



# Interprétation des variations génétiques : les besoins du généticien moléculaire à l'ère de la génomique

*Gaël Nicolas, MD, PhD*

*MCU-PH Génétique*

**Service de Génétique – CHU de Rouen**

**Inserm U 1245 – équipe neurogénétique**

**FHU Centre Normand de Génomique et de Médecine Personnalisée**

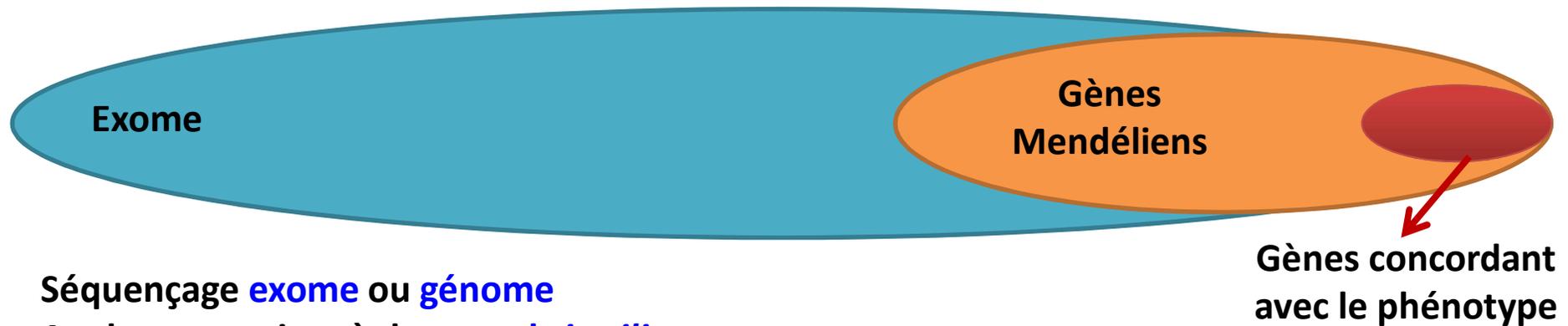
# Utilisation du NGS pour le diagnostic moléculaire : besoins

- 1. Déterminer une stratégie et une cible adaptée à la question clinique**
- 2. Déterminer un pipeline bioinformatique adapté aux besoins**
- 3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires**
- 4. Déterminer les filtres adaptés**
- 5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif**

# Utilisation du NGS pour le diagnostic moléculaire : besoins

- 1. Déterminer une stratégie et une cible adaptée à la question clinique**
2. Déterminer un pipeline bioinformatique adapté aux besoins
3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires
4. Déterminer les filtres adaptés
5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif

# Approches pangénomiques pour le diagnostic des maladies rares



Séquençage **exome** ou **génom**e

Analyse restreinte à des **panels *in silico***

- ✓ **maladies hétérogènes** (nombreux gènes potentiellement impliqués)
- ✓ **syndromes difficilement reconnaissables**
- ✓ **nombreux diagnostics différentiels**

→ Pas de reséquençage si panel n°1 négatif

Risque : **incidental findings / données secondaires**

→ Identification **incidentale** ou **active** d'un variant pathogène dans un gène de susceptibilité héréditaire au cancer, cardiopathie...

→ Actionable genes : **recommandations américaines**

→ Nécessité de préciser :

