'interprétation des variants issus des tests par séquençage à haut débit ont été émises. Les éléments supplémentaires, comme la ségrégation familiale, les explorations du phénotype par des examens complémentaires ou les données d'études fonctionnelles peuvent être nécessaires pour avancer dans la démarche diagnostique. D'autres situations problématiques en diagnostic sont la présence simultanée de plusieurs variants potentiellement pathogènes ou de signification incertaine chez un patient, ou la présence d'un variant pathogène ne corrélant pas avec la présentation clinique du patient. Enfin, le séquençage à haut débit peut révéler la présence d'un variant pathogène dans un gène responsable d'une pathologie différente de l'indication initiale. Devant ces défis d'interprétation, les Réunions de Concertation Pluridisciplinaire (RCP) sont mises en place dans la grande majorité de centres de diagnostic génétique. Elles permettent d'échanger sur les données indispensables pour mieux orienter le diagnostic. Les éléments cliniques et paracliniques recueillis par l'équipe clinique sont nécessaires pour l'interprétation et la priorisation des variants identifiés chez le patient. D'autre part, les résultats du séquençage peuvent révéler des variants dans des gènes qui n'étaient pas encore évoqués chez le patient, orientant alors les explorations phénotypiques vers une autre hypothèse étiologique. Dans le cadre des maladies neuromusculaires, les généticiens de notre Service de Génétique Médicale (Pr. N. Lévy) participent régulièrement aux RCP organisées par le Service des Maladies Neuromusculaires et de la SLA dirigé par le Pr. Attarian, le Service de Neuropédiatrie dirigé par le Pr. Chabrol. Ces réunions ont permis de résoudre un certain nombre d'observations difficiles en aboutissant à des diagnostics moléculaires.

La double face des variants faux-sens exoniques profonds

Malgré les récents progrès de la médecine génomique, la pathogénicité de la plupart des variants identifiés dans le cadre d'un test de séquençage diagnostique n'est souvent pas bien établie. Les outils bio-informatiques visant à prédire l'effet du variant identifié sur la fonction de la protéine ou sur l'épissage de l'ARNm jouent donc un rôle important dans la démarche diagnostique (Hu et al., 2019). Cependant, ces algorithmes ne fonctionnent pas de la même manière pour tous les types de variants, ce qui nécessite en fin de compte une confirmation fonctionnelle de ces prédictions. L'interprétation de la pathogénicité d'un variant est souvent guidée par sa localisation : les variants introniques ainsi que les variants exoniques à proximité des jonctions exon-intron sont généralement supposés perturber l'épissage de l'ARNm, tandis que les variants faux-sens dans les régions codantes sont considérées comme pouvant altérer

la fonction des protéines. Cependant, plusieurs études ont déjà attiré l'attention sur des exemples de variants exoniques distants des sites canoniques et perturbant l'épissage normal (Kergourlay et al., 2014b; Savisaar and Hurst, 2017). Dans cette étude, nous avons systématiquement criblé 16 exons du gène *CAPN3* pour ce type de variants, en utilisant l'approche « minigène » pour tester fonctionnellement les variants faux-sens situés en dehors des régions d'épissage canonique (variants exoniques profonds). Nous avons détecté des anomalies d'épissage pour huit des 21 variants faux-sens testés. Il est intéressant de noter que ces huit variants faux-sens perturbant l'épissage sont aussi prédits délétères sur la fonction de la protéine calpaine 3 en raison du changement d'acide aminé. Il est difficile de distinguer l'effet pathogène sur la protéine de l'effet sur l'épissage de l'ARNm pour ces variants. Il est possible que ces deux mécanismes contribuent à la pathogénicité globale *in vivo*. L'interaction entre l'effet du variant sur la fonction de la protéine et son impact sur l'épissage pourrait potentiellement avoir des conséquences importantes sur la fonction cellulaire et l'expression de la maladie. Par exemple, l'effet d'un variant pathogène dominant négatif au niveau de protéine pourrait être réduit si son effet sur l'épissage était augmenté, réduisant ainsi la quantité du transcrite toxique. Des études récentes ont montré une variabilité non-négligeable dans l'épissage de l'ARNm entre les individus et les différentes populations en général (Lu et al., 2012; Park et al., 2018). Ainsi, on peut imaginer qu'un variant connu pour avoir un effet pathogène dominant négatif pourrait avoir une fréquence plus élevée dans une population particulière si ce variant avait plus d'impact sur l'épissage dans cette population. Au contraire, si le principal effet pathogène d'un variant provient de son impact sur l'épissage, les transcrits anormaux pourraient être exprimés à des niveaux plus faibles chez certains individus ou populations porteurs de variants modifiant l'épissage, conduisant à une expression plus légère de la maladie (Cooper et al., 2013). Cela peut être le cas de c.551C>T:p.(Thr184Met), l'un des variants montrant un épissage anormal dans notre étude par « minigène ». Ce variant faux-sens a longtemps été considéré comme pathogène, responsable de formes plus légères de LGMDR1 (Anderson et al., 1998; Barp et al., 2020; Fanin et al., 2004; Richard et al., 1999). Initialement, ce variant avait une très faible fréquence dans les sous-populations testées (0,001 % chez les Européens, 2 allèles sur 129164 dans gnomADv2.1.1). Toutefois, avec les données de séquençage plus nombreuses actuellement disponibles pour différentes populations, il s'avère que ce variant a une fréquence de 3,3 % dans la population Africaine et de 0,09 % dans la population Latino, conduisant à sa reclassification comme bénin ou probablement bénin par de nombreux laboratoires de diagnostic. Compte tenu de nos résultats montrant un effet sur l'épissage de ce variant, d'autres études sont nécessaires pour établir son rôle dans le LGMDR1. Un autre exemple de variant affectant l'épissage ayant une fréquence allélique élevée dans la population générale (1% en population finnoise) est le variant c.1746-20C>G (Nascimbeni et al., 2010). Il est intéressant de noter que des différences dans l'expression de transcrits anormaux ont été détectées entre les patients porteurs de ce variant, montrant ainsi la variabilité inter-individuelle dans les patterns d'épissage (Nascimbeni et al., 2010).

Les défis de la prédiction bio-informatique de l'effet d'un variant sur l'épissage

Étant donné la grande importance clinique de la détection précise des variants perturbant l'épissage, un grand nombre d'algorithmes de prédiction de l'effet sur épissage ont été développés. Des outils plus anciens, tels que MaxEntScan et HSF (Desmet et al., 2009; Yeo and Burge, 2004), sont encore largement utilisés dans le diagnostic, tandis que des approches plus récentes basées sur *machine learning*, telles que SpliceAI, S-CAP et MMSplice, gagnent

également en importance (Rowlands et al., 2019). Cependant, ces outils de prédiction ont été conçus et testés en utilisant les variants d'épissage déjà connus, dont la plupart sont introniques ou situés sur les sites d'épissage canoniques. Dans cette étude, nous nous sommes concentrés sur les variants faux-sens susceptibles de perturber l'épissage et situés plus profondément dans les exons, groupe de variants qui ont été relativement peu explorés (Savisaar and Hurst, 2017). Nous avons constaté que les performances des outils testés étaient sous-optimales, car ils ne permettaient pas de prédire l'effet sur l'épissage pour plus de la moitié des variants pour lesquels un impact sur l'épissage a été démontré dans notre étude par « minigène ». L'ensemble des résultats suggère que les variants exoniques ayant un effet sur l'épissage pourraient être plus difficiles à prédire en utilisant les algorithmes actuellement disponibles. Les variants exoniques perturbant l'épissage identifiés dans cette étude pourraient contribuer à la description d'un nouveau recueil de données pour ce type de variants et ainsi améliorer les outils de prédiction d'épissage.

Les tests fonctionnels de l'effet sur épissage dans le cadre du diagnostic des maladies neuromusculaires

L'accessibilité et l'utilisation large du séquençage à haut débit pour le diagnostic de la dystrophie musculaire des ceintures ont permis d'identifier de multiples variants de pathogénicité incertaine, augmentant la nécessité d'études fonctionnelles efficaces. Pour la forme LGMDR1, la présence de la protéine calpaïne 3 sur le Western Blot, ou la détection de transcripts anormaux à partir de l'ARN provenant du muscle atteint, est très utile pour avancer dans le diagnostic moléculaire. Cependant, ces deux approches nécessitent l'accès au matériel de biopsie musculaire du patient. Une alternative *in vitro* est l'approche par l'étude « minigène ». Il a été précédemment démontré que les résultats du test « minigène » sont concordants avec l'analyse de l'ARN du patient (Bonnet et al., 2008; Théry et al., 2011; Tournier et al., 2008). Ici, nous avons développé l'approche « minigène » pour la majorité des exons du gène *CAPN3* et l'avons utilisé pour tester 24 variants. L'épissage anormal dans le test « minigène » a permis d'attribuer des points supplémentaires de preuve de pathogénicité selon les recommandations de l'ACMG (Richards et al. 2015) et a entraîné un changement de classification de l'ACMG pour sept de ces 24 variants. Ces résultats démontrent l'utilité du test fonctionnel « minigène » pour le diagnostic de la calpaïnopathie et renforcent la nécessité d'intégrer les études fonctionnelles dans la démarche diagnostique à plus grande échelle. L'essai fonctionnel décrit ici peut être directement utilisé par les laboratoires de diagnostic pour tester les variants *CAPN3* avec un effet d'épissage suspecté, permettant d'établir un diagnostic parfois longtemps attendu pour de nombreux patients atteints de LGMDR1.

Nécessité d'un diagnostic rapide pour bénéficier de traitements potentiels

Pendant longtemps, le traitement des maladies neuromusculaires a été considéré comme purement symptomatique. Grâce aux nouvelles technologies de ces dernières années, de nouvelles formes de traitement ont pu être développées. Des thérapies géniques pour l'amyotrophie spinale, la dystrophie musculaire de Duchenne, certaines formes de dystrophie musculaire des ceintures, la myopathie myotubulaire et la neuropathie motrice et sensorielle héréditaire de type 1A sont actuellement en cours d'évaluation dans le cadre d'essais cliniques, ou même pour certaines en utilisation clinique. Par exemple, pour la dystrophie musculaire de Duchenne, les thérapies géniques encourageantes utilisent l'apport de mini/micro-dystrophine avec un vecteur AAV ou l'induction pharmacologique d'une translecture du codon stop prématuré chez les patients avec des variants pathogènes non-

sens. D'autre part, les approches de thérapie génique par saut d'exon, grâce à des oligonucléotides antisens qui se lient à l'ARN pré-messager afin d'exclure le ou les exons décalant le cadre de lecture a aussi montré des résultats satisfaisants (2019a). Le Nusinersen (ou Spinraza) est le premier traitement à avoir obtenu une autorisation de mise sur le marché pour l'amyotrophie spinale proximale liée à SMN1 en Europe et aux États-Unis. Un autre médicament de thérapie génique, le Zolgensma®, est autorisé aux États-Unis pour traiter les enfants âgés de moins de 2 ans atteints d'amyotrophie spinale liée à SMN1. Une demande d'autorisation de mise sur le marché (AMM) est en cours d'examen en Europe.

Pour pouvoir bénéficier d'un traitement par des molécules ayant une AMM ou pour être inclus dans les essais cliniques, le diagnostic moléculaire précis doit être connu chez le patient. Le diagnostic moléculaire précoce a donc aujourd'hui un véritable intérêt chez les patients ayant une affection neuromusculaire. Les efforts des centres de diagnostic dans le domaine des maladies neuromusculaires affiliés à FILNEMUS, et bientôt les résultats des séquençages des génomes dans le cadre du plan « France Médecine Génomique 2025 », pourront aider de nombreux patients à réduire l'errance diagnostique et accéder aux éventuels traitements plus rapidement.

Conclusion

Depuis son arrivée dans le domaine du diagnostic génétique il y a dix ans, la technologie de séquençage à haut débit a pris rapidement une place importante dans la pratique clinique. Cette approche permet d'effectuer l'analyse simultanée de la séquence de dizaines ou centaines de gènes d'intérêt (analyse par « panel de gènes »), des séquences codantes de tous les gènes du génome (approche « exome ») voire même de la totalité des séquences d'un génome humain. En permettant d'obtenir des informations de séquence pour de nombreux gènes en même temps, le séquençage à haut débit accélère grandement le processus de diagnostic des maladies génétiques par rapport à l'approche classique « gène-après gène » avec la technique de séquençage direct automatisé selon la méthode de Sanger. Les pathologies avec une grande variabilité phénotypique et génétique, tel que les maladies neuromusculaires, ont bénéficié largement de cette technologie. En effet, plus de 300 gènes associés à des pathologies neuromusculaires peuvent être analysés simultanément avec des délais de diagnostic plus courts, réduisant grandement l'errance diagnostic de patients atteints de ces maladies.

