Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS		
Date de création : 20/12/2017	Date de révision : 11/02/2021	Version : 2
Date de 1ère Application : 01/06/2018	N° document : NGSDIAG_001	
Approbation par le board du réseau le 24/03/2021		

La version 2 de ce document a été préparée par les membres du Groupe de Travail du Réseau NGS-Diag listés ci-dessous :

Cécile ACQUAVIVA-BOURDAIN (Lyon ; ANPGM, FNMR G2M)
Stéphanie BAERT-DESURMONT (Rouen ; représentant GGC)
Sandrine CAPUTO (Paris Curie ; représentant GGC)
Nicolas CHATRON (Lyon ; représentant Réseau AChro-Puce)
Florence COULET (Paris ; représentant GGC)
Martine DOCO-FENZY (Reims ; représentant ACLF)
Boris KEREN (La Pitié ; représentant Réseau AChro-Puce)
Cédric LE MARECHAL (Brest ; ANPGM, FNMR SENSGENE)
Jean MULLER (Strasbourg ; représentant Réseau NGS-Diag)
Gaël NICOLAS (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience)
Vincent PROCACCIO (Angers ; ANPGM, MitoDIAG)
Pascale RICHARD (Paris Pitié Salpêtrière ; ANPGM, FNMR CARDIOGEN)
Pauline ROMANET (Marseille ; ANPGM, FNMR FIRENDO, OSCAR)
Cécile ROUZIER (Nice ; représentant Réseau NGS-Diag, réseau MitoDIAG)
Sarah SNANOUDJ (Rouen ; FNMR G2M)

Coordination :

Martin KRAHN (Marseille ; ANPGM, FNMR FILNEMUS)
Pascale SAUGIER-VEBER (Rouen ; ANPGM, FNMR AnDDi-Rares et DefiScience)
Svetlana GOROKHOVA (Marseille ; ANPGM, BioInfoDiag)

Sommaire

Contexte de l'initiative Version 1 (2018). 4
Contexte de l'initiative Version 2 (2020). 5
Recommandations pour l'interprétation de variants générés par le séquençage à haut débit. 6
1 Liste des arguments à prendre en compte pour l'interprétation des variants 6
1-1 Données épidémiologiques 8
1-1-1 Données de fréquence allélique dans les populations contrôles 8
1-1-2 Données de fréquence des variants chez les sujets atteints 11
1-1-3 Bases de données locales 13
1-2 Critères cliniques 13
1-3 Co-occurrence de variants 14
1-4 Données de ségrégation 15
1-5 Données prédictives et computationnelles 17
1-6 Données fonctionnelles 23
1-7 Analyses somatiques 23
1-8 Analyse des variants mitochondriaux 24
1-9 Interprétation des variants structuraux (SV), dont CNV 25
1-10 Modifications de l'argumentaire par rapport aux recommandations 25
2 Classification du variant 26
3 Utilisation dans le cadre d'un compte-rendu d'analyses 26
4 Bibliographie et Annexes 27

Abréviations utilisées dans ce document :

ACLF : Association des Cytogénéticiens 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 Héréditaires 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
ExAC : Exome Aggregation Consortium
FILNEMUS : Filière de Santé des Maladies Rares Neuromusculaires
FIRENDO : Filière Maladies Rares Endocriniennes
FNMR : Filières de soin Nationales de Maladies Rares
G2M : Filière Maladies Rares Maladies héréditaires du métabolisme
GGC : Groupe Génétique et Cancer d'UNICANCER
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
Réseau AChroPuce: Réseau d'Analyse Chromosomique sur Puce à ADN
Réseau MitoDIAG : Réseau des laboratoires impliqués dans le diagnostic des maladies mitochondriales
SENSGENE : Filière de Maladies Rares Sensorielles
SV : Structural variant (englobant délétions, duplications, translocations, insertions, inversions…)

Contexte de l'initiative Version 1 (2018).

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*) (Richards et al., Genet Med 2015 ; Amendola et al., Am J Hum Genet 2016) 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-DIAG 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

Le document présenté ici reprend les recommandations ACMG/AMP (Richards et al., Genet Med 2015 ; Amendola et al., Am J Hum Genet 2016), et incorpore les éléments d'autres publications (références intégrales dans la section « bibliographie ») visant à actualiser ces recommandations, dont des recommandations émises par l'Association for Clinical Genetic Science (Ellard et al., 2017 http://www.acgs.uk.com/media/1092626/uk_practice_guidelines_for_variant_classification_2017.pdf).

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.

Ce document, dans sa 1ère version, ne s'applique qu'aux maladies à transmission mendélienne. La réflexion sur la classification des variants dans les maladies à transmission non mendélienne est en cours et sera intégrée dans une prochaine version du document.

La réflexion sur la classification des SV (Structural Variants) est en cours, et sera finalisée dans un premier temps pour les CNV (Copy Number Variations) (ACLF et AchroPuce). Les recommandations portant sur cette thématique seront intégrées dans une prochaine version du document.

La réflexion sur la classification des variants de l'ADN mitochondrial est en cours et les recommandations portant sur cette thématique seront intégrées dans une prochaine version du document.

Contexte de l'initiative Version 2 (2020).

La mise en place du séquençage à haut débit (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 n'avaient pas d'expertise spécifique. 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 & Association for Molecular Pathology*), EuroGentest/ESHG (*European Society of Human Genetics*) ont émis des recommandations pour guider les laboratoires dans l'interprétation [1–3]. Le système de calcul de pathogénicité des variants proposé par les recommandations d'ACMG/AMP a été ensuite adapté par ACGS (*Association for Clinical Genomic Science*, UK) en 2016 (voir Appendix 1 de [4]) pour devenir aujourd'hui le standard mondial pour l'interprétation des analyses par séquençage à haut débit. 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). La première version de ce document a été publiée en juin 2018 (NGSDiag_001) 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. Ces recommandations ont été adoptées par la grande majorité des laboratoires diagnostiques en France.

Depuis la diffusion de NGSDiag_001, un certain nombre de mises-à-jours et de précisions a été publié par le groupe de travail sur l'interprétation des variants au sein de ClinGen (*ClinGen Sequence Variant Interpretation* (SVI) *Working Group*) ainsi que par les groupes internationaux d'experts dans différents domaines de la génétique médicale (*Gene Curation Expert Panels* et *Variant Curation Expert Panels*). Il s'agit de précisions générales sur les critères de classification (PVS1, PS2/PM6, PS3/BS3, PM3, PP5/BP6 ou BA1) [5–10] ainsi que de 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*) [11–20]. Les recommandations des groupes de travail ClinGen sont recensées sur le site de ClinGen [21]. **La 2ème version de ce document prend en compte les modifications des recommandations publiées depuis la première version.**

La nouvelle version de ce document inclut maintenant l'interprétation des CNV intragéniques et des variants de l'ADN mitochondrial. Ce présent document contient également des éléments de certaines recommandations gène-spécifiques à titre d'exemple. Néanmoins, l'objectif de ce présent document n'est pas de répertorier de manière exhaustive toutes les recommandations gène-spécifiques émises. Ces dernières peuvent être importantes selon les thématiques et sont du ressort des laboratoires ayant l'expertise adéquate, en lien avec les réseaux professionnels correspondants (groupes, filières, etc.).

Note d'utilisation : **Niveau de preuve pour le lien « gène – pathologie ».**
Les recommandations décrites dans ce document ne sont applicables qu'aux gènes clairement impliqués dans la pathologie pour laquelle l'analyse a été effectuée. Si un variant est identifié dans un gène pour lequel le niveau de preuve de son implication dans la pathologie concernée est insuffisant, la classification maximale attribuable à ce variant est « VSI » (Variant de Signification Incertaine) [22,23]. Le lien gène-pathologie peut être établi par un groupe d'experts internationaux (e.g. paires gène-pathologie classées « Moderate », « Strong » ou « Definitive » par ClinGen [24], « Vert » par PanelApp [25] ou « Confirmed » dans DDG2P [26]) ou sur une décision consensuelle d'un groupe d'experts national. La description de la meilleure façon d'établir le lien « gène-pathologie » fera objet d'autres recommandations.

Recommandations pour l'interprétation de variants générés par le séquençage à haut débit.

Domaine d'application :

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

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

Conditions préalables à l'utilisation :
• Il est nécessaire de vérifier que le variant à interpréter n'est pas un artéfact de séquençage (pour plus d'information, se référer au BP-GT4 [27].
• Il est nécessaire de s'assurer que le variant à interpréter ne fait pas partie d'un « variant multiple » (MNV ; *Multiple Nucleotide Variant* ou delins) correspondant à deux variants en cis, espacés ou répartis sur quelques nucléotides. Dans ce cas, l'interprétation de l'impact global de ces variants est nécessaire.

