| Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS | | |
|---|---|---|
| Date de création : 20/12/2017 | Date de révision : | Version : 1 |
| Date de 1ère Application : 01/06/2018 | N° document : NGSDIAG_001 | |
| Approbation par le board du réseau le 08/02/2018 | | |

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


- 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),

# Sommaire

1 Liste des arguments à prendre en compte pour l'interprétation des variants ..... 5
1-1- Données épidémiologiques ..... 6
1-2- Données structurales ..... 11
1-3- Données bibliographiques ..... 14
1-4- Critères cliniques et généalogiques ..... 14
1-5- Variants associés ..... 15
1-6- Données de ségrégation ..... 15
1-7- Données fonctionnelles ..... 16
1-8- Analyses somatiques ..... 17
1-9- Modifications de l'argumentaire par rapport aux recommandations ..... 17
2 Classification du variant ..... 18
3 Utilisation dans le cadre d'un compte-rendu d'analyses ..... 19
4 Bibliographie et Annexes ..... 19

**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
GGC : Groupe Génétique et Cancer
HI index : Haploinsufficiency Score
MAF : Minor Allele Frequency
NGS : Next Generation Sequencing
PFMG2025 : Plan France Médecine Génomique 2025
pLI: probability of Loss of function intolerance
Réseau AChroPuce: Réseau d'Analyse Chromosomique sur Puce à ADN
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*) (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 de certains items de l'interprétation, ...).

**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.**

# 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, ou autres bases), Z-score de contrainte protéique dans ExAC, sensibilité à l'haplo-insuffisance (par exemple pLI dans ExAC)
- 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 (Richards *et al., Genet Med* 2015 ; Amendola *et al., Am J Hum Genet* 2016). 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 ou PS4.**

## 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.

Il est important de prendre en compte l'origine ethnique du patient pour interpréter les données des bases de données.

### 1-1-1 Etude des bases de données de 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 :**

- **1000 Genomes :**

  Les données sont issues de 3511 génomes de sujets non apparentés (2016). Cette base regroupe 26 populations différentes. Elle est principalement adaptée à l'interprétation de variants non exoniques.

  http://www.internationalgenome.org/

- **ESP (*NHLBI Exome Sequencing Project*) :**

  Les données sont issues de 6503 exomes de sujets non apparentés (2016). Les individus séquencés sont stratifiés en deux populations : EA (european-americans) et AA (african-americans.

  http://evs.gs.washington.edu/EVS/

  Attention :

  Cette base de données comprenant à la fois des individus sains et des patients atteints de diverses pathologies (notamment cardio-vasculaires et pulmonaires), une grande prudence doit être accordée à son utilisation.

- **ExAC (*Exome Aggregation Consortium*) :**

  Les données sont issues de 60706 exomes de sujets non apparentés (Lek *et al., Nature* 2016), et incluent les données de 1000 génomes et partiellement d'ESP. La population générale est génétiquement stratifiée en 6 populations (AFR : population africaine et afro-américaine, AMR : population d'origine mixte américaine hispanique, EAS : population d'Asie de l'est, FIN : population finlandaise, NFE : population européenne non finlandaise, SAS : population d'Asie du sud, OTH : autres). La fréquence allélique du variant à interpréter est détaillée pour chacune de ces populations. Il est opportun de retenir la fréquence allélique correspondante à la « pop max », c'est-à-dire la fréquence allélique la plus élevée décrite dans ces 6 populations. Cette base est principalement adaptée à l'interprétation de variants exoniques.

  http://exac.broadinstitute.org/

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

  Les données sont issues de 123 136 exomes de sujets non apparentés et de 15 496 génomes (2017). Ces données incluent les données de ExAC, 1000 génomes et partiellement d'ESP. Il s'agit actuellement de la base de données la plus puissante en termes de nombre d'individus séquencés. En particulier, la présence de 15 496 génomes entiers permet d'appréhender la fréquence des variants non-exoniques de manière plus précise que 1000 Genomes.

  Elle intègre des populations qui jusque-là n'étaient pas représentées dans les bases de données comme la population juive ashkenaze.

  Comme ExAC, gnomAD consiste en une aggrégation de données génomiques d'individus non apparentés issus de diverses études populationnelles ou spécifiques de certaines maladies. 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 n'est pas précisée.

https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/

- **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).

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.