Aux avantages de l'augmentation considérable des quantités de données mutationnelles générées, ont été associés de nouveaux défis de traitement et d'interprétation des résultats générés par les tests de séquençage à haut débit. Tout d'abord, des efforts importants ont été déployés pour améliorer la qualité des données obtenues afin de satisfaire aux exigences du diagnostic moléculaire. Par exemple, ces nouvelles technologies nécessitent la mise en place de critères de qualité comportant notamment l'analyse de la « couverture » et de la « profondeur de lecture » de la séquence d'intérêt, qui doivent être pris en compte dans l'interprétation des résultats. Dans le cadre de cette thèse, nous avons développé et mis en place dans notre laboratoire un outil de visualisation de la couverture des panels diagnostics de gènes, CovReport. Ce logiciel permet d'évaluer d'un seul coup d'œil la performance du test de séquençage pour aider le prescripteur dans la démarche diagnostique. Par exemple, si une partie d'un gène susceptible d'être en cause sur la base de la présentation clinique n'est pas bien couvert, un test de re-séquençage supplémentaire peut être nécessaire pour s'assurer de l'absence d'un variant pathogène dans cette région.

Un autre enjeu majeur est la problématique de l'interprétation des données de séquence obtenues. Des recommandations nationales et internationales ont décrit des arguments utilisés pour évaluer l'impact clinique du variant, tel que le type de variant identifié, la présence de ce variant dans la population générale ou chez les patients, la prédiction bio-informatique de la pathogénicité, la concordance des données de ségrégation ou les analyses fonctionnelles de pathogénicité. Pour mettre à jour les recommandations françaises, nous avons effectué un travail de synthèse bibliographique prenant en compte les nombreuses publications récentes avec des précisions pour les critères initiaux et de recommandations gène-spécifiques. La nouvelle version des recommandations françaises est actuellement en cours de validation par les réseaux NGS-Diag et ANPGM avant d'être publiée et diffusée aux laboratoires de diagnostic génétique en France.

Enfin, devant le nombre de variants de signification inconnue identifiés par les tests de séquençage à haut débit, il y a une nécessité majeure d'établir des tests fonctionnels pour déterminer l'impact de ces variants sur le fonctionnement de la protéine ou sur l'épissage

d'ARN messenger. Nous avons mis en place une analyse d'épissage *in vitro* par l'analyse de « minigènes » pour la majorité des exons de *CAPN3*, le gène responsable d'une forme majeure des dystrophies musculaires des ceintures. Pour la majorité des variants ayant un effet sur l'épissage et détectés par ce test, la classification clinique des variants a été modifiée suite au résultat de ce test fonctionnel, démontrant ainsi l'importance de cette approche dans la démarche diagnostique.

En résumé, nos efforts dans la mise en place de la technologie de séquençage à haut débit à visée diagnostique, en coordination avec des initiatives nationales et internationales, permettent d'améliorer le diagnostic des maladies neuromusculaires et d'en diminuer l'errance diagnostique pour les patients atteints de ces maladies rares.

Bibliographie

- 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Abou Tayoun, A.N., Pesaran, T., DiStefano, M.T., Oza, A., Rehm, H.L., Biesecker, L.G., Harrison, S.M., and ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI) (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation* 39, 1517–1524.
- Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S.H., Kiemeny, L.A.L.M., Veltman, J.A., Gilissen, C., and Hoischen, A. (2017). Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *The American Journal of Human Genetics* 101, 50–64.
- Allot, A., Peng, Y., Wei, C.-H., Lee, K., Phan, L., and Lu, Z. (2018). LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. *Nucleic Acids Res.* 46, W530–W536.
- Amendola, L.M., Jarvik, G.P., Leo, M.C., McLaughlin, H.M., Akkari, Y., Amaral, M.D., Berg, J.S., Biswas, S., Bowling, K.M., Conlin, L.K., et al. (2016). Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *American Journal of Human Genetics* 98, 1067–1076.
- Anderson, L.V., Davison, K., Moss, J.A., Richard, I., Fardeau, M., Tomé, F.M., Hübner, C., Lasa, A., Colomer, J., and Beckmann, J.S. (1998). Characterization of monoclonal antibodies to calpain 3 and protein expression in muscle from patients with limb-girdle muscular dystrophy type 2A. *Am. J. Pathol.* 153, 1169–1179.
- BA1_Exception_List (2018). BA1 Exception List, July 2018.
- BA1_Exception_List_Nomination_Form BA1 Exception List Nomination Form.
- Barp, A., Laforet, P., Bello, L., Tasca, G., Vissing, J., Monforte, M., Ricci, E., Choumert, A., Stojkovic, T., Malfatti, E., et al. (2020). European muscle MRI study in limb girdle muscular dystrophy type R1/2A (LGMDR1/LGMD2A). *J. Neurol.* 267, 45–56.
- Bean, L.J.H., Funke, B., Carlston, C.M., Gannon, J.L., Kantarci, S., Krock, B.L., Zhang, S., Bayrak-Toydemir, P., and ACMG Laboratory Quality Assurance Committee (2019). Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.*
- Biesecker, L.G., Harrison, S.M., and ClinGen Sequence Variant Interpretation Working Group (2018). The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet. Med.* 20, 1687–1688.
- Birgmeier, J., Deisseroth, C.A., Hayward, L.E., Galhardo, L.M.T., Tierno, A.P., Jagadeesh, K.A., Stenson, P.D., Cooper, D.N., Bernstein, J.A., Haeussler, M., et al. (2020). AVADA: toward automated pathogenic variant evidence retrieval directly from the full-text literature. *Genet. Med.* 22, 362–370.
- Bonnet, C., Krieger, S., Vezain, M., Rousselin, A., Tournier, I., Martins, A., Berthet, P., Chevrier, A., Dugast, C., Layet, V., et al. (2008). Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *Journal of Medical Genetics* 45, 438–446.
- Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Medicine* 12, 3.
- Buermans, H.P.J., and den Dunnen, J.T. (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* 1842, 1932–1941.
- Cheng, J., Nguyen, T.Y.D., Cygan, K.J., Çelik, M.H., Fairbrother, W.G., Avsec, Ž., and Gagneur, J. (2019). MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* 20, 48.
- ClinGen_Lysosomal_Storage_Disorders_Expert_Panel (2019). ClinGen Lysosomal Storage Disorders Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1.

- ClinGen_TP53_Expert_Panel (2019). ClinGen TP53 Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1.
- Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics* 132, 1077–1130.
- Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., and Bérout, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37, e67.
- Desvignes, J.-P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Bérout, C., and Salgado, D. (2018). VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res.* 46, W545–W553.
- DiStefano, M.T., Hemphill, S.E., Oza, A.M., Siegert, R.K., Grant, A.R., Hughes, M.Y., Cushman, B.J., Azaiez, H., Booth, K.T., Chapin, A., et al. (2019). ClinGen expert clinical validity curation of 164 hearing loss gene–disease pairs. *Genetics in Medicine* 21, 2239–2247.
- Doig, K.D., Fellowes, A., Bell, A.H., Seleznev, A., Ma, D., Ellul, J., Li, J., Doyle, M.A., Thompson, E.R., Kumar, A., et al. (2017). PathOS: a decision support system for reporting high throughput sequencing of cancers in clinical diagnostic laboratories. *Genome Med* 9, 38.
- Ebbert, M.T.W., Jensen, T.D., Jansen-West, K., Sens, J.P., Reddy, J.S., Ridge, P.G., Kauwe, J.S.K., Belzil, V., Pregent, L., Carrasquillo, M.M., et al. (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 20, 97.
- Erikson, G.A., Bodian, D.L., Rueda, M., Molparia, B., Scott, E.R., Scott-Van Zeeland, A.A., Topol, S.E., Wineinger, N.E., Niederhuber, J.E., Topol, E.J., et al. (2016). Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* 165, 1002–1011.
- Fanin, M., Fulizio, L., Nascimbeni, A.C., Spinazzi, M., Piluso, G., Ventriglia, V.M., Ruzza, G., Siciliano, G., Trevisan, C.P., Politano, L., et al. (2004). Molecular diagnosis in LGMD2A: mutation analysis or protein testing? *Hum. Mutat.* 24, 52–62.
- Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frébourg, T., and Tosi, M. (2010). Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. In *Cancer Susceptibility*, M. Webb, ed. (Totowa, NJ: Humana Press), pp. 249–257.
- Gelb, B.D., Cavé, H., Dillon, M.W., Gripp, K.W., Lee, J.A., Mason-Suares, H., Rauen, K.A., Williams, B., Zenker, M., Vincent, L.M., et al. (2018). ClinGen’s RASopathy Expert Panel consensus methods for variant interpretation. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 20, 1334–1345.
- Ghosh, R., Oak, N., and Plon, S.E. (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* 18, 225.
- Ghosh, R., Harrison, S.M., Rehm, H.L., Plon, S.E., Biesecker, L.G., and ClinGen Sequence Variant Interpretation Working Group (2018). Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Human Mutation* 39, 1525–1530.
- Gorokhova, S., Biancalana, V., Lévy, N., Laporte, J., Bartoli, M., and Krahn, M. (2015a). Clinical massively parallel sequencing for the diagnosis of myopathies. *Rev. Neurol. (Paris)* 171, 558–571.
- Gorokhova, S., Cerino, M., Mathieu, Y., Courier, S., Desvignes, J.-P., Salgado, D., Bérout, C., Krahn, M., and Bartoli, M. (2015b). Comparing targeted exome and whole exome approaches for genetic diagnosis of neuromuscular disorders. *Appl Transl Genom* 7, 26–31.
- Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 47, D853–D858.
- Hosseini, S.M., Kim, R., Udupa, S., Costain, G., Jobling, R., Liston, E., Jamal, S.M., Szybowska, M., Morel, C.F., Bowdin, S., et al. (2018). Reappraisal of Reported Genes for Sudden Arrhythmic Death. *Circulation* 138, 1195–1205.
- Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., et al. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation* 33, 1228–1238.

- Hu, Z., Yu, C., Furutsuki, M., Andreoletti, G., Ly, M., Hoskins, R., Adhikari, A.N., and Brenner, S.E. (2019). VIPdb, a genetic Variant Impact Predictor Database. *Human Mutation* 40, 1202–1214.
- Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genetics* 6, e1001154.
- Ingles, J., Goldstein, J., Thaxton, C., Caleshu, C., Corty, E.W., Crowley, S.B., Dougherty, K., Harrison, S.M., McGlaughon, J., Milko, L.V., et al. (2019). Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ Genom Precis Med* 12, e002460.
- Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics* 99, 877–885.
- Jagadeesh, K.A., Paggi, J.M., Ye, J.S., Stenson, P.D., Cooper, D.N., Bernstein, J.A., and Bejerano, G. (2019). S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. *Nat. Genet.* 51, 755–763.
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535-548.e24.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes (*Genomics*).
- Kelly, M.A., Caleshu, C., Morales, A., Buchan, J., Wolf, Z., Harrison, S.M., Cook, S., Dillon, M.W., Garcia, J., Haverfield, E., et al. (2018). Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen’s Inherited Cardiomyopathy Expert Panel. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 20, 351–359.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Kergourlay, V., Raï, G., Blandin, G., Salgado, D., Bérout, C., Lévy, N., Krahn, M., and Bartoli, M. (2014a). Identification of splicing defects caused by mutations in the dysferlin gene. *Human Mutation* 35, 1532–1541.
- Kergourlay, V., Raï, G., Blandin, G., Salgado, D., Bérout, C., Lévy, N., Krahn, M., and Bartoli, M. (2014b). Identification of splicing defects caused by mutations in the dysferlin gene. *Hum. Mutat.* 35, 1532–1541.
- Krahn, M., and Collège national des enseignants et praticiens de génétique (2016). *Génétique médicale: enseignement thématique*.
- Krahn, M., Lévy, N., and Bartoli, M. (2016). Le séquençage de nouvelle génération (*Next-Generation Sequencing* , ou NGS) appliqué au diagnostic de maladies monogéniques hétérogènes: Notions essentielles pour le dialogue entre cliniciens et généticiens. *Les Cahiers de Myologie* 31–33.
- Krahn, M., Biancalana, V., Cerino, M., Perrin, A., Michel-Calemard, L., Nectoux, J., Leturcq, F., Bouchet-Séraphin, C., Acquaviva-Bourdain, C., Campana-Salort, E., et al. (2019). A National French consensus on gene lists for the diagnosis of myopathies using next-generation sequencing. *Eur. J. Hum. Genet.* 27, 349–352.
- Lee, K., Krempely, K., Roberts, M.E., Anderson, M.J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., et al. (2018a). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Human Mutation* 39, 1553–1568.
- Lee, K., Krempely, K., Roberts, M.E., Anderson, M.J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., et al. (2018b). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Human Mutation* 39, 1553–1568.
- Lu, Z.-X., Jiang, P., and Xing, Y. (2012). Genetic variation of pre-mRNA alternative splicing in human populations: Genetic variation of splicing. *Wiley Interdisciplinary Reviews: RNA* 3, 581–592.