**Pour qui ?**

Comment ? **prescription / consentement / suivi psychologique - impact**

Comment ? **expertise moléculaire / RCP**

# Applications au diagnostic des maladies rares

## Stratégies diagnostiques

Séquençage des gènes du panel uniquement

Exome entier

Génome entier

Panels de gène *in silico*

Coût séquençage

Coût bioinformatique

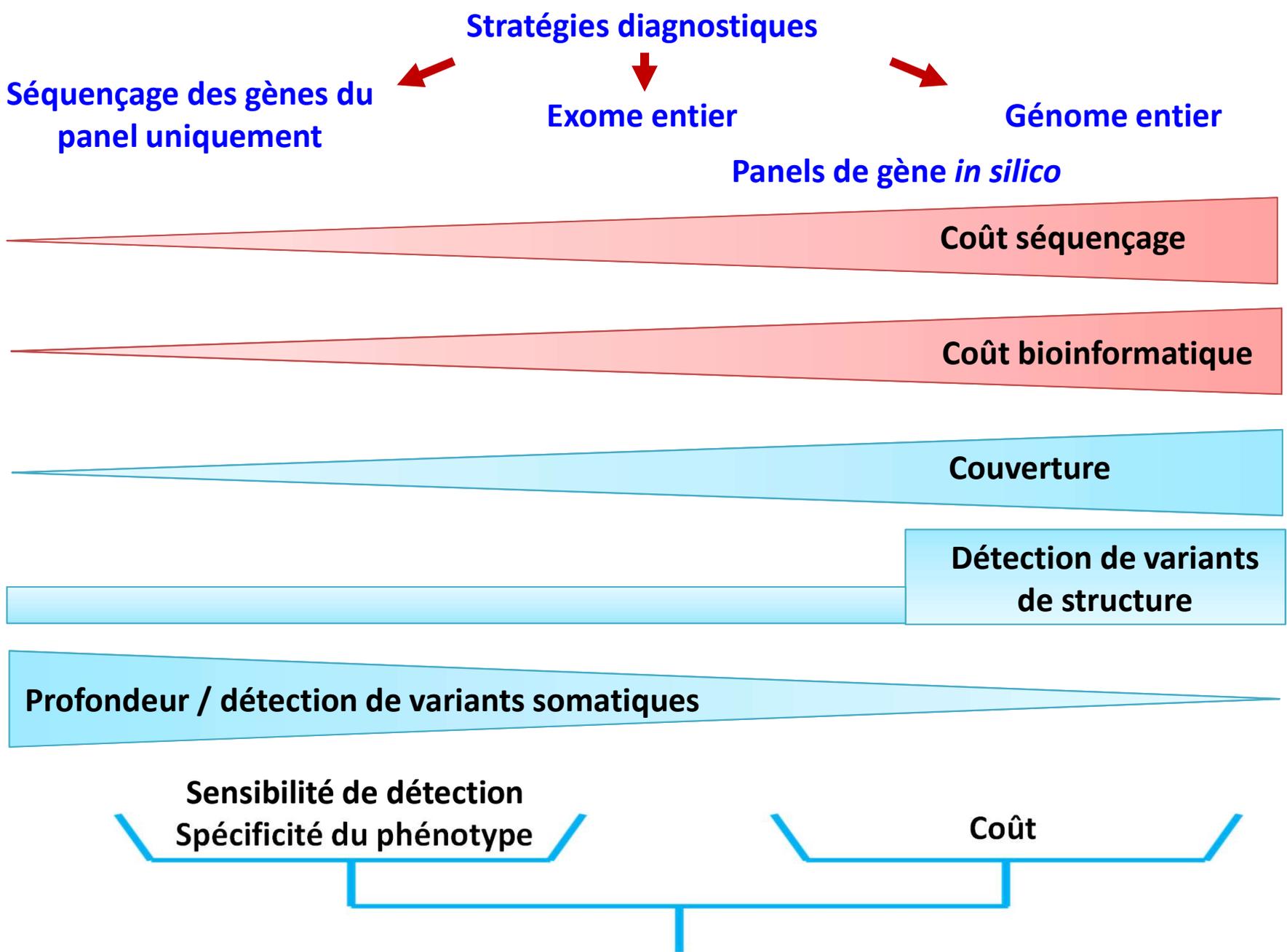
Couverture

Détection de variants de structure

Profondeur / détection de variants somatiques

Sensibilité de détection  
Spécificité du phénotype

Coût

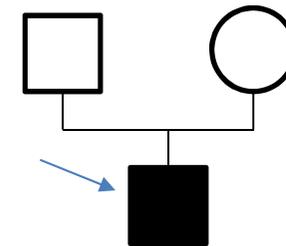


# S'adapter à la question clinique pour choisir la stratégie

Cible choisie → Quel tissu ? Solo ? Trio ?

Éléments de réflexion :

- Probabilité de mosaïque post-zygotique
- Disponibilité du tissu/prélèvement
- Urgence ?

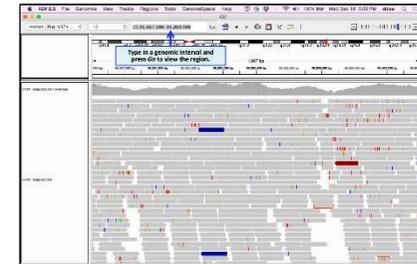


# Utilisation du NGS pour le diagnostic moléculaire : besoins

1. Déterminer une stratégie et une cible adaptée à la question clinique
- 2. Déterminer un pipeline bioinformatique adapté aux besoins**
3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires
4. Déterminer les filtres adaptés
5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif

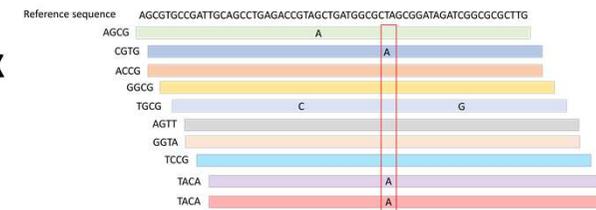
# Alignement, détection et CNV

**Alignement et détection des SNV/indels :**  
Standards maintenant relativement stabilisés :  
BWA / GATK best practices