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 :

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

L'interprétation des résultats consiste à combiner (cf. Annexe 1) 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

A noter, ***le terme « variant de signification inconnue » utilisé dans la version 1 de ce document a***

été changé en « variant de signification incertaine »*, en corrélation avec la terminologie utilisée par l'ACMG/AMP et en lien avec la problématique réelle de classification.*

L'argumentaire prend en compte :

- Des données épidémiologiques
- Des critères cliniques
- L'existence de variants associés
- Des données de ségrégation
- Des données computationnelles/bioinformatiques
- Des données fonctionnelles

Ces règles de combinaison/d'utilisation exclusive des arguments sont récapitulées dans l'annexe 1.

Introduction de la possibilité d'une « modulation de poids » pour certains arguments :
Dans le cadre de l'évolution des recommandations, la notion de « modulation de poids » a été introduite pour certains arguments. Ainsi, certains items auparavant associés à un poids unique (par exemple PM3 : poids moyen) peuvent actuellement être utilisés avec un niveau de détail supplémentaire en fonction des éléments disponibles, et ainsi une « modulation de poids », avec les conséquences associées à la classification finale du variant (par exemple : le poids « moyen » par défaut de PM3 peut ainsi être modulé en poids fort « PM3_fort » ou faible « PM3_faible »).

Lorsqu'un variant est assigné en classe 3 (variant de signification incertaine), cela ne permet ni de l'utiliser dans la prise en charge éventuelle du patient et de sa famille [28], 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 reclasser 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 réexaminer l'interprétation du variant de signification incertaine à la lumière des nouvelles connaissances acquises le concernant.

L'interprétation finale des variants est sous la responsabilité 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 ont été modifiés pour s'adapter aux positions françaises. Les laboratoires de biologie moléculaire (LBM) sont libres de s'inspirer de tout ou partie de ce document de recommandations ou de procéder différemment.

1-1 Données épidémiologiques

Un élément crucial pour l'interprétation est l'analyse des données épidémiologiques. Le biologiste 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.

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.

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[29]. Deux versions de gnomAD sont disponibles actuellement :

version 2.1.1	15 708 génomes	125 748 exomes	alignés sur GRCh37
version 3.1	76 156 génomes		alignés sur GRCh38

La consultation de gnomAD.v3 peut être intéressante car elle contient les données des populations qui, jusque-là, n'étaient pas bien représentées dans les bases de données. La base exclut 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. Dans gnomAD v2.1.1, il est possible de ne visualiser que les variants issus de cohortes d'individus sains en sélectionnant « controls » dans l'onglet « Dataset » (avec le désavantage de n'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 neurologiques ou « non-cancer » pour les analyses dans le domaine des cancers.

GnomAD peut contenir des variants pathogènes présents à l'état de mosaïque chez les sujets asymptomatiques. Dans ces cas, il est conseillé d'évaluer la possibilité d'un mosaïsme avec l'aide de la fenêtre IGV sur la page du variant. Certains variants pathogènes peuvent aussi n'être présents en mosaïque que dans la lignée sanguine des sujets sains en lien avec le phénomène d'hématopoïèse clonale [30]. 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.
Certaines bases de données inclues dans la version 1 de ce document font maintenant partie de gnomAD (ExAC, ESP (NHLBI exome Sequencing Project), 1000 Genomes).
https://gnomad.broadinstitute.org/

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

Cette base de données contient les variants issus de génomes de 132 345 individus du projet TOPMed (Trans-Omics for Precision Medicine) centré sur les maladies cardiaques, pulmonaires, sanguines et du sommeil [31]. Une partie de données dans cette base est également inclue dans gnomAD (~9% pour la version 5).
https://bravo.sph.umich.edu/freeze8/hg38/

• Allele Frequency Aggregator (ALFA) :

Un projet du NCBI vise à génotyper 1 million d'individus avec une croissance de 100-200 000 individus par trimestre. A ce jour, les variants provenant de 98 000 individus (2 000 exomes, 5 600 génomes et 87 000 SNP-array) sont présents dans cette base de données.
https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/

D'autres bases avec des données de séquençage de la population générale existent (voir Tableau 1)

Tableau 1 : Autres bases de données de séquençage de la population générale

	Nombre d'individus		Référence
FLOSSIES (*Fabulous Ladies Over Seventy*)	10000 (27gènes)	*https://whi.color.com/about*	Boomsma et al. 2014, PMID: 23714750
FREX (*French Exome* Project)	600	*http://med-laennec.univ-brest.fr/FrExAC/*	
The Greater Middle East (GME) Variome Project	2497	*http://igm.ucsd.edu/gme*	Scott et al., 2016, PMID: 27428751
SAGE (South Asian Genomes and Exomes)	1213	*http://clingen.igib.res.in/sage/*	Hariprakash et al. 2018, PMID: 30184194
Japan human genetic variation database	1172	*https://www.hgvd.genome.med.kyoto-u.ac.jp/*	
Medical Genome Reference Bank	2845	*https://sgc.garvan.org.au/*	Pinese et al. 2020, PMID: 31974348
ABraOM: Brazilian genomic variants	1780	*http://abraom.ib.usp.br/*	Naslavsky et al., 2020 doi.org/10.1101/2020.09.15.298026
Wellderly	511		Erikson et al., 2016, PMID: 27114037
2000 Danois	2000		Lohmueller et al. 2013, PMID: 24290377
Genome of the Netherland		*http://www.nlgenome.nl/ (does not work)*	Francioli et al. 2014, PMID: 24974849

De manière générale, on considère qu'un variant dans un gène responsable d'une maladie de transmission autosomique dominante avec une pénétrance complète ne peut pas avoir une fréquence allélique supérieure à la prévalence de la maladie. Un calcul théorique, selon la loi de Hardy-Weinberg pour les maladies récessives, 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/ [32]. Cet outil permet en outre de prendre en compte le mode de transmission, l'hétérogénéité allélique et génétique, ainsi que la pénétrance incomplète de certaines pathologies.

BA1

Fréquence allélique > 5% (ou définie par le laboratoire)

Critère à lui seul en faveur du caractère bénin du variant

BS1

Fréquence allélique trop importante par rapport à la fréquence de la pathologie

A noter, une diminution de seuil de BA1 a été proposée pour un certain nombre de gènes et pathologies (voir Tableau 2).

PM2

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

*Devant l'existence d'un grand nombre de variants très rares mais bénins, les recommandations de ClinGen SVI Working Group (publiées en Septembre 2020 [33]) proposent de diminuer le poids de PM2 en « faible », accompagné par une modification de calcul final de critères (PVS1 + **un PP** = classe 4 ou lieu de PVS1 + **un PM** = classe 4). Comme d'autres modifications futures de calcul ont été évoquées dans ces recommandations, notre GT a décidé que l'implémentation de diminution de PM2 en poids « faible » est encore prématurée.*

Précautions :

• Il est important de veiller à l'existence d'un minimum de 1000 allèles dans la sous-population contrôle considérée. Certains auteurs préconisent un minimum de 2000 allèles [34–36].

• Pour les maladies de transmission autosomique récessive, la fréquence allélique de certains variants pathogènes peut être >1%. Des listes de variants pathogènes ayant une fréquence allélique élevée (listes d'exceptions) ont déjà été élaborées pour certains gènes par les groupes de travail gène-spécifique [8,16,37]. D'autres variants pathogènes avec des fréquences alléliques très élevées peuvent être proposés à ClinGen [38].

• 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.

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

• Certaines bases de données peuvent contenir des variants pathogènes présents à l'état de mosaïque chez les sujets asymptomatiques.

• 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. Une évaluation de la qualité de séquençage est également conseillée en cas d'un variant pathogène retrouvé dans la population générale pour savoir s'il s'agit d'un faux-positif.

• Certains variants de type insertion/délétion de 1-50 nucléotides peuvent être absents des bases de données car ils sont mal annotés ou mal détectés par séquençage à haut débit.

• La fréquence allélique du variant dans la population d'origine du patient peut ne pas être connue.

• Il existe une redondance de données entre certaines de ces bases de données, notamment gnomAD et BRAVO (4% d'échantillons dans gnomADv.2 et 9% d'échantillons dans BRAVO sont présents dans les deux bases de données).

• Dans la publication princeps de l'ACMG/AMP[3], 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 SNV/indel 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 décrits 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 est indispensable d'utiliser la véritable fréquence allélique du variant et non la MAF telle qu'elle est classiquement définie. C'est la raison pour laquelle nous utilisons la fréquence allélique dans les présentes recommandations.

• En 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 classer est retrouvé avec une fréquence très élevée dans une sous-population (sans 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. Il est aussi important de noter la présence dans la population générale des individus porteurs du variant étudié à l'état homozygote.

• Il est conseillé d'adapter les seuils de fréquence pour attribuer les arguments PM2, BS1 ou BA1 en fonction du gène et de la pathologie.

A titre d'exemple :

Tableau 2. Les seuils de fréquences pour les arguments PM2/BS1/BA1 issus de dix groupes de travail gène-spécifiques.