- Luo, X., Feurstein, S., Mohan, S., Porter, C.C., Jackson, S.A., Keel, S., Chicka, M., Brown, A.L., Kesserwan, C., Agarwal, A., et al. (2019). ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Advances* 3, 2962–2979.
- Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., Race, V., Sistermans, E., Sturm, M., Weiss, M., et al. (2016). Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet* 24, 2–5.
- Mester, J.L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K.S., Costa, H.A., Lachlan, K., Ngeow, J., Barnholtz-Sloan, J., et al. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Human Mutation* 39, 1581–1592.
- Mort, M., Sterne-Weiler, T., Li, B., Ball, E.V., Cooper, D.N., Radivojac, P., Sanford, J.R., and Mooney, S.D. (2014). MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol.* 15, R19.
- Münz, M., Mahamdallie, S., Yost, S., Rimmer, A., Poyastro-Pearson, E., Strydom, A., Seal, S., Ruark, E., and Rahman, N. (2018). CoverView: a sequence quality evaluation tool for next generation sequencing data. *Wellcome Open Res* 3, 36.
- Nascimbeni, A.C., Fanin, M., Tasca, E., and Angelini, C. (2010). Transcriptional and translational effects of intronic CAPN3 gene mutations. *Hum. Mutat.* 31, E1658-1669.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276.
- Oza, A.M., DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K., Shen, J., Chapin, A., Boczek, N.J., Schimmenti, L.A., et al. (2018). Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Human Mutation* 39, 1593–1613.
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *The American Journal of Human Genetics* 102, 11–26.
- Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34, 867–868.
- Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47, 11.12.1-34.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* 372, 2235–2242.
- Richard, I., Roudaut, C., Saenz, A., Pogue, R., Grimbergen, J.E., Anderson, L.V., Beley, C., Cobo, A.M., de Diego, C., Eymard, B., et al. (1999). Calpainopathy—a survey of mutations and polymorphisms. *Am. J. Hum. Genet.* 64, 1524–1540.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 17, 405–424.
- Romanet, P., Odou, M.-F., North, M.-O., Saveanu, A., Coppin, L., Pasmant, E., Mohamed, A., Goudet, P., Borson-Chazot, F., Calender, A., et al. (2019). Proposition of adjustments to the ACMG-AMP framework for the interpretation of MEN1 missense variants. *Human Mutation* 40, 661–674.
- Rowlands, C.F., Baralle, D., and Ellingford, J.M. (2019). Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells* 8, 1513.
- Savisaar, R., and Hurst, L.D. (2017). Estimating the prevalence of functional exonic splice regulatory information. *Hum. Genet.* 136, 1059–1078.

Scott, E.M., Halees, A., Itan, Y., Spencer, E.G., He, Y., Azab, M.A., Gabriel, S.B., Belkadi, A., Boisson, B., Abel, L., et al. (2016). Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat. Genet.* **48**, 1071–1076.

Seifert, B.A., McGlaughon, J.L., Jackson, S.A., Ritter, D.I., Roberts, M.E., Schmidt, R.J., Thompson, B.A., Jimenez, S., Trapp, M., Lee, K., et al. (2019). Determining the clinical validity of hereditary colorectal cancer and polyposis susceptibility genes using the Clinical Genome Resource Clinical Validity Framework. *Genetics in Medicine* **21**, 1507–1516.

Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., Seifert, B.A., Sneddon, T.P., et al. (2017a). Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895–906.

Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., Seifert, B.A., Sneddon, T.P., et al. (2017b). Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895–906.

SVI_Recommendation_for_PM3_Criterion (2019). SVI Recommendation for in trans Criterion (PM3) - Version 1.0.

SVI_Recommendation_for_PS2_and_PM6_criteria (2018). PS2/PM6: Recommendation for de novo PS2 and PM6 ACMG/AMP criteria (Version 1.0).

Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program (Genomics).

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034.

the NDD Exome Scoping Review Work Group, Srivastava, S., Love-Nichols, J.A., Dies, K.A., Ledbetter, D.H., Martin, C.L., Chung, W.K., Firth, H.V., Frazier, T., Hansen, R.L., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genetics in Medicine* **21**, 2413–2421.

Théry, J.C., Krieger, S., Gaildrat, P., Révillion, F., Buisine, M.-P., Killian, A., Duponchel, C., Rousselin, A., Vaur, D., Peyrat, J.-P., et al. (2011). Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur. J. Hum. Genet.* **19**, 1052–1058.

TOPMed (2018). The NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. BRAVO variant browser: University of Michigan and NHLBI; 2018. Available from: <https://bravo.sph.umich.edu/freeze5/hg38/>.

Tournier, I., Vezain, M., Martins, A., Charbonnier, F., Baert-Desurmont, S., Olschwang, S., Wang, Q., Buisine, M.P., Soret, J., Tazi, J., et al. (2008). A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum. Mutat.* **29**, 1412–1424.

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A.H., Karczewski, K., Ing, A.Y., Barton, P.J.R., Funke, B., Cook, S.A., MacArthur, D., et al. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med.* **19**, 1151–1158.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394.

Zastrow, D.B., Baudet, H., Shen, W., Thomas, A., Si, Y., Weaver, M.A., Lager, A.M., Liu, J., Mangels, R., Dwight, S.S., et al. (2018). Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): The ClinGen IEM Working Group and the Phenylalanine Hydroxylase Gene. *Human Mutation* **39**, 1569–1580.

Ziegler, A., Colin, E., Goudenège, D., and Bonneau, D. (2019). A snapshot of some pLI score pitfalls. *Human Mutation*.

(2019). Avancées dans Dystrophies Musculaires de Duchenne et Becker, Savoir et Comprendre, Avancées de la recherche.

Annexes

Annexe 1: Recommandations françaises pour l'interprétation de variations nucléotidiques générées par les analyses en séquençage haut-débit, Version 2



Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS

Date de création : 6/3/2020

Date de révision :

Version : 2

Date de 1^{ère} Application : en atteinte

N° document : NGSDIAG_001

Approbation par le board du réseau : en cours

Ce document a été préparé par les membres du Groupe de Travail du Réseau NGS-Diag listé cidessous :

- Jean MULLER (Strasbourg ; représentant Réseau NGS-Diag)
- Cécile ROUZIER (Nice; représentant Réseau NGS-Diag)
- Nicolas CHATRON (Lyon ; représentant Réseau AChro-Puce)
- Boris KEREN (La Pitié ; représentant Réseau AChro-Puce)
- Caroline ROORYCK-THAMBO (Bordeaux ; représentant Réseau AChro-Puce)
- Martine DOCO FENZY (Reims ; représentant ACLF)
- Sandrine CAPUTO (Paris Curie ; représentant GGC)
- Nicolas DERIVE (Paris Curie ; représentant GGC)
- Claude HOUDAYER (Curie ; représentant GGC)
- Sylvie BANNWARTH (Nice ; ANPGM, mitochondrie)
- Anne BARLIER (Marseille ; ANPGM, FNMR FIRENDO)
- Svetlana GOROKHOVA (Marseille ; ANPGM)
- Claude JARDEL (La Pitié ; ANPGM, mitochondrie)
- Cédric LE MARECHAL (Brest ; ANPGM, FNMR SENSGENE)
- Gaël NICOLAS (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience)
- Pascale RICHARD (Paris La Pitié ; ANPGM, FNMR CARDIOGEN)

Coordination :

Martin KRAHN (Marseille; ANPGM, FNMR FILNEMUS)

Pascale SAUGIER-VEBER (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience),

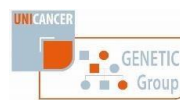
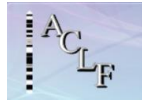


Table des matières

| | |
|--|------------------------------------|
| Contexte de l'initiative | - 80 - |
| 1 - Liste des arguments à prendre en compte pour l'interprétation des variants..... | - 83 - |
| 1-1- Données épidémiologiques | Erreur ! Signet non défini. |
| 1-1-1 Données de fréquence allélique dans les populations contrôles..... | Erreur ! Signet non défini. |
| 1-1-2 Données de fréquence allélique chez les sujets atteints | Erreur ! Signet non défini. |
| 1-1-3 Base de données locale | Erreur ! Signet non défini. |
| 1-4- Critères cliniques et généalogiques | Erreur ! Signet non défini. |
| 1-5- Données alléliques | Erreur ! Signet non défini. |
| 1-6- Données de ségrégation | Erreur ! Signet non défini. |
| 1-2- Données prédictifs | Erreur ! Signet non défini. |
| 1-2-1 Nature du variant nucléotidique | Erreur ! Signet non défini. |
| 1-2-2- Prédiction bio-informatiques | Erreur ! Signet non défini. |
| 1-3- Données bibliographiques | Erreur ! Signet non défini. |
| 1-7- Données fonctionnelles | Erreur ! Signet non défini. |
| 1-8- Analyses somatiques | Erreur ! Signet non défini. |
| 1-9- Modifications de l'argumentaire par rapport aux recommandations | Erreur ! Signet non défini. |
| 2 - Classification du variant | - 97 - |
| 3 - Utilisation dans le cadre d'un compte-rendu d'analyses | - 98 - |
| 4 - Bibliographie et Annexes..... | - 98 - |

Abréviations utilisées dans ce document :

ACLF: Association des Cytogénétiens de Langue Française
ACMG-AMP : American College of Medical Genetics & the Association for Molecular Pathology
AnDDi-Rares : Filière de Santé Anomalies du Développement et Déficience intellectuelle de Causes Rares
ANPGM : Association Nationale des Praticiens de Génétique Moléculaire
CAD : Collecteur Analyseur de Données
CARDIOGEN : Filière de Santé des Maladies Cardiaques Héritaires ou Rares
CNV : Copy Number Variations
CREFIX : Centre de référence, d'innovation, d'expertise et de transfert
DefiScience : Filière de Santé Maladies Rares du Développement Cérébral et de la Déficience Intellectuelle
FILNEMUS : Filière de Santé des Maladies Rares Neuromusculaires
FIRENDO : Filière Maladies Rares Endocriniennes
FNMR : Filières de soin Nationales de Maladies Rares
GGC : Groupe Génétique et Cancer
HI index : Haploinsufficiency Score
LOEUF : loss-of-function observed/expected upper bound fraction
MAF : Minor Allele Frequency
NGS : Next Generation Sequencing
o/e : observed/expected
PFMG2025 : Plan France Médecine Génomique 2025
pLI: probability of Loss of function intolerance
SENSGENE : Filière de Maladies Rares Sensorielles
SV : Structural variant (englobant délétions, duplications, translocations, insertions, inversions...)

Contexte de l'initiative

La mise en place du NGS (Next Generation Sequencing) à visée diagnostique a confronté les laboratoires de diagnostic génétique à une augmentation des problématiques de l'interprétation des variants de séquence, en particulier pour des gènes dont les laboratoires respectifs n'avaient

pas d'expertise spécifique. Sur le plan international, les recommandations émises par l'ACMG-AMP (*American College of Medical Genetics & the Association for Molecular Pathology*) (Amendola et al., 2016; Richards et al., 2015) sont aujourd'hui largement utilisées par les laboratoires de diagnostic génétique, y compris en France. Néanmoins, nous n'avons pas à ce jour en France de consensus national des laboratoires concernés.

Puisque l'interprétation des variants constitue une étape essentielle et critique pour l'homogénéisation inter-laboratoire des rendus de résultats, le Réseau NGS-DI

AG a souhaité initier un Groupe de Travail spécifique dans l'objectif d'aboutir à un consensus national, dans le contexte actuel de la mise en place prochaine des premières plateformes PFMG2025 (Plan France Médecine Génomique).

Cette problématique a été abordée en particulier dans le cadre des commissions de diagnostic génétique des FNMR (Filières de soin Nationales de Maladies Rares) AnDDi-Rares (Filière de Santé Anomalies du Développement et Déficience intellectuelle de Causes Rares), DefiScience (Filière de Santé Maladies Rares du Développement Cérébral et de la Déficience Intellectuelle) et FILNEMUS (Filière de Santé des Maladies Rares Neuromusculaires), où des travaux d'homogénéisation de l'interprétation des variants de séquence ont été débutés, et seront disponibles comme base de travail. **Les recommandations ACMG-AMP étant déjà largement appliquées par de nombreux laboratoires de diagnostic génétique en France, l'objectif qui nous semble à la fois important et réalisable à court terme serait une adoption officielle, au nom du Réseau NGS-Diag, de celles-ci, pour une application consensuelle sur le plan national.**

Pour cette raison, un groupe de travail représentatif des différentes parties constituantes du Réseau NGS-Diag a été mis en place pour procéder selon la démarche suivante :

1. Accord du GT de prendre comme base les recommandations de l'ACMG ;
2. Elaboration du présent document de recommandations en français, issu d'un document très complet déjà réalisé par l'équipe de Rouen ;
3. Adoption par le Réseau NGS-DIAG du document de recommandations en français, et contact/lien à établir avec l'ACMG (notamment pour suivre les évolutions des recommandations de l'ACMG, et d'être associé à des versants d'évolution de ces recommandations). A ce stade aussi, des interactions avec le CREFIX (Centre de référence, d'innovation, d'expertise et de transfert) et le CAD (Collecteur Analyseur de Données) pourront être établies.
4. Relai des guidelines au sein des FNMR/réseaux d'oncogénétique, et travaux spécifiques éventuels au niveau des FNMR/réseaux d'oncogénétique (échange de fichiers vcf, spécifications si nécessaire de certains items de l'interprétation, ...).