- **Wellderly :**

The Wellderly 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). 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**)

- Fréquence allélique trop importante par rapport à la fréquence de la pathologie (**BS1**) ou présence du variant chez les contrôles incohérente avec la pénétrance de la pathologie (**BS2**)

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.

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

- 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**)

- 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 détectées dans la population générale par séquençage haut débit (ne pas utiliser les données ESP pour cette raison).
- Le recrutement des sujets pour ces bases de données n'inclut pas uniquement des sujets dits contrôles sains. Dans la base de données ExAC (et donc dans GnomAD), des cohortes de patients schizophrènes et atteints de formes héréditaires de cancer sont incluses. Dans ExAC (et donc dans GnomAD), des données du TCGA (séquençage de tumeurs) sont aussi incluses.
- Ces bases de données sont internationales et ne sont donc pas représentatives des variants spécifiques de sous-populations (régionales ou autres).
- Il existe une redondance de données entre certaines de ces bases de données, notamment ESP, ExAC et GnomAD
- Certains variants pathogènes reconnus ont une Fréquence allélique >1%.
- Ne peuvent être considérées que les données de fréquences alléliques issues de cohortes d'au moins 1000 individus.
- 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.

### 1-1-2 Bases de données de patients

Il est nécessaire de consulter les bases de données de patients comme ClinVar, HGMD et LOVD (bases spécifiques de gènes), COSMIC. En fonction des pathologies, il peut être important de consulter les bases nationales concernées (Ex : Base 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 important de veiller à l'origine ethnique du patient.
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 :**

- Source documentée classant ce variant comme pathogène (**PP5**)

- Source documentée classant ce variant comme bénin (**BP6**)

La base de données ClinVar donne d'emblée des éléments pour pondérer le poids de l'argument. Ainsi, ClinVar**** ou ClinVar*** donne un **PS**, ClinVar** un **PM** et ClinVar* un **PP**. Il est cependant important de garder à l'esprit que certains variants sont rapportés dans ClinVar avec des informations fausses ou incomplètes. Le biologiste doit donc vérifier la source des variants rapportés.

- 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 faux-sens pathogène connu (**PM5**)

**Précautions :**

- 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 cas 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- Données structurales

### 1-2-1 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**)

- 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
- Cet argumentaire 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 par exemple le score pLi (*probability of Loss of function intolerance*), fourni par la base de données ExAC : un gène est intolérant à la perte de fonction si sa pLi est supérieure ou égale à 0,9. Cette pLI n'est pas un argument suffisant car elle comporte 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.). Ainsi, une analyse qualitative de la charge en variants tronquants dans les bases de données est utile afin d'aborder la pLI de manière critique.
  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.
- 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 3ème 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*).

## 1-2-2- 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).

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. Les logiciels les plus fréquemment utilisés sont les logiciels MaxEntScan, SpliceSiteFinder, NNSPLICE, GeneSplicer et Human Splicing Finder (Houdayer *et al., Hum Mutat* 2012). 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 entrainés 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.

Les informations utilisées par certains des logiciels les plus fréquemment utilisés de prédiction des variants faux-sens sont les suivantes :

| **Logiciel** | **Informations utilisées** |
|---|---|
| SIFT | Conservation au cours de l'évolution |
| FATHMM | Conservation au cours de l'évolution |
| Mutation Assessor | Conservation au cours de l'évolution |
| Align GVGD | Structure et fonctions protéiques<br>Conservation au cours de l'évolution |
| Polyphen-2 | Structure et fonctions protéiques<br>Conservation au cours de l'évolution |
| Mutation Taster | Structure et fonctions protéiques<br>Conservation au cours de l'évolution<br>Epissage |
| UMD-predictor | Structure et fonctions protéiques<br>Conservation au cours de l'évolution<br>Epissage |

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.

## 1-3- Données bibliographiques

Il est nécessaire de faire une revue de la littérature concernant le variant à classer. La base de données bibliographique communément utilisée est PubMed (NCBI : *National Center for Biotechnology Information*).

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

## 1-4- 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.

**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.

## 1-5- Variants associés

**Arguments :**

- Pour une pathologie à transmission autosomique récessive, variant observé en *trans* (détermination de la phase par analyse familiale) avec un variant pathogène (**PM3**)

- Pour une pathologie à pénétrance complète de transmission autosomique dominante ou liée à l'X, co-occurence 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-occurence 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-occurence 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 :**

- Le critère PM3 peut être modifié en PP si la phase n'est pas déterminée.
- 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*

## 1-6- 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**)

- Variant co-ségrégant avec la pathologie chez plusieurs membres atteints dans la même famille et/ou dans plusieurs familles : 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 ≤ 1/16 : argument fort (PS) ; N ≤ 1/8 : argument moyen (PM) ; N ≤ 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).