	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_faible								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_faible									> 0,07%		
BA1	> 0,1%	> 0,05%	> 1%	> 0,1%	> 0,2%	> 0,76%	> 0,15%	> 0,1%	> 0,5%	> 1,5%	> 1%

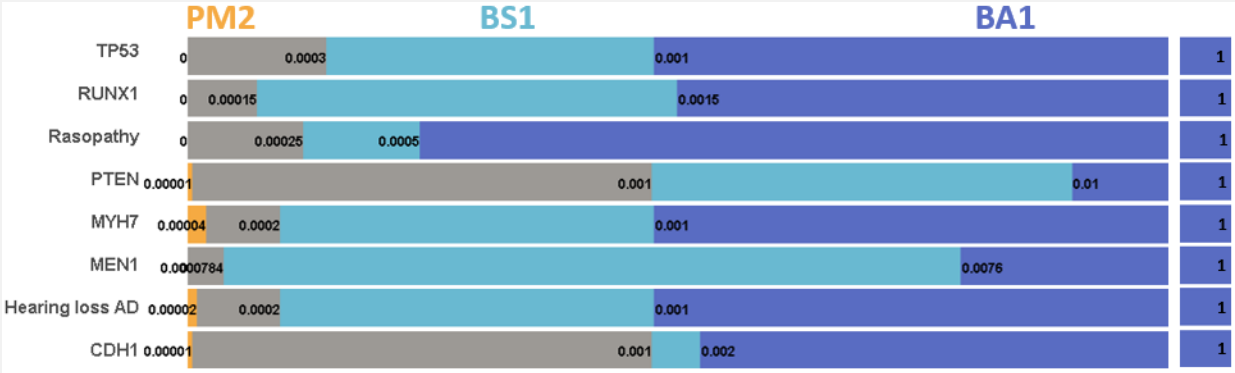

Figure 1. Une représentation graphique des seuils de fréquences pour les arguments PM2/BS1/BA1 issus de dix groupes de travail gène-spécifiques.

La Figure 1 et le Tableau 1 résument les seuils de fréquences d'un variant dans la population générale pour l'attribution des arguments PM2/BS1/BA1 selon les recommandation de dix groupes de travail gène-spécifiques : MYH7 [12], Rasopathie [11], PTEN [15], Surdités [16], CDH1 [13], MEN1 [17], RUNX1 [14], TP53 [19], PAH [18], GAA [20].

1-1-2 Données de fréquence des variants chez les patients

Il est nécessaire de consulter les bases de données de patients telles que ClinVar, LOVD, Decipher, Varsome, HGMD et COSMIC. Dans la mesure du possible, une base de données locus-spécifique (si possible avec consensus par thématique) pourra être consultée. La diversité des bases consultées et leur fiabilité sont à prendre en compte. Parallèlement, il est important de consulter 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 homogène pour tous les patients porteurs de ce variant.

Il est également utile et recommandé 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 bibliographiques pour le variant. Pour compléter cette liste, il est possible d'utiliser PubMed (NCBI, *National Center for Biotechnology Information*), Google Scholar, LitVar [39], AVADA (Automatic Variant evidence DAtabase) [40], 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 [41,42].

Il est possible de contacter les laboratoires ayant une expertise sur le gène en question.

PS4 PS4_moyen PS4_faible

• le variant est présent chez différents patients non-apparentés ET la fréquence allélique de ce variant dans la population générale est compatible avec la prévalence et la pénétrance de la maladie (BS1 n'est pas attribué à ce variant)

PS4

Prévalence du variant chez les individus atteints significativement supérieure à celle des contrôles

• études cas-témoins avec Odd Ratio > 3, borne inférieure de l'intervalle de confiance excluant 1

En l'absence d'études cas-contrôles, il est possible de moduler le poids de l'argument PS4 en fonction du nombre de patients non-apparentés confirmés avec ce variant : PS4, PS4_moyen (argument modulé en poids modéré, PM), PS4_faible (argument modulé en poids faible, PP). Le nombre de patients nécessaires doit être défini en fonction du gène et de la pathologie étudiés. Pour compter le nombre de patients porteurs du variant, il est nécessaire de consulter l'ensemble des bases données de patients (ClinVar, LOVD, bases spécifique) et faire une recherche bibliographique extensive. Le nombre de patients porteurs du variant en question n'inclut pas le patient en cours d'analyse.

Il est proposé de n'utiliser l'argument PS4 que pour les gènes responsables des pathologies dominantes et d'utiliser l'argument PM3 pour les pathologies récessives [4,16].

Cette façon d'attribuer PS4 n'est possible que si le variant à classer est extrêmement rare dans la population générale. Pour des maladies à pénétrance incomplète ou à révélation tardive, le poids PM2 peut être attribué à un variant même quand il existe quelques occurrences de ce variant en population générale. Ainsi, pour ces variants, il est possible de combiner les scores PM2 et le PS4/PS4_moyen/PS4_faible.

A titre d'exemple :

Tableau 3. Seuils pour l'attribution de l'argument PS4 issus de huit recommandations gène-spécifiques.

	MYH7	Rasopathie	PTEN	Surdités AD	CDH1	MEN1	RUNX1	TP53
PS4_très_fort			> 16		> 16			
PS4_fort	> 15	> 5	4-15	> 15	4-15	étude cas-témoin	> 4	> 4
PS4_moyen	6-14	3-4	2-3	6-14	2-3	3	2-3	2-3
PS4_faible	2-5	1-2	1	2-5	1	2	1	1

Ces recommandations gène-spécifiques s'appuient sur les nombres de patients porteurs du variant en question nécessaires pour modifier le poids de l'argument PS4. Ce tableau résume les recommandations récentes pour l'interprétation des variants dans les gènes responsables de maladies dominantes : MYH7 [12], Rasopathie [11], PTEN [15], Surdités [16], CDH1 [13], MEN1 [17], RUNX1 [14], TP53 [19]. Pour les variants dans les gènes responsables de maladies récessives, il est conseillé d'utiliser l'argument PM3. Selon les recommandations pour PTEN et TP53, les évaluations phénotypiques sont aussi calculées en parallèle, avec nécessité d'avoir plus de patients pour un niveau de PS4 si les phénotypes des patients sont moins spécifiques.

Précautions :

• Il est possible que le même patient soit décrit dans plusieurs sources (e.g. PubMed et ClinVar ou LOVD). Il est conseillé de vérifier que le même patient ne soit pas compté plusieurs fois, entrainant un 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. De plus, dans certains de ces articles, la nomenclature du variant peut être incorrecte (ancienne nomenclature vs. nomenclature HGVS). Certaines bases de données peuvent intégrer ces variants 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.

Suite à la recommandation de ClinGen [6], les critères PP5 (*Source documentée classant le variant comme pathogène*) et BP6 (*Source documentée classant le variant comme bénin*) 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 dans l'attribution des arguments PS4 et PM3. Ce document (BP-NGSDiag_001_v2) laisse la possibilité aux laboratoires d'utiliser ces deux arguments, mais recommande fortement d'utiliser les arguments PS4 et PM3 à la place.

1-1-3 Bases de données locales

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 nationales et internationales.

1-2 Critères cliniques

Au-delà des informations cliniques indispensables lors de la demande d'analyse, une confrontation clinico-biologique est nécessaire. Elle peut être établie à partir des informations cliniques disponibles auprès du biologiste, fournies par le clinicien lors de la prescription. Dans certaines situations, ces informations sont suffisantes pour l'interprétation du variant. Dans d'autres cas, une Réunion de Concertation Pluridisciplinaire (RCP) avec le généticien clinicien et/ou clinicien spécialiste prenant en charge le patient peut être recommandée afin de préciser la corrélation phénotype-génotype (pour plus d'information sur les RCP, se référer au document de la HAS [43]).

PP4_moyen PP4

Phénotype ou histoire familiale spécifique en faveur de la pathologie associée au gène comprenant le variant.

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. Idéalement, les phénotypes spécifiques attendus pour chaque gène doivent être définis par le laboratoire [16]. Pour ces phénotypes très spécifiques, il est possible de moduler le poids de l'argument. Par exemple, l'argument PP4_moyen (PP4 renforcé en PM) peut être attribué dans le cadre d'une maladie métabolique si le dosage enzymatique correspondant est compatible avec l'anomalie du gène avec le variant étudié. Un autre exemple de données spécifiques au patient analysé est une absence de la protéine sur Western blot ou Immunohistochimie de la biopsie.

A noter, pour les données se rapportant au variant, c'est l'argument PS3 (dans le chapitre « données fonctionnelles ») qui sera utilisé (et non PP4). Ainsi, les résultats d'analyse par minigène sont utilisés pour l'attribution de l'argument PS3 et non PP4.

Précautions :

• Certains phénotypes sont peu spécifiques (e.g. prédisposition héréditaire au cancer). De plus, certaines pathologies explorées par des panels de gènes peuvent avoir une hétérogénéité génétique considérable (e.g. déficience intellectuelle non syndromique). 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 d'analyse est élevée. En revanche, lorsqu'un grand panel ou un séquençage d'exome est utilisé pour une maladie à forte hétérogénéité génétique, la probabilité d'identifier le variant causal est la somme de toutes les probabilités pour chaque gène individuellement, accompagné par une réduction de la spécificité. 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 récentes n'utilisent plus cet argument dans la classification [11–15,19].