Ce document de recommandations générales a vocation à être complété par d'autres recommandations spécifiques selon les thématiques. Il peut s'agir d'ajouter des arguments d'interprétation pour les métaboliciens (poids des dosages enzymatiques), les oncogénéticiens (poids des analyses somatiques), les cardiogénéticiens (poids de l'ECG), etc... Ces recommandations spécifiques seront à discuter au sein de chaque FNMR/réseaux d'oncogénétique de manière à harmoniser le poids accordé à ces éléments spécifiques dans la classification des variants.

La 2^{ème} version de ce document prend en compte les modifications des recommandations ACMG publiées depuis la sortie de la première version de ce document (NGSDIAG_001). Il s'agit des précisions générales sur les critères de classification (PVS1, PS2/PM6, PS3/BS3, PM3, PP5/BP6 ou BA1) (Abou Tayoun et al., 2018; Biesecker et al., 2018; Brnich et al., 2019; Ghosh et al., 2018; SVI_Recommendation_for_PM3_Criterion, 2019; SVI_Recommendation_for_PS2_and_PM6_criteria, 2018) et les recommandations gène-spécifiques élaborées par des Groupes d'experts du réseau

ClinGen ou autres experts dans le domaine (MYH7, Rasopathies, PAH, Surdités, CDH1, PTEN, MEN1, RUNX1, GAA, TP53)(Gelb et al., 2018; Kelly et al., 2018; Lee et al., 2018a; Luo et al., 2019; Mester et al., 2018; Oza et al., 2018; Romanet et al., 2019; Zastrow et al., 2018)
(ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019; ClinGen_TP53_Expert_Panel, 2019)

Un autre point important qui n'était pas abordé dans la première version de ce document est le choix des gènes sur les panels diagnostiques. L'objectif d'un panel de gènes à visée diagnostique est de maximiser la sensibilité clinique et de minimiser les explorations inutiles ainsi que le stress secondaires à la découverte des variants dont la signification clinique est incertaine. Les recommandations récentes de American College of Medical Genetics and Genomics (ACMG) visent à clarifier quels gènes sont adaptés pour être inclus dans les panels diagnostiques (Bean et al., 2019). Les gènes sont définis comme GAD (Gene associated with Mendelian disorder) ou GUS (Gene of uncertain significance), les gènes pour lesquels il n'y a pas encore assez de preuves pour établir un lien gène-pathologie. Si l'indication est de confirmer un diagnostic clinique, le panel de gènes utilisé doit contenir que des gènes pour lesquels le lien gène-maladie est clairement établi (GAD, Gene associated with Mendelian disorder). Si l'indication est le diagnostic des cas cliniquement complexes, les GUS (Gene of uncertain significance) peuvent être inclus. Dans ce cas, le consentement type exome/génome est requis. De plus, dans le compte rendu, il est nécessaire de clairement séparer les variants identifiés dans les GUS des variants identifiés dans les GAD. La classification maximale attribué à un variant identifié dans un gène sans lien gène-pathologie établie (GUS) ne peut pas être plus qu'une classification VSI (variant d'une signification incertaine).

Recommandations pour l'interprétation de variants générés par le séquençage NGS

Domaine d'application :

Ce document ne s'applique qu'aux maladies à transmission mendélienne.

Dans sa version actuelle, il n'intègre pas les recommandations portant sur les variants de type SV (dont les CNV), et sur les variants de l'ADN mitochondrial.

L'interprétation est réalisée dans l'état actuel des connaissances et des informations disponibles.

Avant d'interpréter un variant dans un gène donné, il est recommandé de recueillir des informations concernant les items suivants :

- La pathologie moléculaire de ce gène : spectre mutationnel des variants pathogènes rapportés pour le gène considéré (HGMD-Pro, ClinVar, LOVD ou autres bases, bibliographie), o/e ou Z-score de contrainte protéique dans gnomAD, sensibilité à l'haplo-insuffisance (par exemple o/e ou pLI dans gnomAD)
- Les domaines fonctionnels de la protéine (bibliographie gène spécifique, UniProt, GeneCards,...)
- Les épissages physiologiques connus
- Le mode de transmission de la ou des maladies mendéliennes associées à ce gène
- Les données de pénétrance s'il y en a
- Les bases de données (locus-spécifiques et populations-contrôles)
- Les équipes référentes (biologistes et cliniciens) s'il y en a
- Les tests fonctionnels validés s'il y en a

1 - Liste des arguments à prendre en compte pour l'interprétation des variants

L'interprétation des résultats repose sur un faisceau d'arguments. Ces arguments ont un poids plus ou moins important dans l'interprétation du variant mis en évidence :

PVS/PS/PM/PP : Argument très fort (Pathogenic Very Strong) / fort (Pathogenic Strong) / moyen (Pathogenic Moderate) / faible (Pathogenic Poor) en faveur de la pathogénicité du variant.

BA/BS/BP : Argument suffisant (Benign stand Alone)/fort (Benign Strong) /faible (Benign Poor) en faveur du caractère bénin du variant.

L'interprétation des résultats est sous la responsabilité exclusive du biologiste et consiste à combiner ces arguments pondérés afin d'assigner une des 5 classes suivantes au variant étudié :

- **Classe 1 : Variant bénin**

- **Classe 2 : Variant probablement bénin**

- **Classe 3 : Variant de signification inconnue**

- **Classe 4 : Variant probablement pathogène**

- **Classe 5 : Variant pathogène**

Lorsqu'un variant est assigné en classe 3, cela ne permet ni de l'utiliser dans la prise en charge éventuelle du patient et de sa famille, ni d'exclure son caractère causal dans la pathologie du patient. Le variant peut faire l'objet d'études complémentaires visant à apporter des arguments permettant de le re-classer en variant d'une autre classe (études fonctionnelles *in vitro*, étude de l'épissage, étude de ségrégation familiale). Le dossier peut aussi être repris ultérieurement afin de

ré-examiner l'interprétation du variant de signification inconnue à la lumière des nouvelles connaissances acquises le concernant.

L'argumentaire pondéré détaillé ci-dessous correspond aux recommandations internationales publiées conjointement par l'ACMG (*American College of Medical Genetics*) et l'AMP (*Association for Molecular Pathology*) en 2015 et 2016 (Amendola et al., 2016; Richards et al., 2015), incorporant les modifications et les mises à jour récentes. Il a pour objectif de limiter la variabilité inter- et intra-laboratoire dans l'interprétation des variants en génétique moléculaire. Afin d'homogénéiser l'utilisation des items de la liste d'arguments, il apparaît nécessaire d'apporter certaines précisions comme la valeur seuil de fréquences alléliques, l'utilisation de scores combinés de prédiction de pathogénicité de différents algorithmes, l'utilisation d'un score unique issu d'un algorithme sélectionné consensuel, etc...

L'interprétation finale des variants reste à l'appréciation du biologiste en charge du dossier, éventuellement en aval d'une réunion de concertation pluridisciplinaire. L'interprétation est réalisée dans l'état actuel des connaissances et des informations disponibles. Certains points des recommandations, détaillés à la fin du chapitre, ont été modifiés pour s'adapter aux positions françaises.

L'argumentaire prend en compte :

- Des données épidémiologiques
- Des données structurales
- Des données bibliographiques
- Des critères cliniques ou généalogiques
- L'existence de variants associés
- Des données de ségrégation
- Des données fonctionnelles
- Des données somatiques

Ces éléments d'argumentation sont récapitulés dans l'annexe 1. Il est important de souligner que, pour chaque argument, il ne faut sélectionner qu'un seul critère. Ainsi, dans les données épidémiologiques, choisir BS1 ou BS2 ou PM2.

1-1- Données épidémiologiques

Un élément crucial pour l'interprétation est l'analyse des données épidémiologiques. Le généticien sera conduit à consulter les bases de données de populations contrôles, les bases de données de patients mais également la base de données locale qui collecte tous les résultats du laboratoire.

a) 1-1-1 Données de fréquence allélique dans les populations contrôles

Il est nécessaire de consulter les bases de données de populations afin de déterminer la fréquence allélique du variant mis en évidence dans la population générale. Une fréquence allélique importante est en faveur du caractère bénin du variant. De manière générale, on considère qu'un variant dans un gène responsable de maladie de transmission autosomique dominante ne peut pas avoir une fréquence allélique supérieure à la prévalence de la maladie.

Les principales bases de données de populations consultées sont :

- **gnomAD (*Genome Aggregation Database*) :**

GnomAD consiste en une agrégation de données génomiques d'individus non apparentés issus de diverses études populationnelles ou spécifiques de certaines maladies (Karczewski et al., 2019). Deux versions de gnomAD sont disponibles actuellement : v.2.1.1 et v.3. La version 2.1.1 contient les données des 125,748 exomes et de 15,708 génomes alignés sur le génome de référence GRCh37. La version 3 contient les variants de 71 702 génomes alignés sur le génome de référence GRCh38. Consultation de gnomAD.v3 est importante car elle contient les données des

populations qui jusque-là n'étaient pas bien représentées dans les bases de données. La base exclue les individus avec des maladies pédiatriques sévères et leurs apparentés au premier degré. En revanche, il est possible qu'elle contienne les données génomiques d'individus avec des maladies Mendéliennes ou complexes, comme des cancers, des maladies psychiatriques, des démences, ou des maladies inflammatoires. Le nom des séries incluses est consultable sur le site internet mais l'information sur la présence de malades est précisée d'une manière globale. Il est possible de visualiser que les variants issues des cohortes des individus sains en sélectionnant « controls » dans l'onglet « Dataset » (avec un désavantage d'avoir que 60K échantillons au lieu de 141K). Il est conseillé de sélectionner les données « non-neuro » pour une analyse des variants dans les maladies neurologique ou « non-cancero » pour les analyses dans le domaine des cancers.

GnomAD peut contenir des variants pathogènes présents en état mosaïque chez les sujets asymptomatiques. Dans ces cas, il est conseillé d'évaluer la possibilité d'un mosaïcisme avec l'aide de la fenêtre IGV sur la page du variant. Certains variants pathogènes peuvent aussi être présent en mosaïque que dans la lignée sanguine des sujets sains due au phénomène d'hématopoïèse clonale (Acuna-Hidalgo et al., 2017). Etant donné que cette situation est plus probable chez les sujets âgés, il est recommandé de consulter l'âge des porteurs du variant étudié dans gnomAD.

<https://gnomad.broadinstitute.org/>

- **BRAVO, projet TOPMed (*Trans-Omics for Precision Medicine*)**

Cette base de données contient les positions génomiques, les prédictions fonctionnelles et les fréquences alléliques pour les variants issue de génomes de 62784 individus du projet TOPMed (Trans-Omics for Precision Medicine) centré sur les maladies cardiaques, pulmonaires, sanguins et du sommeil (Taliun et al., 2019; TOPMed, 2018)

<https://bravo.sph.umich.edu/freeze5/hg38/>

- **FLOSSIES (*Fabulous Ladies Over Seventy*) :**

Cette base collecte les variants identifiés dans une population contrôle de 10.000 femmes de plus de 70 ans pour 27 gènes impliqués (ou possiblement impliqués) dans les prédispositions au cancer du sein.

<https://whi.color.com/about>

- **FREX (*French Exome Project*) :**

Cette base comporte les données de séquençage exome entier issus de 600 individus français de six régions françaises.

<http://med-laennec.univ-brest.fr/FrExAC/>

- **The Greater Middle East (GME) Variome Project :**

Cette base collecte les variants issus du séquençage d'exome entier d'une population de 2497 individus originaires du Nord de l'Afrique et du Moyen-Orient (séquençage effectué dans le cadre de différents projets portant sur des maladies mendéliennes et infectieuses) (Scott et al., 2016).

<http://igm.ucsd.edu/gme/>

Attention : Cette base de données comprenant à la fois des individus sains et des patients atteints de maladies mendéliennes, une grande prudence doit être accordée à son utilisation.

- **Welllderly :**

The Welllderly cohort (considérée à 95% européenne) apporte des données de génomes entiers de 511 individus séquencés avec la technologie Complete Genomics (Erikson et al., Cell 2016), et 163 individus séquencés sur plateforme Illumina) (Erikson et al., 2016). Ces individus sont considérés comme étant en bonne santé et âgés de plus de 80 ans à l'inclusion.