## Utilisation des UMI ?

Intérêt majeur pour la recherche de mosaïque à faible taux  
Peu/pas d'intérêt pour les applications de recherche de mutation germinales



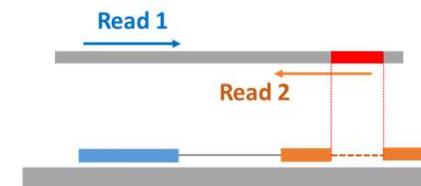
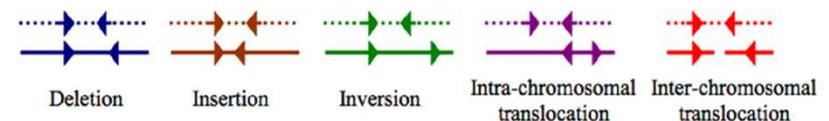
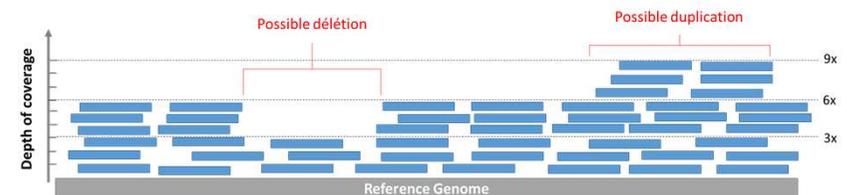
## Détection de variations structurales

Capture exonique :

- DOC comparison +++
- Intérêt des autres outils discutables

WGS / capture des introns/UTR :

- Multiplicité des méthodes
- DOC + read pairs + split reads
- Risque de nombreux faux positifs



# Utilisation du NGS pour le diagnostic moléculaire : besoins

1. Déterminer une stratégie et une cible adaptée à la question clinique
2. Déterminer un pipeline bioinformatique adapté aux besoins
- 3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires**
4. Déterminer les filtres adaptés
5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif

# Annotation

## Niveau du Gène

Carte d'identité du gène

Pathologie(s) associée(s) au gène

Autres

## Niveau du variant

Coordonnées et nomenclature

Données liées au séquençage

Variations populationnelles sans annotation phénotypique

Bases de données de variants en contexte pathologique

Informations prédictives sur la variation

## Annotations spécifiques des SV

## Données de ségrégation

# Groupe de travail annotation – plan France Médecine Génomique

**Liste non exhaustive (presque ?) des annotations potentiellement utiles pour l'interprétation de WGS en solo ou trio dans la Déficience Intellectuelle (DEFIDIAG)**

## **Niveau du Gène**

### **Carte d'identité du gène**

Les identifiants et la taille du ou des transcrits, de la ou les protéines, de la ou les structures 3D (RefSeq, ensembl, Uniprot <http://www.uniprot.org/>, PDB [http://www.rcsb.org/...](http://www.rcsb.org/))

La liste du ou des domaines protéiques, des éventuels motifs (PFAM <http://pfam.xfam.org>, SMART <http://smart.embl-heidelberg.de>) ([Letunic and Bork 2018](#); [Punta et al. 2012](#))

L'existence et le nombre de pseudogènes (<http://www.gencodegenes.org>)

La localisation cellulaire présumée (ex : <http://compartments.jensenlab.org>)

L'expression dans les tissus comme le "Human protein Atlas" ([www.proteinatlas.org](http://www.proteinatlas.org)), GTex (<https://www.gtexportal.org>) ou TISSUES (<http://tissues.jensenlab.org>)

L'information si le gène est soumis à l'empreinte

L'information si le gène est inactivé sur l'X

La Gene Ontology (GO) (<http://www.geneontology.org/>)

Les voies métaboliques telles que KEGG (<http://www.genome.jp/kegg/pathway.html>) et BioCyc (<https://biocyc.org/>)

Les réseaux d'interactions protéine-protéine tels que STRING ([Szklarczyk et al. 2016](#)) ou GRAVITY (<http://gravity.pasteur.fr>)

Les orthologues et paralogues ([Huerta-Cepas et al. 2016](#); [Altenhoff et al. 2015](#); [Linard et al. 2015](#))

Un alignement multiple de la famille de protéine codée par le gène incluant des mesures de la conservation protéique

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du Gène

### Pathologie(s) associée(s) au gène :

**Le nom de la pathologie (OMIM)**

**Le mode de transmission (OMIM)**

Le nombre et les types de mutations (mécanisme moléculaire et/ou sélection) (via ClinVar/HGMD/DDG2P)

Le nombre de patients rapportés si disponible

Les symptômes clés qui lui sont associés si possible en terme HPO ([Köhler et al. 2017](#))

### Autres informations :

La sensibilité du gène aux variations par type et par localisation

(<https://decipher.sanger.ac.uk/gene/MYH3#overview/protein-info>).