1-3 Co-occurrence de variants

PM3_fort **PM3** **PM3_faible**
pour une pathologie à transmission autosomique récessive, un variant observé en *trans* avec un variant pathogène sur le même gène

Conditions pour l'utilisation :
• L'argument PM3 n'est valable que pour les variants avec une fréquence allélique compatible avec la prévalence et la pénétrance de la maladie (donc à condition que BS1 ne soit pas attribué).

Les nouvelles recommandations du groupe de travail ClinGen SVI proposent le système de calcul suivant pour le score PM3 [9]. Il devient ainsi possible de renforcer l'argument en poids fort (PS) si le nombre de patients porteurs de ce variant est important (PM3_fort) ou de l'abaisser en poids faible (PP) si ce nombre est peu important (PM3_faible).

Tableau 4 : Calcul du score PM3 pour un variant observé en *trans* avec un variant pathogène dans la situation d'une pathologie de transmission autosomique récessive.

Classification/Zygotie du variant en face	Points par Cas Index	
	Confirmé *en trans*	Phase inconnue
Pathogène (classe 5) Probablement pathogène (classe 4)	1	0,5 0,25
Variant de Signification Incertaine (classe 3) *(maximum 0,5 point par variant)*	0,25	0
Variant homozygote *(maximum 1 point par variant)*	0,5	

PM3_fort	2+ points
PM3	1 point
PM3_faible	0.5 point

Si un variant a été observé chez de nombreux cas index, chacun avec un variant pathogène différent associé, 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/AMP initiaux).

• Pour une pathologie de pénétrance incomplète et de transmission autosomique dominante, la co-occurrence de 2 variants *en trans* dans un gène ou la co-dominance de 2 gènes peut être corrélée à l'âge de début et la sévérité du phénotype (ex : Cardiomyopathies).

• 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/AMP 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/AMP initiaux).

Précautions :

• Dans certains gènes, des transmissions dominantes et récessives ont été décrites (e.g. CLCN1/myotonie congénitale, BRCA2/cancer du sein et de l'ovaire/Fanconi-like).

• Le phénotype du patient peut être secondaire aux anomalies dans deux gènes différents (diagnostics doubles) [44–47].

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

1-4 Données de ségrégation

PS2_très_fort **PS2** **PM6** **PM6_faible**

Variant *de novo* chez un patient sans antécédent familial
(le poids de l'argument est à attribuer en fonction de la confirmation de la paternité/maternité et en fonction de la spécificité du phénotype).

Les nouvelles recommandations de groupe du travail ClinGen SVI proposent le système de calcul des scores PS2/PM6 [10] suivant :

Tableau 5 : Calcul du score PS2/PM6 pour un variant survenu *de novo*.

Le phénotype est …	Points par Cas Index	
	Confirmé *de novo*	Supposé *de novo*
… très spécifique pour le gène	2	1
… compatible pour le gène mais pas très spécifique	1	0,5
… compatible pour le gène mais pas très spécifique + une hétérogénéité génétique*	0,5	0,25
… n'est pas compatible avec le gène	0	0

**maximum 1 point par variant*

PS2_très_fort 4 points
PS2 2 points
PM6 1 point
PM6_faible 0.5 point

• 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 cohérente avec le phénotype

PP1_fort **PP1_moyen** **PP1**

Variant co-ségrégeant avec la pathologie chez plusieurs membres atteints dans la même famille et/ou dans plusieurs familles (le poids de l'argument est à attribuer en fonction du nombre de ségrégations).

Plus le nombre de sujets analysés attestant de la co-ségrégation du génotype avec le phénotype est important plus l'argument peut avoir un poids élevé. Le poids faible de PP1 peut alors être augmenté en PM (PP1_moyen) ou PS (PP1_fort) ou rester un poids faible (PP1). Une façon de préciser ce critère est d'utiliser 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 ($N = (1/2)^m$, « m » étant le nombre de méioses informatives concernant la co-ségrégation) [48]. Il est aussi d'utiliser le calcul de LOD score simplifié [22] qui est directement lié à la probabilité « N ». Cette approche a été adoptée par les groupes ClinGen (les instructions détaillées sont disponibles à [49]). Le poids de l'argument PP1 est ensuite attribué en fonction de la probabilité N ou LOD score.

Par exemple, pour une maladie dominante, 5 ségrégations de la maladie avec le variant étudié correspondent à N ≤ 1/32 et LOD score = 1.5, permettant d'attribuer le poids PP1_fort pour ce variant. Pour une maladie récessive, si en plus du patient porteur du variant il y a deux autres membres de la famille atteints ET porteurs du variant, le poids PP1_moyen peut être attribué.

Maladies AR :

Atteints*

Non-atteints	0	1	2	3	4	5	6	7	8	9	10
0	0	0.6	1.2	1.81	2.41	3.01	3.61	4.21	4.82	5.42	6.02
1	0.12	0.73	1.33	1.93	2.53	3.14	3.74	4.34	4.94	5.54	6.15
2	0.25	0.85	1.45	2.06	2.66	3.26	3.86	4.46	5.07	5.67	6.27
3	0.37	0.98	1.58	2.18	2.78	3.39	3.99	4.59	5.19	5.79	6.4
4	0.5	1.1	1.7	2.31	2.91	3.51	4.11	4.71	5.32	5.92	6.52
5	0.62	1.23	1.83	2.43	3.03	3.63	4.24	4.84	5.44	6.04	6.65
6	0.75	1.35	1.95	2.56	3.16	3.76	4.36	4.96	5.57	6.17	6.77
7	0.87	1.48	2.08	2.68	3.28	3.88	4.49	5.09	5.69	6.29	6.9
8	1	1.6	2.2	2.81	3.41	4.01	4.61	5.21	5.82	6.42	7.02
9	1.12	1.73	2.33	2.93	3.53	4.13	4.74	5.34	5.94	6.54	7.15
10	1.25	1.85	2.45	3.06	3.66	4.26	4.86	5.46	6.07	6.67	7.27

$$Z\ (\text{LOD score}) = \log_{10} \frac{1}{(0.25)^{\#\ \text{atteints}-1} \times (0,75)^{\#\ \text{non-atteints}}}$$

Maladies AD / liées à l'X :

Ségrégations

2	3	4	5	6	7	8	9	10
0.6	0.9	1.2	1.5	1.8	2.1	2.4	2.7	3.0

$$Z\ (\text{LOD score}) = \log_{10} \frac{1}{(0.5)^{\text{Segregations}}}$$

PP1_fort
PP1_moyen
PP1

** en excluant le cas faisant l'objet de l'analyse diagnostique ; tous les sujets atteints doivent être porteurs des variants analysés*

Figure 2 : Calcul du Z-score en fonction du nombre de sujets atteints et non atteints pour les maladies de transmission autosomique réces²sive (AR) et pour maladies de transmission autosomique dominantes (AD) ou liées à l'X. Le code couleur, fonction du Z-score, indique le poids à accorder à l'argument : en beige, argument PP1 ; en orange, argument PP1-modéré ; en rouge, argument PP1 fort.

BS2

Variant observé chez un individu sain pour une maladie récessive (variant homozygote ou variants hétérozygotes composites), dominante (variant hétérozygote) ou récessive liée à l'X (variant hémizygote chez un garçon), avec une pénétrance complète attendue à un âge précoce.

BS4

Variant ne ségrégeant pas avec la pathologie chez des apparentés atteints.

Précautions :
• Seuls des individus bien phénotypés peuvent être considérés pour l'attribution du l'argument BS2 (ne pas utiliser les données de population).

• Attention aux phénocopies pouvant mimer un défaut de ségrégation.

• Attention aux pathologies à variabilité d'expression phénotypique et de pénétrance incomplète : lorsqu'un variant est hérité d'un parent apparemment asymptomatique, une réévaluation 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 et d'effectuer une étude de biais d'inactivation de l'X chez les femmes hétérozygotes.

• Attention aux défauts de ségrégation dus aux fausses paternités, mosaïques somatiques, aux disomies uniparentales ou au phénomène d'empreinte parentale ainsi qu'aux variants *de novo* pour des pathologies récessives.

1-5 Données prédictives et computationnelles

Variant avec un effet nul prédit:

PVS1 **PVS1_fort** **PVS1_moyen** **PVS1_faible**
Variant ayant un effet nul prédit (non-sens, frameshift, site canonique d'épissage (±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 un mécanisme pathogène connu.