Arguments :

- Fréquence allélique > 5% = critère à lui seul en faveur du caractère bénin du variant (**BA1**). A noter, ce seuil a été baissé pour un certain nombre de gènes et pathologies (voir Tableau 1).
- Fréquence allélique trop importante par rapport à la fréquence de la pathologie (**BS1**)

Attention : pour les maladies de transmission autosomique récessive, la fréquence allélique de certains variants pathogènes fréquents peut être >1%. Un calcul théorique selon la loi de Hardy-Weinberg peut être utilisé pour ajuster le seuil en fonction du gène et des données de fréquences alléliques correspondantes disponibles. Le calcul utilisant le cadre statistique pour filtrage de variants candidats basée sur la fréquence peut être réalisé avec l'aide de <http://cardiodb.org/allelefrequencyapp/> (Whiffin et al., 2017)

Attention : il est important de veiller à l'existence d'un minimum de 1000 allèles dans la souspopulation contrôle considérée.

- Variant absent des bases de données de populations contrôles (ou très faible fréquence si récessif) (**PM2**)

Attention : pour les maladies fréquentes ou de révélation tardive de transmission dominante, certains variants pathogènes peuvent être présents à de faibles fréquences dans les bases de populations contrôles.

Attention : certaines bases de données comprennent des individus malades ou des individus à risque d'une pathologie donnée (par exemple pathologies cardio-vasculaires).

Attention : l'absence d'un variant dans une base de données doit être associée à la vérification de la profondeur moyenne au niveau de la base considérée, et l'évaluation de la qualité de séquençage. Cette dernière peut être approchée par le nombre de génotypes disponibles à la position génomique considérée ou au niveau des positions adjacentes (dénominateur variable de calcul des fréquences alléliques).

Précautions :

- Les indels (en particulier de grande taille) peuvent être mal détectés dans la population générale par séquençage haut débit.
- Ces bases de données sont internationales et ne sont donc pas représentatives des variants spécifiques de sous-populations (régionales ou autres). Il est important de connaître l'origine ethnique du patient pour pouvoir estimer si la sous-population d'où le patient est originaire est représentée dans gnomAD. Si ce n'est pas le cas, l'argument PM2 doit être interprété avec une caution
- Il existe une redondance de données entre certaines de ces bases de données, notamment TopMED et GnomAD
- Certains variants pathogènes reconnus ont une Fréquence allélique >1%. Les listes de ce type de variants sont déjà été élaborés pour certains gènes par les groupes de travail gène-spécifique (BA1_Exception_List, 2018; Ghosh et al., 2018; Oza et al., 2018). D'autres variants

pathogènes avec des fréquences allélique très élevées peuvent être proposé à ClinGen (BA1_Exception_List_Nomination_Form).

- Dans la publication de l'ACMG-AMP, les fréquences alléliques des variants sont appelées probablement par abus de langage MAF (Minor Allele Frequency). Le terme MAF désigne la fréquence allélique de l'allèle mineur. Selon la définition utilisée par la base dbSNP, archive internationale des petits variants et notamment des polymorphismes du génome humain, (https://www.ncbi.nlm.nih.gov/projects/SNP/docs/rs_attributes.html) cette MAF correspond à la fréquence du second allèle le plus fréquent. Ainsi, si à une position génomique donnée 3 variants sont décrites avec les fréquences suivantes 0,50/0,49/0,01 la MAF de cette position sera 0,49. La MAF est donc la même pour tous les variants à cette position. Cette donnée doit ainsi être utilisée avec précaution car elle ne correspond pas nécessairement à la fréquence du variant en cours d'étude. Il semble judicieux d'utiliser plutôt la véritable fréquence allélique du variant à la place de la MAF. C'est la raison pour laquelle nous utilisons la fréquence allélique dans les présentes recommandations.
- Un plus de la fréquence globale dans une base de données comme gnomAD, il est nécessaire d'évaluer la fréquence allélique dans les sous-populations. Si le variant à classé est retrouvé avec une fréquence très élevée dans une sous-population (sans l'effet fondateur connu), l'argument BS1 peut être attribué même si la fréquence allélique de ce variant est très basse dans les autres sous-populations

Les seuils de fréquence pour attribuer les arguments PM2, BS1 ou BA1 doivent être définies en fonction du gène et de la pathologie. Depuis la publication des Recommandations ACMG de 2015 ainsi que la première version de ce document, dix groupes de travaux ont proposé des seuils précis pour ces critères dans le cadre des recommandations gène-spécifiques. Le résumé de ces données est présenté dans le Tableau 1 :

| | Pathologies Dominantes | | | | | | | | Pathologies Récessives | | |
|----------------|------------------------|------------|----------|-------------|------------------------|------------|----------|---------|------------------------|--------------|--------|
| | MYH7 | Rasopathie | PTEN | Surdités AD | CDH1 | MEN1 | RUNX1 | TP53 | Surdités AR | PAH | GAA |
| PM2 | < 0,004% | 0 | < 0,001% | < 0,002% | < 0,001% (ou < 0,002%) | 0-1 | 0 | | < 0,007% | < 0,02% | < 0,1% |
| PM2_sup | | | | | | | | 0 | < 0,07% | | |
| BS1 | > 0,02% | > 0,025% | > 0,1% | > 0,02% | > 0,1 % | > 0,00784% | > 0,015% | > 0,03% | > 0,3% | > 0,2 - 1,5% | > 0,5% |
| BS1_sup | | | | | | | | | > 0,07% | | |
| BA1 | > 0,1% | > 0,05% | > 1% | > 0,1% | > 0,2% | > 0,76% | > 0,15% | > 0,1% | > 0,5% | > 1,5% | > 1% |

Tableau 1 : Seuils de fréquences alléliques du variant à classer pour attribuer PM2, BS1 ou BA1.

Un résumé des recommandations issue des panels d'experts pour les gène ou pathologies suivants : MYH7 (Kelly et al., 2018), Rasopathie (Gelb et al., 2018), PTEN (Mester et al., 2018), Surdités (Oza et al., 2018), CDH1 (Lee et al., 2018a), MEN1 (Romanet et al., 2019), RUNX1 (Luo et al., 2019), TP53 (ClinGen_TP53_Expert_Panel, 2019), PAH (Zastrow et al., 2018), GAA (ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019).

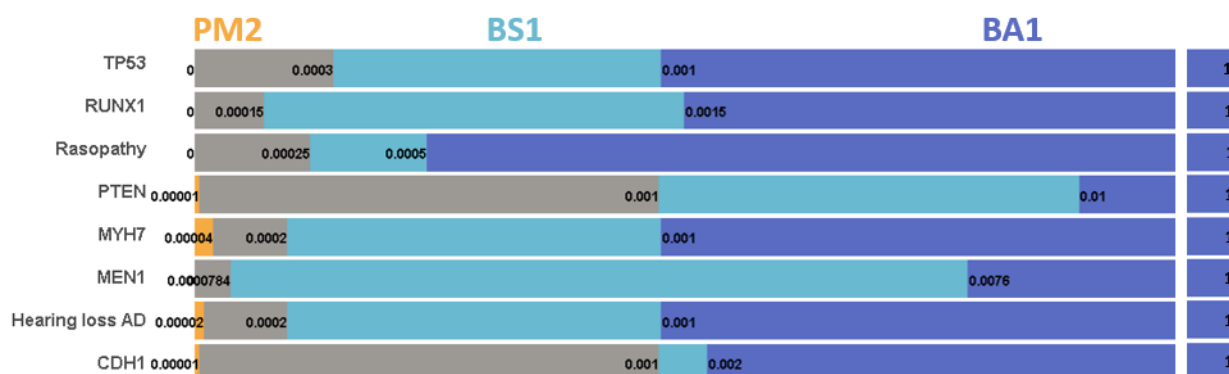


Figure 1 : Un résumé graphique des seuils des fréquences alléliques pour attribuer PM2/BS1/BA1 selon les recommandations gène-spécifique (pathologies AD)

1-1-2 Données de fréquence allélique chez les sujets atteints

Il est nécessaire de consulter les bases de données de patients comme ClinVar, LOVD (bases spécifiques de gènes), COSMIC. En fonction des pathologies, il peut être important de consulter les bases nationales concernées (Ex : Bases de données MLH1/MSH2/MSH6 et BRCA1/BRCA2/PALB2). Dans la mesure du possible, une base de données locus-spécifique (si possible avec consensus par thématique) devra être consultée.

La diversité des bases consultées et leur fiabilité est à prendre en compte.

Il est également important de consulter nos propres résultats concernant ce variant dans la base de données du laboratoire. Cela permet de s'appuyer sur la description de ce variant chez d'autres patients pour affiner l'interprétation. L'objectif est que l'interprétation d'un variant soit la même pour tous les patients porteurs de ce variant.

Il est possible de contacter les laboratoires ayant une expertise sur la thématique.

Arguments :

- Prévalence du variant chez les individus atteints significativement supérieure à celle des contrôles de la même origine ethnique que les patients, Odd Ratio > 3 (test exact de Fisher) (**PS4**)

En absence des études cas-contrôles, il est possible d'attribuer PS4 en fonction de nombre de patients confirmés avec ce variant. Le nombre de patients nécessaires doit être défini en fonction du gène et la pathologie étudié. Pour compter le nombre de patients porteurs du variants, il est nécessaire de consulter l'ensemble des bases données de patients (ClinVar, LOVD, bases spécifique) et faire une recherche bibliographique extensive.

Huit recommandations gène-spécifiques récentes propose des seuils pour l'attribution de l'argument PS4 (Tableau 2). Cette façon d'attribuer PS4 est possible que si le variant à classé est absent ou extrêmement rare dans la population générale (PM2 attribué). Le variant à classé lui-même ne compte pas dans le score PS4.

Il était proposé de calculer PS4 de cette façon que pour les gènes responsables des pathologies dominantes et d'utilisant plutôt PM3 pour les pathologies récessives (Oza et al., 2018). Comme la majorité de recommandations gène-spécifiques ont été établies pour des gènes et pathologies dominantes, plus de réflexion est nécessaire pour mieux adapter ces critères pour les gènes récessifs.

| | MYH7 | Rasopathie | PTEN | Surdités AD | CDH1 | MEN1 | RUNX1 | TP53 |
|----------------|------|------------|------|-------------|------|---------------------|-------|------|
| PS4_VeryStrong | | | > 16 | | > 16 | | | |
| PS4_Strong | > 15 | > 5 | 4-15 | > 15 | 4-15 | case-control needed | > 4 | > 4 |
| PS4_Moderate | 6-14 | 3-4 | 2-3 | 6-14 | 2-3 | 3 | 2-3 | 2-3 |
| PS4_Supporting | 2-5 | 1-2 | 1 | 2-5 | 1 | 2 | 1 | 1 |

Tableau 2. Seuils de nombre de patients confirmés porteurs de variant pour attribuer PS4.

Un résumé des recommandations pour huit pathologies ou groupes des pathologies dominantes : MYH7 (Kelly et al.,

2018), Rasopathie (Gelb et al., 2018), PTEN (Mester et al., 2018), Surdités (Oza et al., 2018), CDH1 (Lee et al., 2018a), MEN1 (Romanet et al., 2019), RUNX1(Luo et al., 2019), TP53 (ClinGen_TP53_Expert_Panel, 2019). Selon les recommandations pour PTEN et TP53, les évaluations phénotypiques sont aussi calculés en parallèle, avec nécessité d'avoir plus de patients pour un niveau de PS4 si les phénotypes des patients sont moins surs.

Suite la recommandation de ClinGen (Biesecker et al., 2018), les critères PP5 et BP6 ont été retirés de la grande majorité des recommandations gène-spécifiques. En effet, la présence ou absence d'un variant dans les bases de données de patients est déjà évaluée pendant l'attribution de PS4.

- Variant à l'origine du même changement d'acide aminé qu'un variant pathogène connu (**PS1**) sauf si le variant n'est pas un variant faux-sens mais a un effet sur l'épissage

- Variant à l'origine d'un changement d'acide aminé différent à la même position qu'un variant fauxsens pathogène connu (**PM5**)

Précautions :

- Il est possible que le même patient est décrit dans plusieurs sources (e.g. PubMed et ClinVar ou LOVD). Il est indispensable de vérifier que le même patient n'est pas compté plusieurs fois, entraînant le score PS4 faussement élevé
- Attention aux articles trop anciens ne bénéficiant pas d'une population contrôle ou d'une population de patients de taille suffisante et/ou d'ethnies suffisamment variées pour s'affranchir des biais possibles.
- Il convient de rester prudent avec la classification des variants issus de ces bases de données puisque ces bases intègrent des variants issus d'articles parfois anciens avec une classification qui ne tient pas toujours compte de tous les éléments requis pour l'interprétation. Ainsi, certains variants dans ces bases sont mal classés. Ces bases devraient être utilisées comme un outil permettant de retrouver une publication rapportant le variant considéré et le biologiste devrait refaire l'interprétation du variant sur la base des arguments disponibles dans la/les publications et en tenant compte des recommandations.