- Variant ne ségrégant 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 asymptomatique, 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,…

## 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**)

**Chaque FNMR/réseau d'oncogénétique est amené à déterminer le poids de ces arguments dans l'interprétation.**

## 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


Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, Cooper GM, Dorschner MO, Dulik MC, Ghazani AA, Ghosh R, Green RC, Hart R, Horton C, Johnston JJ, Lebo MS, Milosavljevic A, Ou J, Pak CM, Patel RY, Punj S, Richards CS, Salama J, Strande NT, Yang Y, Plon SE, Biesecker LG, Rehm HL. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. Am J Hum Genet. 2016 Jul 7;99(1):247.

Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. Am J Hum Genet. 2016 Jun 2;98(6):1077-1081.

Ellard S, Baple EL, Owens M, Cannon S, Eccles DM, Abbs S, Scott R, Deans ZC, Lester T, Campbell J, Newman WG, McMullan DJ; ACGS Best Practice Guidelines for Variant Classification 2017.
http://www.acgs.uk.com/media/1092626/uk_practice_guidelines_for_variant_classification_2017.pdf

Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, Topol SE, Wineinger NE, Niederhuber JE, Topol EJ, Torkamani A. Whole-Genome Sequencing of a Healthy Aging Cohort. Cell. 2016 May 5;165(4):1002-11.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285-91.

Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Léone M, Mazoyer S, Muller D, Remenieras A, Révillion F, Rouleau E, Sokolowska J, Vert JP, Lidereau R, Soubrier F, Sobol H, Sevenet N, Bressac-de Paillerets B, Hardouin A, Tosi M, Sinilnikova OM, Stoppa-Lyonnet D. 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. 2012 Aug;33(8):1228-38.

Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, Walkiewicz M, Bi W, Xiao R, Ding Y, Xia F, Beaudet AL, Muzny DM, Gibbs RA, Boerwinkle E, Eng CM, Sutton VR, Shaw CA, Plon SE, Yang Y, Lupski JR. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N Engl J Med. 2017 Jan 5;376(1):21-31.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. 2015 May;17(5):405-24.

**Annexe 1 : Poids de chaque argument et les précautions d'interprétation (Richards *et al., Genet Med* 2015)**

**Annexe 2 : Tableau récapitulant les différents paramètres d'interprétation et leur poids (Richards *et al., Genet Med* 2015, modifié d'après Jarvik and Browning, *Am J Hum Genet* 2016)**