Conditions pour l'utilisation :
• L'argument PVS1 n'est valable que pour les gènes avec un lien « gène-pathologie » bien établi (voir Note d'utilisation : **Niveau de preuve pour le lien** « gène-pathologie » sur la page 4).
• L'argument PVS1 n'est valable que pour les gènes où la perte de fonction est un mécanisme pathogène connu. Il existe plusieurs bases de données où cette information est disponible :
- ClinGen Dosage [50] – l'haploinsuffisance de chaque gène est évaluée par des groupes d'experts[51]
- Gene2Phenotype [52] (aussi disponible sur le site de DECIPHER [26,53])

Si l'intolérance d'un gène aux variants de type « perte de fonction » n'a pas encore été évaluée par un groupe d'experts, elle peut être estimée en utilisant les scores pLI (*probability of Loss of function Intolerance*) et o/e (*observed/expected*) disponibles dans gnomAD[29]. Le score pLI est largement utilisé depuis plusieurs années, même si ce score comporte un certain nombre des 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.) [54]. 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 LOEUF) 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[29]. Pour un gène donné, ce score correspond au ratio entre le nombre de variants « perte de fonction » observés dans la population et le nombre de variants attendus dans ce gène (*observed/expected*). Un gène est considéré comme intolérant à la perte de fonction si la borne supérieure d'intervalle de confiance du score o/e est < 0.35. L'utilisation du HI index (*Haploinsufficiency Score*) de la base de données DECIPHER [26] présente le même intérêt et les mêmes limites mais sur un jeu de données indépendant[55]. Une analyse qualitative de la charge en variants tronquants dans les bases de données est utile afin d'aborder ces scores de manière critique. L'utilisation de la fréquence des variants tronquants dans gnomAD est moins adaptée pour estimer la sensibilité à la perte de fonction pour les gènes responsables des maladies avec une pénétrance incomplète ou tardive. En effet, les variants tronquants pathogènes responsables de ces maladies peuvent avoir des fréquences plus élevées dans la population générale par rapport aux variants responsables des maladies pédiatriques sévères.

- PVS1 ne peut pas être combiné avec l'argument PP3 (*Effet d'un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés*).

Attribution de l'argument PVS1 pour les variants ayant un effet nul prédit

Un arbre décisionnel a été proposé afin d'attribuer l'argument PVS1 pour les variants ayant un effet nul prédit : non-sens, frameshift, site canonique d'épissage (±1,2), codon d'initiation de la traduction (adapté de [5]).

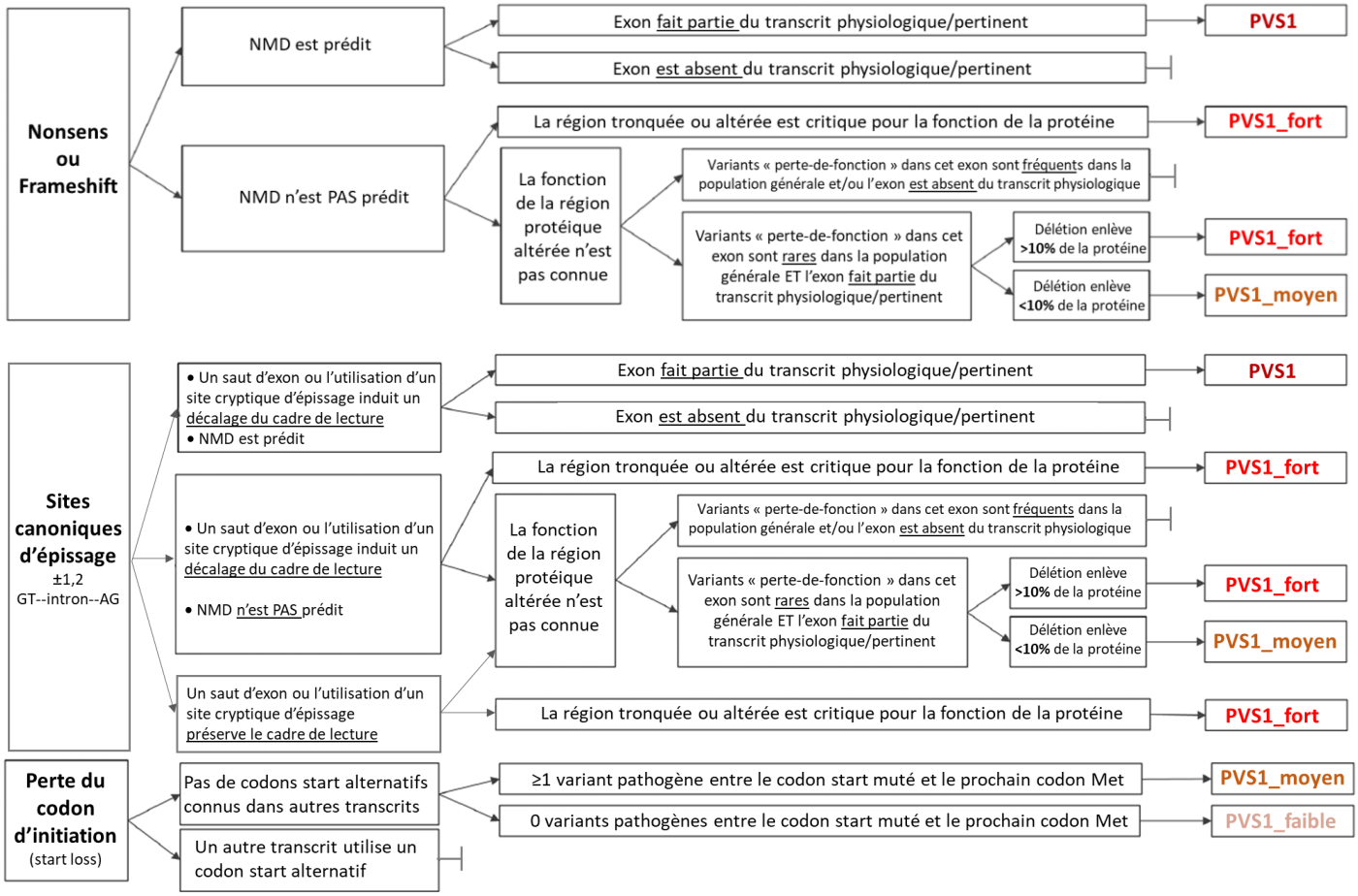

Figure 3 : arbre décisionnel pour l'attribution de l'argument PVS1 pour les variants ayant un effet nul prédit (adapté de [5]).

Attribution de l'argument PVS1 pour les CNV intragéniques

Un arbre décisionnel a été proposé afin d'attribuer l'argument PVS1 pour les CNV intragéniques (adapté de [5]) :

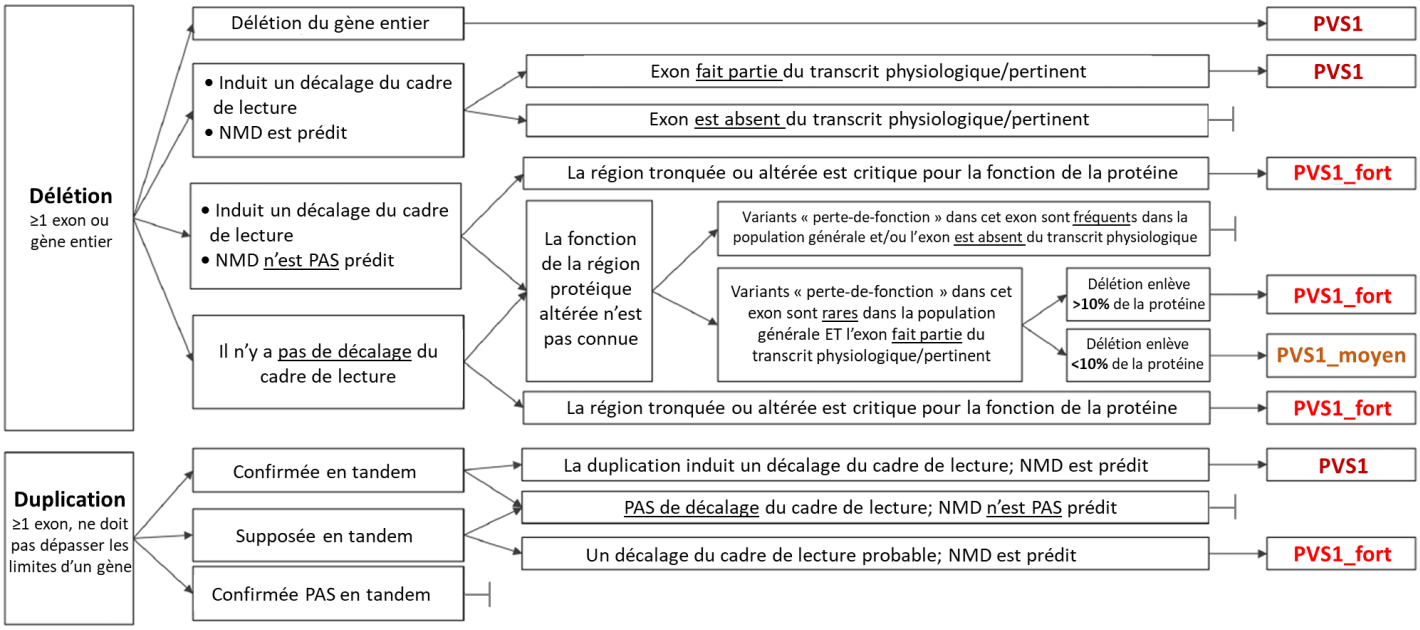

Figure 4 : arbre décisionnel pour l'attribution de l'argument PVS1 pour les CNV intragéniques (adapté de [5]).