1-1-3 Base de données locale

La consultation de la base de données locale de chaque centre de génétique est également d'une grande utilité. Dans le cadre du PFMG2025, le CAD permettra d'accéder à des données au niveau national. Il est fortement recommandé de soumettre les variants classés dans des bases de données internationales.

1-2- Critères cliniques et généalogiques

Au-delà des informations cliniques indispensables lors de la demande d'analyse, une confrontation clinico-biologique dans le cadre d'une Réunion de Concertation Pluridisciplinaire avec le généticien clinicien et/ou clinicien spécialiste prenant en charge le patient est fortement recommandée afin de préciser la corrélation phénotype-génotype.

Arguments :

- Phénotype ou histoire familiale spécifique en faveur de la pathologie associée au gène comprenant le variant. Cependant, l'analyse se faisant par définition sur des sujets atteints, la force de l'argument est fonction de la spécificité du phénotype observé, de l'hétérogénéité génétique pour ce phénotype et du nombre de gènes analysés. Elle est classée **PP4** dans les critères ACMG/AMP. En idéal, les phénotypes spécifiques attendus pour chaque gène doivent être définis par le laboratoire, comme dans les recommandations pour les gènes de surdité (Oza et al., 2018).

Précautions :

- Certains phénotypes sont peu spécifiques ou certains panels explorent l'hétérogénéité génétique d'un groupe phénotypique (ex : épilepsie). Il n'est alors pas possible d'utiliser l'argument de la compatibilité de la présentation phénotypique. En effet, la probabilité que

le variant soit causal est haute si un gène unique est exploré et que la spécificité du test est élevée. En revanche, lorsque, pour une condition avec hétérogénéité génétique, un grand panel ou un séquençage d'exome est utilisé, la probabilité d'identifier le variant causal est la somme de toutes les probabilités pour chaque gène individuellement mais la spécificité est réduite. Ainsi, si le phénotype n'est pas spécifique, l'argument PP4 ne peut pas être considéré. C'est pour cette raison que six sur dix recommandations gène-spécifiques récents n'utilisent plus cet argument dans la classification (ClinGen_TP53_Expert_Panel, 2019; Gelb et al., 2018; Kelly et al., 2018; Lee et al., 2018b; Luo et al., 2019; Mester et al., 2018).

1-3 Données alléliques

Arguments :

Pour une pathologie à transmission autosomique récessive, variant observé *en trans* (détermination de la phase par analyse familiale) avec un variant pathogène (**PM3**). Les nouvelles recommandations de groupe de travail ClinGen SVI propose le système de calcul de score PM3 suivant (SVI_Recommendation_for_PM3_Criterion, 2019)

Table 1. Points awarded per *in trans* proband

| Classification/Zygosité of other variant ¹ | Points per Proband | |
|---|---------------------------|----------------------|
| | Confirmed in <i>trans</i> | Phase unknown |
| Pathogenic or Likely pathogenic variant | 1.0 | 0.5 (P) 0.25 (LP) |
| Homozygous occurrence (max point 1.0) | 0.5 | N/A |
| Uncertain significance variant (max point 0.5) | 0.25 | 0.0 |

¹All variants should be sufficiently rare (meet PM2 specification); P - Pathogenic; LP - Likely pathogenic

Table 2. Recommendation for determining the appropriate evidence strength level for PM3

| PM3_Supporting | PM3 | PM3_Strong | PM3_VeryStrong |
|----------------|-----|------------|----------------|
| 0.5 | 1.0 | 2.0 | 4.0 |

Si un variant a été observé dans de nombreux cas index, chacun avec un variant pathogène différent sur l'autre allèle, la probabilité que les variants soient *en trans* est plus élevée même si la confirmation de la phase n'a pas été effectuée dans tous les cas. Dans ces cas, les valeurs peuvent être ajustées en fonction de la confiance dans l'occurrence *en trans*.

- Pour une pathologie à pénétrance complète de transmission autosomique dominante ou liée à l'X, co-occurrence du variant avec un variant pathogène *en trans* (dans le même gène): poids de l'argument laissé à l'appréciation du laboratoire (classé **BP2** dans les critères ACMG initiaux)
- Co-occurrence du variant avec un variant pathogène *en cis* (dans le même gène): poids de l'argument laissé à l'appréciation du laboratoire (classé **BP2** dans les critères ACMG initiaux)
- Co-occurrence du variant avec un variant pathogène dans un autre gène impliqué dans la même pathologie : poids de l'argument laissé à l'appréciation du laboratoire (classé **BP5** dans les critères ACMG initiaux)

Précautions :

- Dans certains gènes, des transmissions dominantes et récessives ont été décrites avec le même phénotype (ex : CLCN1/myotonie congénitale).
- Dans certains gènes, des transmissions dominantes et récessives ont été décrites avec des phénotypes distincts (ex : BRCA2/cancer du sein et de l'ovaire/Fanconi-like)
- Attention aux doubles diagnostics (Posey *et al.*, N Engl J Med 2017)

- Attention aux allèles complexes où 2 variants pathogènes ségrégent en *cis*

1-4- Données de ségrégation

Arguments :

- Variant *de novo* chez un patient sans antécédent familial :
avec confirmation de la paternité et de la maternité (**PS2**)
sans confirmation de la paternité et de la maternité si la pénétrance est considérée complète (**PM6**)

Les nouvelles recommandations de groupe de travail ClinGen SVI propose le système de calcul des scores PS2/PM6 (SVI_Recommendation_for_PS2_and_PM6_criteria, 2018) suivant

| Phenotypic consistency | Points per Proband | |
|--|--------------------|-----------------|
| | Confirmed de novo | Assumed de novo |
| Phenotype highly specific for gene | 2 | 1 |
| Phenotype consistent with gene but not highly specific | 1 | 0.5 |
| Phenotype consistent with gene but not highly specific and high genetic heterogeneity* | 0.5 | 0.25 |
| Phenotype not consistent with gene | 0 | 0 |

*Maximum allowable value of 1 may contribute to overall score

| Supporting (PS2_Supporting or PM6_Supporting) | Moderate (PS2_Moderate or PM6) | Strong (PS2 or PM6_Strong) | Very Strong (PS2_VeryStrong or PM6_VeryStrong) |
|--|-----------------------------------|-------------------------------|---|
| 0.5 | 1 | 2 | 4 |

- Variant co-ségréant avec la pathologie chez plusieurs membres atteints dans la même famille et/ou dans plusieurs familles (**PP1**). Plus le nombre de sujets testés attestant de la co-ségrégation du génotype avec le phénotype est important ou plus les sujets testés sont éloignés dans l'arbre généalogique, plus l'argument peut avoir un poids élevé. Afin de préciser ce critère, il a été proposé (Jarvik and Browning, Am J Hum Genet 2016) de calculer la probabilité (« N ») que la présence du ou des variants d'intérêt (selon le mode de transmission présumé) chez les individus atteints, soit due au hasard plutôt que due à une co-ségrégation. Par exemple, dans le cadre d'un modèle dominant, cette probabilité est de $N = (1/2)^m$, « m » étant le nombre de méioses informatives concernant la co-ségrégation.

En fonction de la probabilité N calculée, une pondération associée est proposée de la manière suivante :

Données issues d'une seule famille : $N \leq 1/32$: argument fort (PS) ; $N \leq 1/16$: argument moyen (PM) ; $N \leq 1/8$: argument faible (PP).

Données issues de plusieurs familles : $N \leq 1/16$: argument fort (PS) ; $N \leq 1/8$: argument moyen (PM) ; $N \leq 1/4$: argument faible (PP).

Des exemples de calcul pour la détermination de la probabilité « N », selon les différents modes de transmission, sont donnés dans l'article de Jarvik et Browning (Jarvik and Browning, Am J Hum Genet 2016).

Une façon plus simple d'évaluer la ségrégation est d'utiliser les tableaux avec des LOD scores pré-calculés. Cette approche est adoptée par les groupes ClinGen (les instructions détaillées disponibles à <https://clinicalgenome.org/docs/summary-of-updates-to-the-clingen-gene-clinical-validity->

[curation-sop-version-7/](#)).

Figure 5: Dominant/X-linked LOD score table

| | | | | | | | | | | | | |
|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Dominant Segregations | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 |
| Estimated LOD | 4.5 | 4.2 | 3.9 | 3.6 | 3.3 | 3.0 | 2.7 | 2.4 | 2.1 | 1.8 | 1.5 | 1.2 |

Figure 6: Recessive estimated LOD (eLOD) score table

| | | Unaffecteds | | | | | | | | | | |
|-----------|----|-------------|------|------|------|------|------|------|------|------|------|------|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Affecteds | 3 | 1.20 | 1.32 | 1.45 | 1.50 | 1.70 | 1.82 | 1.95 | 2.07 | 2.20 | 2.33 | 2.45 |
| | 4 | 1.81 | 1.93 | 2.06 | 2.18 | 2.31 | 2.43 | 2.56 | 2.68 | 2.81 | 2.93 | 3.06 |
| | 5 | 2.41 | 2.53 | 2.66 | 2.78 | 2.91 | 3.03 | 3.16 | 3.28 | 3.41 | 3.53 | 3.66 |
| | 6 | 3.01 | 3.14 | 3.26 | 3.39 | 3.51 | 3.63 | 3.76 | 3.88 | 4.01 | 4.13 | 4.26 |
| | 7 | 3.61 | 3.74 | 3.86 | 3.99 | 4.11 | 4.24 | 4.36 | 4.49 | 4.61 | 4.74 | 4.86 |
| | 8 | 4.21 | 4.34 | 4.46 | 4.59 | 4.71 | 4.84 | 4.96 | 5.09 | 5.21 | 5.34 | 5.46 |
| | 9 | 4.82 | 4.94 | 5.07 | 5.19 | 5.32 | 5.44 | 5.57 | 5.69 | 5.82 | 5.94 | 6.07 |
| | 10 | 5.42 | 5.54 | 5.67 | 5.79 | 5.92 | 6.04 | 6.17 | 6.29 | 6.42 | 6.54 | 6.67 |

Les recommandations gènes spécifiques utilisent les seuils suivants pour attribuer l'argument PP1 :

| | MYH7, Rasopathy PTEN, CDH1, RUNX1, TP3 | MEN1 | Hearing loss AD | PAH | Hearing loss AR |
|----------------|--|---|--|--|-----------------|
| PP1_Strong | > 7 meioses (likelihood ratio 100, LOD 2,1) | $N \leq 1/32$ if 1 family $N \leq 1/16$ if 2+ families | 5 affected (likelihood ratio 1:32, LOD 1,5) | 3 affected + 0 non-affected 2 affected + 3 non-affected | 3 affected |
| PP1_Moderate | 5-6 meioses (likelihood ratio 30, LOD 1,5) | $N \leq 1/16$ if 1 family $N \leq 1/8$ if 2+ families | 4 affected (likelihood ratio 1:16, LOD 1,2) | 2 affected + 0 non-affected | 2 affected |
| PP1_Supporting | 3-4 meioses (likelihood ratio 10, LOD 0,9) | $N \leq 1/8$ if 1 family $N \leq 1/4$ if 2+ families | 2 affected (likelihood ratio 1:4, LOD 0,6) | 1 affected + 3 non-affected | 1 affected |

- Variant ne ségrégeant pas avec la pathologie chez des apparentés atteints (**BS4**)

Précautions :

- Le critère de novo (PM6 ou PS2) ne peut être utilisé que si le phénotype du patient est concordant.
- Le critère PS2 peut être modifié en PM ou PP si le variant est en mosaïque avec une représentation tissulaire cohérentes avec le phénotype
- Attention aux phénocopies pouvant mimer un défaut de ségrégation
- Attention aux pathologies à variabilité d'expression phénotypique : lorsqu'un variant est hérité d'un parent apparemment asymptotique, une ré-évaluation de son phénotype doit être proposée avant de conclure à tort au caractère bénin du variant.
- Attention aux pathologies à transmission dominante avec prévalence importante. En effet, en raison de la prévalence importante, un individu peut être porteur de deux variants à effet dominant, dans deux gènes différents impliqués dans un même groupe phénotypique (situation de double hétérozygote) ; ou de deux variants à effet dominant dans le même gène (situation hétérozygote composite, avec un variant dominant sur l'allèle maternel, et un variant dominant sur l'allèle paternel).
- Un variant peut sembler co-ségréger avec la pathologie parce qu'il est en *cis* avec un variant pathogène non identifié
- Pour les transmissions liées à l'X : il est utile de rechercher les variants pathogènes éventuels chez les grands-parents maternels; de rechercher des signes cliniques *a minima* chez la mère ; d'effectuer une étude de biais d'inactivation de l'X chez les femmes hétérozygotes.
- Attention aux défauts de ségrégation dus aux fausses paternités, disomies uniparentales, ou phénomène de l'empreinte...