## Annexe 1 : Poids de chaque argument et les précautions d'interprétation (Richards *et al., Genet Med* 2015)

| Argument de pathogénicité | Catégorie |
|---|---|
| Très fort | PVS1 Variant ayant un effet nul prédit (non-sens, décalage du cadre de lecture, 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<br>Attention:<br>- Cet argumentaire n'est valable que pour les gènes où la perte de fonction est le mécanisme pathogène connu<br>- Attention à l'attribution d'un effet perte de fonction aux variants situés en partie 3' extrême d'un gène<br>- Attention à l'attribution d'un effet perte de fonction pour les variants situés au niveau de sites d'épissage entrainant un saut d'exon(s) en phase, laissant intact le reste de la protéine<br>- Un variant nul prédit peut avoir un effet délétère sur un transcrit, mais laisser intact les autres transcrits du même locus, produits par l'épissage alternatif |
| Fort | PS1 Variant à l'origine du même changement d'acide aminé qu'un variant pathogène connu<br>Exemple: Val>Leu est produit soit par G>C ou par G>T au sein du même codon<br>Attention: un variant peut avoir un effet sur l'épissage, indépendamment de son effet sur la protéine<br><br>PS2 Variant de novo chez un patient sans antécédent familial, avec confirmation de la paternité ET de la maternité<br>Remarque : La confirmation seulement de la paternité n'est pas suffisante: le don d'ovocytes, les mères porteuses, les erreurs pendant le transfert d'embryon peuvent être à l'origine de non-maternité<br><br>PS3 Etudes fonctionnelles in vivo ou in vitro bien établies montrant un impact délétère du variant sur le gène ou son produit<br>Remarque : Les études fonctionnelles validées, robustes et reproductibles dans les conditions d'un laboratoire diagnostique sont considérées comme les plus établies.<br><br>PS4 Prévalence du variant chez les individus atteints significativement supérieure à celle des contrôles de la même origine ethnique que les patients<br>Remarque 1 : Risque relatif ou Odd Ratio (pour les études cas-témoins) > 5 ET l'intervalle de confiance autour de RR ou OR estimé n'inclut pas 1<br>Remarque 2 : Pour des variants très rares, les études types cas-témoins peuvent ne pas atteindre le seuil de significativité. Dans ce cas, la présence d'un variant chez plusieurs patients non-apparentés, ainsi que l'absence de ce variant chez les contrôles, peut être utilisé comme un argument de poids moyen |
| Moyen | PM1 Variant non-tronquant situé sur un hot spot mutationnel et/ou un domaine fonctionnel critique bien établi (e.g. un site actif d'une enzyme) exempt de variants bénins<br><br>PM2 Variant absent des bases de données de populations contrôles (ou très faible fréquence si récessif) GnomAD, ExAC, 1000 génomes<br>Attention : Les insertions/délétions (en particulier de grande taille) peuvent être mal détectées dans la population générale par séquençage haut débit<br><br>PM3 Variant observé en trans avec un variant pathogène distinct, si la pathologie a une transmission récessive<br>Remarque : nécessité de tester les parents pour déterminer la phase<br><br>PM4 Variant affectant la longueur de la protéine : insertions/délétions en phase dans une région non-répétée ou perte du codon stop<br><br>PM5 Variant à l'origine d'un changement d'acide aminé différent à la même position qu'un variant faux-sens pathogène connu<br>Exemple: Arg156His est connu pathogène, le variant identifié est Arg156Cys<br>Attention: un variant peut avoir un effet sur l'épissage, indépendamment de son effet sur la protéine<br><br>PM6 Variant de novo sans confirmation de la paternité et de la maternité |
| Faible | PP1 Variant qui ségrège avec la pathologie chez de multiples apparentés atteints<br>Remarque : cet item a été adapté dans les recommandations françaises selon Jarvik et Browning (Am J Hum Genet 2016)<br><br>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<br><br>PP3 Effet d'un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés (conservation, effet sur épissage, etc.)<br>Attention : PP3 ne peut être attribué qu'une seule fois pendant l'évaluation d'un variant, car des nombreux logiciels de prédiction utilisent des données similaires<br><br>PP4 Phénotype ou histoire familiale spécifique en faveur de la pathologie associée au gène comprenant le variant<br><br>PP5 Source documentée classant ce variant comme pathogène, dans la situation où les arguments permettant de définir la pathogénicité ne sont pas disponibles pour une ré-évaluation indépendante |

| Argument de caractère bénin | Catégorie |
|---|---|
| Suffisant | BA1 Fréquence allélique > 5% dans les bases de données de populations contrôles |
| Fort | BS1 Fréquence allélique trop importante par rapport à la fréquence de la pathologie<br><br>BS2 Présence chez un individu sain d'un variant homozygote pour une maladie récessive, hétérozygote pour une maladie dominante ou hemizygote pour une maladie liée à X, si une pénétrance complète de cette maladie est attendue en bas âge<br><br>BS3 Etudes fonctionnelles bien établies montrent un impact non délétère du variant<br><br>BS4 Variant ne ségrégant pas avec la pathologie chez des apparentés atteints<br>Attention : L'existence de phénocopies peut mimer un défaut de ségrégation pour les maladies avec un phénotype commun (e.g. cancer, épilepsie). Pour certaines familles, plus d'un variant peut contribuer à une maladie autosomique dominante, produisant un effet de non-ségrégation |
| Faible | BP1 Variant faux-sens dans un gène où seuls les variants tronquants sont décrits comme associées à la pathologie<br><br>BP2 Variant observé en trans avec un variant pathogène pour une maladie dominante à pénétrance complète; variant observé en cis avec un variant pathogène<br>Remarque : cet item a été adapté dans les recommandations françaises, avec un poids de l'argument laissé à l'appréciation du laboratoire |