Pour aider à l'attribution de PVS1, un outil de classification automatique est maintenant disponible[56,57].

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. L'information sur les transcrits et les données d'expression par exon peut être obtenue dans la littérature ou dans les études d'expression à grande échelle tel que le *Genotype-Tissue Expression Project (GTEx)* [58–60].

• Attention à l'apparition de nouveaux sites cryptiques d'épissage (par exemple des variants d'épissage rajoutant un ou plusieurs codons résultant en une indel en phase). Avant d'interpréter les variants sur les sites canoniques d'épissage, il est conseillé de rechercher les sites cryptiques forts autour du variant (±20nt).

• Attention au saut d'un ou de plusieurs exon(s) en phase, laissant intact le reste de la protéine.

• Pour les variants affectant les codons d'initiation de la traduction, il est important de rechercher l'existence de codons d'initiation alternatifs éventuels (analyse ORFinder [61]).

• Des variants dans la région 5'UTR peuvent introduire un codon d'initiation avec un changement du cadre de lecture, conduisant à une perte de fonction du gène [62].

• Pour les variants de type « perte de codon stop », il est important de rechercher l'existence de codons stop alternatifs éventuels (analyse ORFinder [61]). S'il n'y a pas de codons stop dans la région 3'UTR, le transcrit est prédit d'être soumis au NSD (*nonstop mediated decay*) [63], permettant

l'attribution de l'argument PVS1. Si un codon stop alternatif est identifié, l'argument PM4 (changement de la taille de la protéine) est plus adapté.

• Il est important de rechercher un éventuel effet sur l'épissage pour tout variant, qu'il soit intronique ou exonique (comprenant les variants synonymes, non-sens, frameshift, faux-sens, indel).

Variant affectant la longueur de la protéine:

PM4
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.

BP3
Variant affectant la longueur de la protéine, si l'altération est en phase dans une région répétée sans fonction connue.

Type et contexte du variant :

PP2
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.

BP1
Variant faux-sens dans un gène où seuls les variants tronquants sont décrits comme associés à la pathologie.

BP7
Variant synonyme sans impact prédit sur l'épissage **ET** pour lequel la séquence nucléotidique n'est pas très conservée chez les vertébrés.

Précautions :
• Avant d'interpréter la conservation nucléotidique, il est important de vérifier que la représentation de génomes connus est appropriée. Il est à noter également que la dégénérescence du code génétique impliquant la 3ème base du codon induit une conservation souvent 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*).

Localisation du variant :

PS1
Variant à l'origine du même changement d'acide aminé qu'un variant pathogène connu, sauf si le nouveau variant a un possible effet sur l'épissage.
Le poids de l'argument peut être modulé en « modéré » si le variant connu est classe 4.

PM5
Variant à l'origine d'un changement d'acide aminé différent à la même position qu'un variant faux-sens pathogène connu.

Le poids de l'argument peut être modulé en « faible » si le variant connu est classe 4

Il est important de vérifier que le variant « pathogène connu » est classé selon la classification ACMG/AMP.
Un changement pour un acide aminé similaire au variant pathogène déjà décrit (Gratham score bas) a plus de risque d'être également pathogène.

PM1
Variant non-tronquant situé sur un hot spot mutationnel et/ou un domaine fonctionnel critique bien établi **exempt de variants bénins.**

Prédictions bio-informatiques :

PP3
Effet d'un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés.
BP4
Effet d'un variant prédit bénin par l'ensemble des logiciels de prédiction de pathogénicité interrogés.

Conditions pour l'utilisation :
• PP3 ne peut pas être combiné avec l'argument PVS1.

Choix des logiciels de prédiction des variants faux-sens

En cas de variant prédit faux-sens, des logiciels de prédiction de pathogénicité sont utilisés pour estimer l'impact du variant sur la fonction protéique. De nombreux outils de prédiction de la pathogénicité sont recensés dans la base de données VIPdb [64,65]. Des algorithmes dits « **non-intégrés** » attribuent un score et une classification en faveur du caractère délétère ou du caractère bénin d'un variant faux-sens en fonction de différents critères : conservation de la base nucléotidique et de l'acide aminé dans l'évolution, distance physico-chimique entre l'acide aminé d'origine et muté (score de Grantham [66]), position dans un domaine fonctionnel connu de la protéine. Les outils classiquement utilisés sont : SIFT [67], Polyphen-2 [68], Mutation Taster [69], Mutation Assessor [70], FATHMM[71], UMD-Predictor [72] ainsi que Align GVGD [73] pour certains gènes.

Plus récemment, des scores appelés « **intégrés** » (e.g. Condel [74], CADD [75], MataSVM et MetaLR [76], MCAP [77], REVEL [78], Eigen [79]) ont été mis au point afin d'améliorer la performance en regroupant les prédictions d'autres algorithmes. Certains de ces logiciels ont l'avantage de permettre des prédictions pour des variants exoniques et non exoniques. En revanche, comme tous les logiciels de prédiction bioinformatique, ils sont dépendants des variants utilisés lors de l'entraînement (*machine learning*) pour les mettre au point ainsi que du type de modèle d'apprentissage.

Des études récentes ont testé un certain nombre d'outils pour l'application en diagnostic et ont identifié plusieurs algorithmes de haute performance qui ne semblent pas avoir été intégrés dans les pipelines des laboratoires hospitaliers [80,81]. Ces études ont aussi proposé que certains algorithmes « intégrés » puissent être plus performants que l'accumulation de scores individuels. En effet, plusieurs recommandations gène-spécifiques attribuent maintenant l'argument PP3 selon un seul score de prédiction (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 [78]. Pour

l'attribution de PP3, les seuils de REVEL proposés varient entre >0.6 et >0.75 [14,16,34,78]. Certaines recommandations proposent aussi l'attribution de l'argument pour le caractère bénin (BP4) pour des variants avec un score REVEL bas (<0.15 ou <0.5).

A ce jour, il n'existe pas de consensus sur le choix de l'algorithme de prédiction ou une combinaison des outils bio-informatiques pour attribuer l'argument PP3. Si les résultats des outils de prédiction sont combinés, il est important d'estimer le niveau de redondance entre les logiciels afin de ne pas avoir un effet de surprédiction préjudiciable au résultat final. Idéalement, la performance des outils de prédiction devrait être testée pour des gènes analysés par les laboratoires en routine car il peut y avoir des différences entre les prédictions en fonction des gènes. Pour aider dans la démarche d'évaluation des différents scores de prédiction, il est possible d'utiliser Varsome [82] qui affiche les résultats de 21 outils intégrés et non-intégrés ainsi que le calcul automatisé de PP3 basé sur l'utilisation de certains de ces logiciels. Variant Score Ranker [83] ou ClinGen Pathogenicity Calculator [84] fournissent également les scores de prédiction pour les variants faux-sens.

Choix des logiciels de prédictions du retentissement sur l'épissage

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 pour les variants non-sens, faux-sens et synonymes. Des outils classiques, tels que MaxEntScan [85], SpliceSiteFinder [86], NNSPLICE [87], GeneSplicer [88] et Human Splicing Finder [89] sont encore largement utilisés en diagnostic [90]. Les approches plus récentes, telles que SpliceAI [91] , S-CAP [92], SPiCE [93] et MMSplice [94], commencent également à prendre de l'importance.

En cas de prédiction d'un effet délétère sur l'épissage, 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. Les arguments PS3 ou PVS1 peuvent être ensuite attribués en fonction de la fiabilité du résultat du test fonctionnel et de l'effet sur le transcrit (perte de fonction etc).

Le document *BP-ANPGM_004* décrit les étapes nécessaires pour évaluer l'impact d'un variant sur l'épissage [95] (la nouvelle version de ce document est en cours de développement).

Précautions **:**

• Il n'est pas possible d'exploiter les résultats de logiciels de prédiction 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.

• Même si la majorité des outils de prédiction pour les variants faux-sens semblent avoir la même performance pour les variants perte-de-fonction et les variants gain-de-fonction [80], il a été montré que certains logiciels sont moins performants pour les variants de type gain-de-fonction [96,97].

• Prudence avec les outils de prédiction de variants d'épissage : autant les outils sont fiables pour la création de sites cryptiques et l'abolition de sites d'épissage, autant les prédictions d'éléments enhancer ou silencer d'épissage ne sont pas fiables à ce jour. Certains outils de prédictions peuvent être également moins sensibles à la détection des évènements d'épissage atypiques, tel que microexons ou introns de type U12 [98,99].

• 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.

1-6 Données fonctionnelles

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

• Il est fortement conseillé que les tests fonctionnels utilisés aient été validés par les experts du gène analysé [7].

• Il convient de distinguer les études fonctionnelles qui apportent du poids à l'échelle du gène ou à l'échelle du variant. Les études fonctionnelles qui apportent du poids à l'échelle du variant peuvent utiliser l'argument PS3. Quant aux études fonctionnelles qui apportent du poids à l'échelle du gène, elles utiliseront l'argument PP4 (*Phénotype du patient*). Ainsi, pour une absence de la protéine sur Western blot ou un profil enzymatique altéré pour l'enzyme codée par le gène dans lequel le variant d'intérêt a été détecté, on attribuera un argument PP4 tandis que pour un test fonctionnel sur un variant exprimé dans une lignée cellulaire, on attribuera un argument PS3.