1-5- Données prédictifs

1-5-1 Nature du variant nucléotidique

Arguments :

- Variant ayant un effet nul prédit (non-sens, frameshift, site canonique d'épissage : -2, -1, +1, +2, codon d'initiation de la traduction, délétion d'un ou plusieurs exons hors phase, délétion d'un ou plusieurs exons en phase mais emportant un domaine fonctionnel) dans un gène où la perte de fonction est le mécanisme pathogène connu (**PVS1**). Un arbre décisionnel pour attribuer cet argument a été établi par ClinGen SVI (Abou Tayoun et al., 2018). PVS1 ne peut pas être combiné avec l'argument PP3.
- Variant affectant la longueur de la protéine (Indels en phase, perte du codon stop) : modulation du poids de l'argument : le poids de l'argument en faveur de la pathogénicité du variant est inversement proportionnel à la taille résiduelle de la protéine (**PM4**). Si l'altération est en phase dans une région répétée sans fonction connue, l'argument conféré est faible et en faveur du caractère bénin du variant (**BP3**).
- Variant faux-sens dans un gène avec un faible taux de variants faux-sens bénins et dans lequel les variants faux-sens sont un mécanisme fréquemment responsable de la pathologie (**PP2**)
- Variant faux-sens dans un gène où seuls les variants tronquants sont décrits comme associés à la pathologie (**BP1**)
- Variant synonyme sans impact prédit sur l'épissage **ET** pour lequel la séquence nucléotidique n'est pas très conservée chez les vertébrés (**BP7**)

Précautions :

- Attention à l'attribution d'un effet perte de fonction aux variants situés après la dernière jonction intron/exon ou à moins de 50 nucléotides en amont de celle-ci (ARNm non soumis au *nonsense-mediated decay* ou NMD)
- Attention à l'épissage physiologique avec transcrits multiples
- Attention à l'apparition de nouveaux sites cryptiques d'épissage (par exemple des variants d'épissage rajoutant 1 ou plusieurs codons résultant en une indel en phase)
- Attention au saut d'exon(s) en phase, laissant intact le reste de la protéine. L'argument PVS1 n'est valable que pour les gènes où la perte de fonction est le mécanisme pathogène connu. Il convient de se renseigner sur la pathologie si nécessaire et d'utiliser les scores qui prédisent l'intolérance d'un gène aux variants de type « perte de fonction », disponibles pour chaque gène dans gnomAD. Le score pLI (*probability of Loss of function intolerance*) est largement utilisé depuis plusieurs années, même si ce score comporte un certain nombre de biais (transcrit utilisé pour son calcul, non prise en compte des variants de type décalage du cadre de lecture, méthodologie de prise en compte de la taille du cDNA, etc.)(Ziegler et al., 2019). Un gène est considéré comme intolérant à la perte de fonction si sa pLI est supérieure ou égale à 0,9. Le nouveau score o/e (ainsi que le score associé LOUEF), prend en compte la taille du gène permettant d'avoir une estimation de l'intolérance à la perte de fonction pour des petits gènes avec peu de variants (Karczewski et al., 2019). Ce score correspond au ratio entre le nombre de variants perte de fonction observés dans la population pour un gène et le nombre de variants attendus dans un gène (observed/expected). Des gènes avec o/e inférieurs à 0.35 sont considérés comme intolérants à la perte de fonction. L'utilisation du HI index (*Haploinsufficiency Score*) de la base de données DECIPHER (<https://decipher.sanger.ac.uk/>) présente le même intérêt et les mêmes limites mais sur un jeu de données indépendant (Huang et al., 2010). Une analyse qualitative de la charge en

variants tronquants dans les bases de données est utile afin d'aborder ces scores de manière critique.

- Pour les variants affectant les codons d'initiation de la traduction, il est important de rechercher l'existence de codons d'initiation de la traduction alternatifs éventuels (analyse ORFinder,...)
- Pour les variants de type perte de codon stop, il est important de rechercher l'existence de codons stop alternatifs éventuels (analyse ORFinder,...)
- Il est important de rechercher un éventuel effet sur l'épissage pour tout variant, qu'il soit intronique ou exonique (comprenant synonymes, non-sens, faux-sens, indel)

Conservation nucléotidique

A propos de la conservation nucléotidique, il est important de bien considérer la profondeur évolutive des espèces alignées afin de garantir une représentation la plus appropriée de génomes connus. En général, les espèces alignables pour le génome humain sont des vertébrés. Il est à noter également que la dégénérescence du code génétique, impliquant la 3^{ème} base du codon, induit une conservation souvent bien faible à ces positions. Ainsi, l'observation d'une conservation à ces positions pourrait être éventuellement considérée comme plus significative, suggérant l'existence d'une éventuelle séquence régulatrice à ce niveau (par exemple les ESE pour *exonic splicing enhancer* et les ESS pour *exonic splicing silencer*).

1-5-2- Prédictions bio-informatiques

En cas de variant prédit faux-sens, il est nécessaire d'interroger plusieurs logiciels de prédiction du caractère délétère pour la fonction protéique, souvent appelés logiciels de prédiction de pathogénicité, comme SIFT, MutationTaster, Polyphen-2 ou UMD-Predictor. Ces algorithmes attribuent un score et une classification en faveur de la probabilité du caractère délétère d'un variant faux-sens ou du caractère bénin du variant en fonction de différents critères : conservation de la base nucléotidique et surtout de l'acide aminé dans l'évolution, distance physico-chimique entre le résidu aminoacide sauvage et muté (score de Grantham), position dans un domaine fonctionnel connu de la protéine. Certains outils comme Mutation Taster ou UMD-Predictor intègrent à la fois la prédiction d'un effet sur l'acide aminé et d'un effet sur l'épissage. Néanmoins, pour apprécier un éventuel effet sur l'épissage, il est préférable d'approfondir les prédictions au moyen d'outils spécifiques de prédiction des répercussions sur l'épissage (voir paragraphe ci-dessous). Les nombreux outils de prédiction de la pathogénicité sont recensés dans la base de données VIPdb <https://genomeinterpretation.org/vipdb> (Hu et al., 2019). Une étude récente a testé 25 de ces outils pour l'application en diagnostic et a identifié plusieurs algorithmes de haute performance relativement récents qui ne semblent pas avoir été intégrés dans les pipelines des laboratoires cliniques (tels que REVEL, VEST3, etc.) (Ghosh et al., 2017). En effet, trois des recommandations gène-spécifiques attribuent maintenant l'argument PP3 selon la prédiction d'un seul score (REVEL), qui représente une méthode de prédiction de la pathogénicité des variants faux-sens basée sur une combinaison de scores provenant de 13 outils individuels (MutPred, FATHMM v2.3, VEST 3.0, Polyphen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP et phastCons) (ClinGen_Lysosomal_Storage_Disorders_Expert_Panel, 2019; Ioannidis et al., 2016; Luo et al., 2019; Oza et al., 2018). En idéal, la performance des outils de prédictions devrait être testée pour des gènes analysés par les laboratoires en routine, car il peut avoir des différences entre les gènes dans les outils le plus adaptés. Par exemple, les recommandations pour le gène TP53 préconisent de ne pas utiliser PolyPhen2 et SIFT pour la prédiction de la pathogénicité des variants dans ce gène.

Il est nécessaire d'appliquer des logiciels de prédiction de l'effet potentiel sur l'épissage du variant étudié quelle que soit sa nature, y compris les variants non-sens, faux-sens et synonymes. Des outils classiques, tels que MaxEntScan (Yeo and Burge, 2004), SpliceSiteFinder, NNSPLICE, GeneSplicer et Human Splicing Finder (Desmet et al., 2009) sont encore largement utilisés en diagnostic (Houdayer et al., 2012). Les approches plus récentes basées sur « machine learning », telles que SpliceAI (Jaganathan et al., 2019), S-CAP (Jagadeesh et al., 2019) et MMSplice (Cheng et al., 2019), commencent également à prendre de l'importance. En cas d'effet délétère prédit, des analyses des transcrits chez le patient et/ou des tests *ex vivo* d'épissage peuvent être réalisés au laboratoire afin d'objectiver ce potentiel effet et d'en faire éventuellement un argument fort en faveur de la pathogénicité du variant (PS3).

Au sein d'une même thématique, pour veiller à l'homogénéité de l'interprétation des variants, il serait intéressant de s'accorder sur les outils de prédiction à interroger. La complémentarité des outils tient au fait qu'ils utilisent des algorithmes distincts et qu'ils ont été entraînés et validés sur des sets de variants « vrais positifs » et « vrais négatifs » distincts.

Arguments :

- Effet d'un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés (PP3)
- Effet d'un variant prédit bénin par l'ensemble des logiciels de prédiction de pathogénicité interrogés (BP4)
- Variant non-tronquant situé sur un hot spot mutationnel et/ou un domaine fonctionnel critique bien établi exempt de variants bénins (PM1)

Précautions :

- Il n'est pas possible d'exploiter les résultats de ces logiciels en cas de discordance de prédiction pour les variants faux-sens. La discordance de prédiction entre ces logiciels n'exclut pas la possible pathogénicité ou neutralité du variant
- Prudence avec les outils de prédiction de variants d'épissage : autant les outils sont fiables pour la création de sites cryptiques, l'abolition de sites d'épissage, autant les prédictions d'éléments enhancer ou silencer d'épissage ne sont pas fiables à ce jour.
- Attention aux variants multiples (MNV ; *Multiple Nucleotide Variant* ou delins) correspondant à deux variants en phase, espacés ou répartis sur quelques nucléotides : nécessité d'interpréter l'impact global de ces variants
- En raison du manque de spécificité des prédictions bio-informatiques, la plus grande prudence doit être aussi accordée à l'absence de prédiction d'un effet. Le risque encouru est un faux négatif. Ce résultat doit être interprété avec les autres arguments comme la fréquence du variant dans les populations contrôles ou son caractère hérité.

Choix des logiciels de prédictions du retentissement sur l'épissage ; choix des logiciels de prédiction des variants faux-sens

Il est essentiel de combiner les approches, c'est-à-dire d'interroger plusieurs outils et algorithmes, afin d'obtenir la meilleure sensibilité et la meilleure spécificité possible. Il est néanmoins important d'estimer le niveau de redondance entre les logiciels afin de ne pas avoir un effet de surprédictions préjudiciable au résultat final.

Plus récemment, des scores appelés « intégrés » ont été mis au point afin de regrouper les prédictions d'un nombre donné d'autres logiciels de prédiction (CADD, Condel, KGGseq...). Ces logiciels ont l'avantage de permettre des prédictions sur des variants exoniques et non exoniques. En revanche, comme tous les logiciels de prédiction bioinformatique, ils sont dépendants des sets de variants utilisés lors de l'entraînement (*machine learning*) pour les mettre au point. Ainsi, il

apparaît que l'interrogation de plusieurs outils distincts « non-ensemble » a certainement un intérêt complémentaire par rapport à l'utilisation d'un seul outil « ensemble » associé à un seuil donné, à ce jour.

1-6- Données bibliographiques

Afin d'identifier tous les cas déjà décrits pour un variant à classer et attribuer l'argument PS4, il est nécessaire de faire une revue de la littérature concernant ce variant. Les bases de données des variants identifiés chez les individus atteints (ClinVar, LOVD etc) fournissent souvent une liste de références pour les variants. Pour compléter cette liste, il est conseillé d'utiliser PubMed (NCBI, *National Center for Biotechnology Information*), Google Scholar, LitVar (Allot et al., 2018), AVADA (Automatic Variant evidence DAtabase) (Birgmeier et al., 2020), Genomenon Mastermind et HGMD. Les variants présents dans les bases de données HGMD-public, Genomenon et AVADA sont facilement accessibles à partir d'UCSC Genome Browser (Haeussler et al., 2019; Kent et al., 2002)

La base de données HGMD donne accès rapidement aux publications rapportant un variant donné, le plus souvent non identifiables par interrogation simple par mots clés dans Pubmed ou autre moteur de recherche.

- Variant déjà classé pathogène ou à l'origine du même changement d'acide aminé qu'un variant pathogène connu (**PS1**), sauf si le variant n'est pas un variant faux-sens mais a un effet sur l'épissage

1-7- Données fonctionnelles

Arguments :

- Etudes fonctionnelles *in vivo* ou *in vitro* bien établies montrant un impact délétère du variant sur le gène ou son produit (**PS3**)

- Cet argument peut inclure les résultats d'analyses fonctionnelles spécifiques en fonction de la pathologie concernée. Exemples : caractérisation d'un déficit enzymatique pour l'enzyme codée par le gène dans lequel le(les) variant(s) d'intérêt a (ont) été détecté(s) ; Western-blot et immunofluorescence indirecte signant un déficit en protéine codée par le gène dans lequel le(les) variant(s) d'intérêt a(ont) été détecté(s).