## Annexe 2 : Tableau récapitulant les différents paramètres d'interprétation et leur poids (Richards *et al., Genet Med* 2015, modifié d'après Jarvik and Browning, *Am J Hum Genet* 2016)

| | Bénin | | Pathogène | | | |
|---|---|---|---|---|---|---|
| | BA/BS | BP | PP | PM | PS | PVS |
| | Suffisant/Fort | Faible | Faible | Moyen | Fort | Très fort |
| **Données épidémiologiques** | Fréquence allélique trop importante par rapport à la fréquence de la pathologie (**BA1/BS1**) OU présence du variant chez les contrôles incohérente avec la pénétrance de la pathologie (**BS2**) | | | Variant absent des bases de données de populations contrôles (**PM2**) | Prévalence du variant chez les individus atteints significativement supérieure à celle des contrôles (**PS4**) | |
| **Données structurales** | | Effet d'un variant prédit bénin par l'ensemble des logiciels de prédiction de pathogénicité interrogés (**BP4**)<br><br>Variant faux-sens dans un gène où seules les variants tronquants sont décrits comme associées à la pathologie (**BP1**)<br><br>Variant synonyme sans impact prédit sur l'épissage, et pour lequel la séquence nucléotidique n'est pas très conservée chez les vertébrés (**BP7**)<br><br>Indel en phase dans une région répétée sans fonction connue (**BP3**) | Effet d'un variant prédit délétère par l'ensemble des logiciels de prédiction de pathogénicité interrogés (**PP3**) | Variant à l'origine d'un changement d'acide aminé différent à la même position qu'un variant faux-sens pathogène connu (**PM5**)<br><br>Variant affectant la longueur de la protéine (concerne les indel en phase et pertes de codons stop) (**PM4**) | 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 ayant un effet nul prédit dans un gène où la perte de fonction est un mécanisme pathogène connu (**PVS1**) |
| **Données fonctionnelles** | Etudes fonctionnelles *in vivo* ou *in vitro* bien établies montrent un impact non délétère du variant sur le gène ou son produit (cet argument peut inclure les résultats d'analyses fonctionnelles spécifiques en fonction de la pathologie concernée) (**BS3**) | | Variant faux-sens dans un gène avec un faible taux de variants faux-sens bénins et dans lequel les variats faux-sens sont un mécanisme fréquemment responsable de la pathologie (**PP2**) | Variant non tronquant situé sur un hot spot mutationnel et/ou un domaine fonctionnel critique bien établi exempt de variants bénins (**PM1**) | Etudes fonctionnelles *in vivo* ou *in vitro* bien établies montrant un impact délétère du variant sur le gène ou son produit (cet argument peut inclure les résultats d'analyses fonctionnelles spécifiques en fonction de la pathologie concernée)(**PS3**) | |
| **Données de ségrégation** | Variant ne ségrégant pas avec la pathologie chez des apparentés atteints (**BS4**) | | Données issues d'une seule famille : N ≤ 1/8 : argument faible (**PP**)*<br><br>Données issues de plusieurs familles : N ≤ 1/4 : argument faible (**PP**)* | Données issues d'une seule famille : N ≤ 1/16 : argument moyen (**PM**)*<br><br>Données issues de plusieurs familles N ≤ 1/8 : argument moyen (**PM**)* | Données issues d'une seule famille : N ≤ 1/32 : argument fort (**PS**)*<br><br>Données issues de plusieurs familles : N ≤ 1/16 : argument fort (**PS**)* | |
| **Données de novo** | | | | Variant de novo sans confirmation de la paternité et de la maternité , si la pénétrance est considérée complète (**PM6**) | Variant de novo avec confirmation de la paternité et de la maternité (**PS2**) | |
| **Données alléliques** | | Pour une pathologie à pénétrance complète de transmission autosomique dominante ou liée à l'X, co-occurence 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)<br><br>Co-occurence 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) | | Variant observé en *trans* avec un variant pathogène distinct, si la pathologie a une transmission récessive (poids de l'argument laissé à l'appréciation du laboratoire, pondération **PM3** dans les recommandations ACMG-AMP initiales) | | |
| **Autres bases de données** | | Source documentée classant ce variant comme bénin (**BP6**) | Source documentée classant ce variant comme pathogène (**PP5**) | | | |
| **Données additionnelles** | | Co-occurence 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) | Le phénotype du patient ou les antécédents familiaux sont très spécifiques pour le gène (**PP4**) | | | |

- **Le tableau 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 (référencées 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).**
- **Les items marqués d'un astérisque sont issus d'ajouts selon Jarvik and Browning, *Am J Hum Genet 2016* et ne comportent pas de numéro spécifique. 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.**