• Pour les variants avec un effet sur l'épissage, l'argument PS3 peut être attribué pour les analyses par minigene ainsi que les analyses par RT-PCR sur les prélèvements du patient.

• Si l'impact du variant localisé sur les sites non-canoniques d'épissage est confirmé par les études fonctionnelles et prédit d'entrainer une perte de fonction du gène, il est conseillé de suivre l'arbre décisionnel de l'attribution de PVS1 au lieu d'appliquer PS3 (ne pas combiner PVS1 et PS3).

• Si un effet partiel sur l'épissage a été mis en évidence, la modulation du poids de l'argument PS3 peut être discuté par le groupe de spécialistes de la pathologie.

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

1-7 Analyses somatiques

Des recommandations ACMG/AMP spécifiques intègrent les analyses de variants somatiques [100]. Chaque FNMR/réseau d'oncogénétique est amené à adapter ces recommandations selon leur besoin dans l'interprétation. L'interprétation de variants constitutionnels identifiés peut intégrer les résultats d'analyses somatiques (exemples dans le cadre des prédispositions au cancer avec les analyses tumorales: recherche d'instabilité microsatellitaire et marquage immunohistochimique dans le syndrome de Lynch ou marquage immunohistochimique dans les paragangliomes héréditaires). La réalisation de ces analyses dépend de la disponibilité du matériel tissulaire. De plus, l'identification d'un second événement tumoral peut constituer un argument de pathogénicité du variant constitutionnel (autre variant pathogène (PS), LoH (PM) ou méthylation du promoteur (PM)). Le poids de l'argument accordé à ces analyses tumorales doit être défini au sein des groupes de travail par pathologie. Les laboratoires d'oncogénétique constitutionnelle doivent travailler en étroite collaboration avec les laboratoires d'oncogénétique somatique/tumorale pour constituer des bases de données communes de variants alléliques. Un variant pathogène détecté uniquement au niveau tumoral pourrait être intégré à la réflexion et participer aux arguments à prendre en compte dans la classification de variants constitutionnels. En revanche, la perte d'hétérozygotie (LoH) n'est pas un marqueur fiable de la pathogénicité d'un variant pour tous les gènes. En effet, la perte de certaines régions chromosomiques est fréquente dans les tumeurs de survenue sporadique.

1-8 Analyse des variants de l'ADN mitochondrial

Les maladies mitochondriales sont très hétérogènes et sont secondaires à des variants pathogènes localisés soit dans des gènes nucléaires, soit dans des gènes de l'ADN mitochondrial (ADNmt).

Les modalités d'interprétation des variants nucléaires suivent les recommandations précédemment décrites dans le document général. Des spécificités pour plusieurs gènes nucléaires ont été décrites dans un document élaboré par le panel d'experts « ClinGen Mitochondrial Disease Nuclear and Mitochondrial Variant Curation Expert Panel » [101].

Des recommandations ont également été élaborées par ce groupe concernant les variants de l'ADN mitochondrial selon les critères ACMG/AMP appliqués aux particularités de l'ADNmt [102]. Des recommandations spécifiques pour les ARN de transfert (ARNt) sont également disponibles [103].

Les principales spécificités pour l'ADNmt à prendre en compte sont :

- L'hérédité maternelle : l'ADNmt est transmis par les mères.
- La notion d'hétéroplasmie : on ne parle pas d'homozygotie/hétérozygotie pour les variants de l'ADNmt. En revanche la proportion d'ADNmt muté, c'est-à-dire le taux d'hétéroplasmie, peut être différent suivant les mitochondries, les cellules ou les tissus, les tissus affectés ayant un taux d'hétéroplasmie plus fort.
- L'absence d'épissage et d'introns à l'exception d'une région non codante appelée D-Loop d'environ 1 kb.
- La plupart des variants affectant des protéines mitochondriales sont des variants faux-sens.
- Une séquence de référence Cambridge Reference Séquence (CRS) NC_012920.1 [104].

Les bases de données de populations utilisées pour l'ADNmt, afin de déterminer la fréquence du variant mis en évidence dans la population générale, sont principalement :

-MITOMAP : *https://www.mitomap.org/MITOMAP*

-HmtDB : *https://www.hmtdb.uniba.it/*

La fréquence du variant doit également tenir compte de l'haplogroupe du patient. Pour les classifications d'haplogroupes :

-HaploGrep2 *https://haplogrep.i-med.ac.at/category/haplogrep2/*

-Haplofind *https://haplofind.unibo.it/*

Il existe de nombreux outils d'analyse de variants, parmi les plus utilisés

-Mitimpact *http://mitimpact.css-mendel.it/*

-MITOMASTER *https://www.mitomap.org/foswiki/bin/view/MITOMASTER/WebHome*

-Pour les ARNt : mitoTIP *https://www.mitomap.org/foswiki/bin/view/MITOMAP/MitoTipInfo*

Les différents outils pouvant être utilisés dans les maladies mitochondriales sont recensés [105,106].

1-9 Interprétation des variants structuraux (SV), dont CNV

L'interprétation des CNVs intragéniques est décrite dans la section sur l'argument PVS1.

L'interprétation des grands CNV et des CNV récurrents est décrite dans le document récemment actualisé par le réseau AChroPuce [107].

1-10 Modifications de l'argumentaire par rapport aux recommandations

Modifications entre les recommandations initiales l'ACMG/AMP et NGSDIAG_001 v1 :

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/AMP).
- Ajout de l'actualisation selon Jarvik et Browning [48] 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 n'existaient pas dans les recommandations ACMG/AMP. Chaque FNMR/réseau d'oncogénétique est amenée à déterminer le poids de ces arguments dans l'interprétation.

Modifications depuis la version 1 du document NGSDIAG_001 :

- Mise à jour des informations sur les bases de données et les outils de prédiction bio-informatiques
- Formalisation de la possibilité de moduler le poids des arguments
- Ajout des précisions sur l'attribution des arguments PVS1 (un arbre décisionnel), PM3, PS2/PM6 et PP1 (des tableaux facilitant le calcul pour l'attribution de l'argument)
- Ajout des exemples de seuils de fréquences pour l'attribution des arguments PM2/BS1/BA1, issus de dix groupes de travail gène-spécifiques
- Ajout des recommandations sur l'interprétation des CNV intragéniques
- Ajout des recommandations sur l'interprétation des variants de l'ADN mitochondrial
- Ajout de l'annexe avec un résumé des arguments utilisés pour la classification d'un variant

2 Classification du variant

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

Arguments en faveur de la **pathogénicité** du variant :

Argument Très fort	Exemple : **PVS1**
Fort	Exemples : **PS4, PM3_fort**
Moyen	Exemples : **PM2, PP1_moyen**
Faible	Exemples : **PP3, PM3_faible, PS4_faible**

Arguments en faveur du **caractère bénin** du variant :

Argument suffisant	Exemple : **BA1**
Fort	Exemples : **BS1, BS3**
Faible	Exemples : **BP1, BP2**

☐ PATHOGENE	☐ PROBABLEMENT PATHOGENE	☐ DE SIGNIFICATION INCERTAINE	☐ PROBABLEMENT BENIN	☐ BENIN
☐ PVS1 + PS	☐ PVS1 + PM	☐ Critères insuffisants	☐ BS + BP	☐ BA1
☐ PVS1 + 2PM	☐ PS + PM	☐ Critères contradictoires	☐ 2BP	☐ 2BS
☐ PVS1 + PM + PP	☐ PS + 2PP			
☐ PVS1 + 2PP	☐ 3PM			
☐ 2PS	☐ 2PM + 2PP			
☐ PS + 3PM	☐ 1PM + 4PP			
☐ PS + 2PM + 2PP				
☐ PS + PM + 4PP				

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

- Varsome *https://varsome.com*
- ClinGen Pathogenicity Calculator *http://calculator.clinicalgenome.org*
- InterVar Classify System *http://wintervar.wglab.org/evds.php*
- Genetic Variant Interpretation Tool

http://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/

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

Les recommandations inclues dans la version 1 de ce document sont remplacées par le document NGS-DIAG_006 relatif aux comptes-rendus de NGS.

4 Bibliographie et Annexe

L'annexe 1 reprend le poids de chaque argument.