- Etudes fonctionnelles *in vivo* ou *in vitro* bien établies montrent un impact non délétère du variant (**BS3**)

1-8- Analyses somatiques

Ce chapitre qui n'existe pas dans les recommandations ACMG vise à intégrer les analyses somatiques

Chaque FNMR/réseau d'oncogénétique est amené à déterminer le poids de ces arguments dans l'interprétation. Les laboratoires d'oncogénétique doivent travailler en étroite collaboration avec les laboratoires d'oncosomatique pour partager leurs variants alléliques. Un variant pathogène détecté uniquement au niveau tumoral peut aider à la classification de variants constitutionnels et doit être intégré à la réflexion. Ce critère est à considérer gène par gène, la LoH n'étant pas un marqueur fiable de la pathogénicité d'un variant pour tous les gènes.

L'interprétation de variants constitutionnels peut aussi s'appuyer sur des analyses somatiques quand du tissu tumoral est disponible : marquage immunohistochimique, perte d'hétérozygotie dans la région chromosomique du gène d'intérêt (ex : recherche d'instabilité microsatellitaire et marquage immunohistochimique dans le syndrome de Lynch ou marquage immunohistochimique dans les paragangliomes héréditaires).

1-9- Modifications de l'argumentaire par rapport aux recommandations

La liste initiale des arguments recommandés par l'ACMG-AMP a été modifiée pour s'adapter aux positions françaises. Les arguments qui ont été modifiés sont les suivants :

- Force de l'argument «Co-occurrence du variant avec un variant pathogène (même gène, autre gène pour la même pathologie)» : poids de l'argument laissé à l'appréciation du laboratoire (au lieu de BP2 et BP5 selon les recommandations ACMG).
- Ajout de l'actualisation selon Jarvik et Browning (Am J Hum Genet 2016). pour préciser l'utilisation des données de ségrégation (BS4 et PP1).
- Ajout de l'argument « analyses somatiques », analyses qui ont un poids déterminant dans l'interprétation. Ces analyses somatiques n'existaient pas dans les recommandations ACMG. Chaque FNMR/réseau d'oncogénétique est amenée à déterminer le poids de ces arguments dans l'interprétation.

2 - Classification du variant

Les différents arguments pondérés listés ci-dessus sont ensuite combinés, selon les critères de l'ACMG et de l'AMP, afin de déterminer la classe du variant :

PVS/PS/PM/PP : Argument très fort (*Pathogenic Very Strong*) / fort (*Pathogenic Strong*) / moyen (*Pathogenic Moderate*) / faible (*Pathogenic Poor*) en faveur de la pathogénicité du variant.

BA/BS/BP : Argument suffisant (*Benign Stand Alone*) / fort (*Benign Strong*) / faible (*Benign Poor*) en faveur du caractère bénin du variant.

Classe 5 : Variant

pathogène 1 **PVS** ET ≥ 1

PS OU

≥ 2 **PM** OU

1 **PM** ET 1 **PP** OU

≥ 2 **PP** OU

≥ 2 **PS** OU

1 **PS** ET ≥ 3 **PM** OU

2 **PM** ET ≥ 2 **PP** OU

1 **PM** ET ≥ 4 **PP**

Classe 4 : Variant probablement pathogène

1 **PVS** ET 1 **PM** OU

1 **PS** ET 1–2 **PM** OU

1 **PS** ET ≥ 2 **PP** OU

≥ 3 **PM** OU

2 **PM** ET ≥ 2 **PP** OU

1 **PM** ET ≥ 4 **PP**

Classe 3 : Variant de signification inconnue

Les arguments pondérés ne permettent pas de classer le variant dans une autre classe

Classe 2 : Variant probablement bénin

1 **BS** ET 1 **BP** OU

≥2 **BP**

Classe 1 : Variant bénin

BA OU

≥2 **BS**

NB : afin de faciliter l'établissement des scores pour la classification des variants, différentes ressources sont disponibles en ligne :

- ClinGen Pathogenicity Calculator : <http://calculator.clinicalgenome.org/site/cg-calculator> - InterVar Classify System: <http://wintervar.wglab.org> - Genetic Variant Interpretation Tool:
http://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/
- SPiCE : outil de prédiction des anomalies d'épissage aussi (SPiCE) validé sur 395 variants d'épissage des gènes *BRCA1*, *BRCA2*, *CFTR*, *RHD*

L'annexe 1 reprend le poids de chaque argument et les précautions d'interprétation (Tableaux 3 et 4 de Richards *et al.*, 2015).

3 - Utilisation dans le cadre d'un compte-rendu d'analyses

Il est recommandé de préciser dans la partie méthodologie des comptes-rendus d'analyse la phrase suivante :

« Les variants sont décrits selon la **nomenclature HGVS** (<https://varnomen.hgvs.org>), et classés dans les catégories « pathogène », « probablement pathogène », « bénin », « probablement bénin », ou « de significativité incertaine » 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 XXX , dans les limites des connaissances actuelles. »

Les variants retenus par le généticien responsable de l'analyse figureront avec leur classification dans la partie résultats du compte-rendu.

Il est recommandé par ailleurs que le compte-rendu spécifie dans la partie conclusion si les variants classés en Classe 4 (probablement pathogène) et Classe 5 (pathogène) retenus par le généticien responsable de l'analyse sont considérés comme en lien (causal) avec la pathologie que présente le patient en prenant en compte les informations cliniques et paracliniques disponibles, en interaction avec les cliniciens prescripteurs et éventuellement à l'issue de réunions de concertation pluridisciplinaires.

4 - Bibliographie et Annexes

Abou Tayoun, A.N., Pesaran, T., DiStefano, M.T., Oza, A., Rehm, H.L., Biesecker, L.G., Harrison, S.M., and ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI) (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* 39, 1517–1524.

Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S.H., Kiemeney, L.A.L.M., Veltman, J.A., Gilissen, C., and Hoischen, A. (2017). Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* *101*, 50–64.

Allot, A., Peng, Y., Wei, C.-H., Lee, K., Phan, L., and Lu, Z. (2018). LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. *Nucleic Acids Res.* *46*, W530–W536.

Amendola, L.M., Jarvik, G.P., Leo, M.C., McLaughlin, H.M., Akkari, Y., Amaral, M.D., Berg, J.S., Biswas, S., Bowling, K.M., Conlin, L.K., et al. (2016). Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am. J. Hum. Genet.* *98*, 1067–1076.

BA1_Exception_List (2018). BA1 Exception List, July 2018.

BA1_Exception_List_Nomination_Form BA1 Exception List Nomination Form.

Bean, L.J.H., Funke, B., Carlston, C.M., Gannon, J.L., Kantarci, S., Krock, B.L., Zhang, S., Bayrak-Toydemir, P., and ACMG Laboratory Quality Assurance Committee (2019). Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med. Off. J. Am. Coll. Med. Genet.*

Biesecker, L.G., Harrison, S.M., and ClinGen Sequence Variant Interpretation Working Group (2018). The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *20*, 1687–1688.

Birgmeier, J., Deisseroth, C.A., Hayward, L.E., Galhardo, L.M.T., Tierno, A.P., Jagadeesh, K.A., Stenson, P.D., Cooper, D.N., Bernstein, J.A., Haeussler, M., et al. (2020). AVADA: toward automated pathogenic variant evidence retrieval directly from the full-text literature. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *22*, 362–370.

Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* *12*, 3.

Cheng, J., Nguyen, T.Y.D., Cygan, K.J., Çelik, M.H., Fairbrother, W.G., Avsec, Ž., and Gagneur, J. (2019). MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* *20*, 48.

ClinGen_Lysosomal_Storage_Disorders_Expert_Panel (2019). ClinGen Lysosomal Storage Disorders Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1.

ClinGen_TP53_Expert_Panel (2019). ClinGen TP53 Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., and Bérout, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* *37*, e67.

Erikson, G.A., Bodian, D.L., Rueda, M., Molparia, B., Scott, E.R., Scott-Van Zeeland, A.A., Topol, S.E., Wineinger, N.E., Niederhuber, J.E., Topol, E.J., et al. (2016). Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* *165*, 1002–1011.

Gelb, B.D., Cavé, H., Dillon, M.W., Gripp, K.W., Lee, J.A., Mason-Suares, H., Rauen, K.A., Williams, B., Zenker, M., Vincent, L.M., et al. (2018). ClinGen’s RASopathy Expert Panel consensus methods for variant interpretation. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *20*, 1334–1345.

Ghosh, R., Oak, N., and Plon, S.E. (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* *18*, 225.

Ghosh, R., Harrison, S.M., Rehm, H.L., Plon, S.E., Biesecker, L.G., and ClinGen Sequence Variant Interpretation Working Group (2018). Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum. Mutat.* *39*, 1525–1530.

Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* *47*, D853–D858.

Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., et al. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.* *33*, 1228–1238.

- Hu, Z., Yu, C., Furutsuki, M., Andreoletti, G., Ly, M., Hoskins, R., Adhikari, A.N., and Brenner, S.E. (2019). VIPdb, a genetic Variant Impact Predictor Database. *Hum. Mutat.* *40*, 1202–1214.
- Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet.* *6*, e1001154.
- Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* *99*, 877–885.
- Jagadeesh, K.A., Paggi, J.M., Ye, J.S., Stenson, P.D., Cooper, D.N., Bernstein, J.A., and Bejerano, G. (2019). S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. *Nat. Genet.* *51*, 755–763.
- Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* *176*, 535-548.e24.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes (Genomics).
- Kelly, M.A., Caleshu, C., Morales, A., Buchan, J., Wolf, Z., Harrison, S.M., Cook, S., Dillon, M.W., Garcia, J., Haverfield, E., et al. (2018). Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *20*, 351–359.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* *12*, 996–1006.
- Lee, K., Krempely, K., Roberts, M.E., Anderson, M.J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., et al. (2018a). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum. Mutat.* *39*, 1553–1568.
- Lee, K., Krempely, K., Roberts, M.E., Anderson, M.J., Carneiro, F., Chao, E., Dixon, K., Figueiredo, J., Ghosh, R., Huntsman, D., et al. (2018b). Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum. Mutat.* *39*, 1553–1568.
- Luo, X., Feurstein, S., Mohan, S., Porter, C.C., Jackson, S.A., Keel, S., Chicka, M., Brown, A.L., Kesserwan, C., Agarwal, A., et al. (2019). ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv.* *3*, 2962–2979.
- Mester, J.L., Ghosh, R., Pesaran, T., Huether, R., Karam, R., Hruska, K.S., Costa, H.A., Lachlan, K., Ngeow, J., Barnholtz-Sloan, J., et al. (2018). Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum. Mutat.* *39*, 1581–1592.
- Oza, A.M., DiStefano, M.T., Hemphill, S.E., Cushman, B.J., Grant, A.R., Siegert, R.K., Shen, J., Chapin, A., Boczek, N.J., Schimmenti, L.A., et al. (2018). Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum. Mutat.* *39*, 1593–1613.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *17*, 405–424.
- Romanet, P., Odou, M.-F., North, M.-O., Saveanu, A., Coppin, L., Pasmant, E., Mohamed, A., Goudet, P., Borson-Chazot, F., Calender, A., et al. (2019). Proposition of adjustments to the ACMG-AMP framework for the interpretation of MEN1 missense variants. *Hum. Mutat.* *40*, 661–674.
- Scott, E.M., Halees, A., Itan, Y., Spencer, E.G., He, Y., Azab, M.A., Gabriel, S.B., Belkadi, A., Boisson, B., Abel, L., et al. (2016). Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat. Genet.* *48*, 1071–1076.

SVI_Recommendation_for_PM3_Criterion (2019). SVI Recommendation for in trans Criterion (PM3) - Version 1.0.

SVI_Recommendation_for_PS2_and_PM6_criteria (2018). PS2/PM6: Recommendation for de novo PS2 and PM6 ACMG/AMP criteria (Version 1.0).

Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program (Genomics).

TOPMed (2018). The NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. BRAVO variant browser: University of Michigan and NHLBI; 2018. Available from: <https://bravo.sph.umich.edu/freeze5/hg38/>.

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A.H., Karczewski, K., Ing, A.Y., Barton, P.J.R., Funke, B., Cook, S.A., MacArthur, D., et al. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med. Off. J. Am. Coll. Med. Genet.* *19*, 1151–1158.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* *11*, 377–394.

Zastrow, D.B., Baudet, H., Shen, W., Thomas, A., Si, Y., Weaver, M.A., Lager, A.M., Liu, J., Mangels, R., Dwight, S.S., et al. (2018). Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): The ClinGen IEM Working Group and the Phenylalanine Hydroxylase Gene. *Hum. Mutat.* *39*, 1569–1580.

Ziegler, A., Colin, E., Goudenège, D., and Bonneau, D. (2019). A snapshot of some pLI score pitfalls. *Hum. Mutat.*

SERMENT D'HIPPOCRATE

Au moment d'être admis(e) à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.