1. Amendola, L. M. *et al.* Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am. J. Hum. Genet.* **98**, 1067–1076 (2016).
2. Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet. EJHG* **24**, 1515 (2016).
3. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
4. https://www.acgs.uk.com/media/11285/uk-practice-guidelines-for-variant-classification-2019-v1-0-3.pdf.
5. Abou Tayoun, A. N. *et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* **39**, 1517–1524 (2018).
6. Biesecker, L. G., Harrison, S. M. & ClinGen Sequence Variant Interpretation Working Group. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**, 1687–1688 (2018).
7. Brnich, S. E. *et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **12**, 3 (2019).
8. Ghosh, R. *et al.* Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum. Mutat.* **39**, 1525–1530 (2018).
9. SVI_Recommendation_for_PM3_Criterion. *SVI Recommendation for in trans Criterion (PM3) - Version 1.0*. (2019).
10. SVI_Recommendation_for_PS2_and_PM6_criteria. *PS2/PM6: Recommendation for de novo PS2 and PM6 ACMG/AMP criteria (Version 1.0)*. (2018).
11. Gelb, B. D. *et al.* ClinGen's RASopathy Expert Panel consensus methods for variant interpretation. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**, 1334–1345 (2018).
12. Kelly, M. A. *et al.* 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 (2018).
13. Lee, K. *et al.* Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. *Hum. Mutat.* **39**, 1553–1568 (2018).
14. Luo, X. *et al.* ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv.* **3**, 2962–2979 (2019).
15. Mester, J. L. *et al.* Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum. Mutat.* **39**, 1581–1592 (2018).
16. Oza, A. M. *et al.* Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum. Mutat.* **39**, 1593–1613 (2018).
17. Romanet, P. *et al.* Proposition of adjustments to the ACMG-AMP framework for the interpretation of MEN1 missense variants. *Hum. Mutat.* **40**, 661–674 (2019).
18. Zastrow, D. B. *et al.* 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 (2018).
19. Fortuno, C. *et al.* Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. *Hum. Mutat.* **42**, 223–236 (2021).
20. ClinGen Lysosomal Storage Disorders Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1; https://clinicalgenome.org/site/assets/files/3969/clingen_lsd_acmg_specifications_v1.pdf.
21. https://clinicalgenome.org/curation-activities/variant-pathogenicity/documents/.
22. Strande, N. T. *et al.* 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 (2017).

23. Bean, L. J. H. *et al.* 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.* (2019) doi:10.1038/s41436-019-0666-z.
24. https://search.clinicalgenome.org/kb/curations/.
25. https://panelapp.genomicsengland.co.uk/panels/entities/.
26. https://decipher.sanger.ac.uk/ddd/ddgenes.
27. https://anpgm.fr/media/documents/BP-GT4_Maitrise_Pipeline_Bioinfo_NGS.pdf.
28. https://www.has-sante.fr/upload/docs/application/pdf/2013-02/c_2012_0198_abm_regles_bonnes_pratiques_genetique_constitutionnelle.pdf.
29. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
30. Acuna-Hidalgo, R. *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* **101**, 50–64 (2017).
31. NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
32. Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **19**, 1151–1158 (2017).
33. https://www.clinicalgenome.org/site/assets/files/5182/pm2_-_svi_recommendation_-_approved_sept2020.pdf.
34. ClinGen_Lysosomal_Storage_Disorders_Expert_Panel. *ClinGen Lysosomal Storage Disorders Expert Panel Specifications to the ACMG/AMP Variant Interpretation Guidelines Version 1*. (2019).
35. Luo, X. *et al.* ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood Adv.* **3**, 2962–2979 (2019).
36. Mester, J. L. *et al.* Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum. Mutat.* **39**, 1581–1592 (2018).
37. BA1_Exception_List. BA1 Exception List, July 2018. (2018).
38. https://tinyurl.com/ClinGen-BA1.
39. Allot, A. *et al.* LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. *Nucleic Acids Res.* **46**, W530–W536 (2018).
40. Birgmeier, J. *et al.* AVADA: toward automated pathogenic variant evidence retrieval directly from the full-text literature. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **22**, 362–370 (2020).
41. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
42. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
43. https://www.has-sante.fr/upload/docs/application/pdf/2017-11/reunion_de_concertation_pluridisciplinaire.pdf.
44. Estrada-Cuzcano, A. *et al.* Novel IQCE variations confirm its role in postaxial polydactyly and cause ciliary defect phenotype in zebrafish. *Hum. Mutat.* **41**, 240–254 (2020).
45. Karaca, E. *et al.* Phenotypic expansion illuminates multilocus pathogenic variation. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**, 1528–1537 (2018).
46. Posey, J. E. *et al.* Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *N. Engl. J. Med.* **376**, 21–31 (2017).
47. Smith, E. D. *et al.* A retrospective review of multiple findings in diagnostic exome sequencing: half are distinct and half are overlapping diagnoses. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **21**, 2199–2207 (2019).
48. Jarvik, G. P. & Browning, B. L. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am. J. Hum. Genet.* **98**, 1077–1081 (2016).
49. https://clinicalgenome.org/docs/summary-of-updates-to-the-clingen-gene-clinical-validity-curation-sop-version-7/.
50. https://dosage.clinicalgenome.org/.
51. Riggs, E. R. *et al.* Towards an evidence-based process for the clinical interpretation of copy number variation. *Clin. Genet.* **81**, 403–412 (2012).
52. https://www.ebi.ac.uk/gene2phenotype.
53. Thormann, A. *et al.* Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat. Commun.* **10**, 2373 (2019).
54. Ziegler, A., Colin, E., Goudenège, D. & Bonneau, D. A snapshot of some pLI score pitfalls. *Hum. Mutat.* (2019) doi:10.1002/humu.23763.

55. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet.* **6**, e1001154 (2010).
56. Xiang, J., Peng, J., Baxter, S. & Peng, Z. AutoPVS1: An automatic classification tool for PVS1 interpretation of null variants. *Hum. Mutat.* **41**, 1488–1498 (2020).
57. http://autopvs1.genetics.bgi.com/.
58. Cummings, B. B. *et al.* Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020).
59. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
60. https://www.gtexportal.org/home/.
61. https://www.ncbi.nlm.nih.gov/orffinder.
62. Whiffin, N. *et al.* Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nat. Commun.* **11**, 2523 (2020).
63. Klauer, A. A. & van Hoof, A. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdiscip. Rev. RNA* **3**, 649–660 (2012).
64. Hu, Z. *et al.* VIPdb, a genetic Variant Impact Predictor Database. *Hum. Mutat.* **40**, 1202–1214 (2019).
65. https://genomeinterpretation.org/vipdb.
66. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
67. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
68. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
69. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
70. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
71. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
72. Salgado, D. *et al.* UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. *Hum. Mutat.* **37**, 439–446 (2016).
73. Tavtigian, S. V. *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **43**, 295–305 (2006).
74. González-Pérez, A. & López-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440–449 (2011).
75. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
76. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
77. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
78. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
79. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).
80. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* **18**, 225 (2017).
81. Gunning, A. C. *et al.* Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J. Med. Genet.* (2020) doi:10.1136/jmedgenet-2020-107003.
82. https://varsome.com/.
83. http://vsranker.broadinstitute.org/.
84. https://calculator.clinicalgenome.org/.
85. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **11**, 377–394 (2004).
86. Senapathy, P., Shapiro, M. B. & Harris, N. L. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252–278 (1990).
87. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **4**, 311–323 (1997).

88. Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).
89. Desmet, F.-O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
90. Houdayer, C. *et al.* 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 (2012).
91. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535–548.e24 (2019).
92. Jagadeesh, K. A. *et al.* S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. *Nat. Genet.* **51**, 755–763 (2019).
93. Leman, R. *et al.* Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res.* **46**, 7913–7923 (2018).
94. Cheng, J. *et al.* MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* **20**, 48 (2019).
95. https://anpgm.fr/media/documents/BP-ANPGM_004_Evaluation_impact_variant_sur_pissage.pdf.
96. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet. Test. Mol. Biomark.* **14**, 533–537 (2010).
97. Gnad, F., Baucom, A., Mukhyala, K., Manning, G. & Zhang, Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* **14 Suppl 3**, S7 (2013).
98. Moyer, D. C., Larue, G. E., Hershberger, C. E., Roy, S. W. & Padgett, R. A. Comprehensive database and evolutionary dynamics of U12-type introns. *Nucleic Acids Res.* **48**, 7066–7078 (2020).
99. Sibley, C. R., Blazquez, L. & Ule, J. Lessons from non-canonical splicing. *Nat. Rev. Genet.* **17**, 407–421 (2016).
100. Li, M. M. *et al.* Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn. JMD* **19**, 4–23 (2017).
101. https://www.clinicalgenome.org/affiliation/50027.
102. McCormick, E. M. *et al.* Specifications of the ACMG/AMP standards and guidelines for mitochondrial DNA variant interpretation. *Hum. Mutat.* (2020) doi:10.1002/humu.24107.
103. Wong, L.-J. C. *et al.* Interpretation of mitochondrial tRNA variants. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **22**, 917–926 (2020).
104. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
105. Bris, C. *et al.* Bioinformatics Tools and Databases to Assess the Pathogenicity of Mitochondrial DNA Variants in the Field of Next Generation Sequencing. *Front. Genet.* **9**, 632 (2018).
106. Cappa, R., de Campos, C., Maxwell, A. P. & McKnight, A. J. 'Mitochondrial Toolbox' - A Review of Online Resources to Explore Mitochondrial Genomics. *Front. Genet.* **11**, 439 (2020).
107. https://acpa-achropuce.com/wp-content/uploads/2021/01/Guide-interpretation-des-CNV-2020.pdf.