**L'intolérance du gène aux variations pertes de fonction** (Zscore et pLI de ExAC, RVIS, <http://genic-intolerance.org/about.jsp>)

Et Présence dans une **liste de gènes d'intérêt**, ex. Gènes de déficience intellectuelle

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du variant

### Coordonnées et nomenclature

Choix du transcrit : le plus long ou annotation la plus « grave »

Annotation tous transcrits

Nomenclature selon les recommandations HGVS niveau g., c., p.

Nombre de variants dans ce gène après filtration

### Données liées au séquençage

Le nombre de lectures du variant

Le nombre de lectures totales à la position du variant

Le ratio entre les 2 précédentes informations

Le nombre de lectures en sens et en anti-sens

Le génotype et sa qualité

La séquence flanquante de la variation (+/- 10 pb)

Ces données doivent également être disponibles pour les analyses familiales (ex : trio)

La couverture pour le gène (pourcentage du gène/transcrit/CDS à une certaine profondeur de lecture)

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du variant

### Variations populationnelles sans annotation phénotypique

Ces données sont capitales car elles permettront les premiers filtres lors de l'analyse :

Pour les SNV/Indels :

**gnomAD** (<http://gnomad.broadinstitute.org>) ([Lek et al. 2016](#)) (restreint à GRCh37).

1000 Génomes phase 3 (<http://www.internationalgenome.org/data>)

kaviar (<http://db.systemsbiology.net/kaviar/>) lift-over fait, dataset variés de 35 projets

dont ExAC (<http://db.systemsbiology.net/kaviar/cgi-pub/Kaviar.pl?show=sources>)

French Exome Project (FrEx, <http://med-laennec.univ-brest.fr/FrExAC/>) et le projet pilote

« Population Générale » du Plan France Médecine Génomique 2025

RDVD « Rare disease variant database » (<https://rdvd.fr>), la base de données de la Fondation Maladie Rare.

Les données générées au cours de ce projet DEFIDIAG (base interne)

Pour les SV :

DGV (<http://dgv.tcag.ca/dgv/app/home>) ([MacDonald et al. 2014](#))

Cartographie de CNV par CGH ([Coe et al. 2014](#)), par WGS ([Sudmant et al. 2015](#))

DECIPHER (<http://decipher.sanger.ac.uk>)

Base de données française du réseau AChro-Puce (BANCCO, <https://bancco.fr>)

Les données générées au cours de ce projet DEFIDIAG (base interne)

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du variant

### Bases de données de variants en contexte pathologique

ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar>)

HGMD (<http://www.hgmd.org>)

DECIPHER (<http://decipher.sanger.ac.uk>) pour les SNV et CNV.

BANCCO, (<https://bancco.fr>) pour les CNV

Les variations possiblement *de novo* déjà identifiées ([Turner et al. 2017](#)) (<http://denovo-db.gs.washington.edu/denovo-db/>)

Locus spécifique

LSDB si disponible (LOVD <http://www.lovd.nl> et UMD <http://www.umd.be>)

Rettbase (<http://mecp2.chw.edu.au/>)

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du variant

### Informations prédictives sur la variation

#### Scores intégrés

CADD (<http://cadd.gs.washington.edu>)

MutationTaster ([Schwarz et al. 2010](#))

REVEL

etc.

#### Impact sur les faux sens

FATHMM (<http://fathmm.biocompute.org.uk>)

M-CAP (<http://bejerano.stanford.edu/mcap/>) ([Jagadeesh et al. 2016](#))

PolyPhen-2 ([Adzhubei, Jordan, and Sunyaev 2013](#))

Sift ([Sim et al. 2012](#))

#### Impact de faux-sens et synonymes

UMD-Predictor (<http://umd-predictor.eu>)

# Groupe de travail annotation – plan France Médecine Génomique

## Niveau du variant

### Informations prédictives sur la variation

Impact d'une indel in-frame :

Sift\_indel2 ([Hu and Ng 2013, 2012](#)),

DDIG-in ([Zhao et al. 2013](#))

Provean ([Choi et al. 2012](#))

Impact sur l'épissage

HSF (<http://www.umd.be/HSF3/>)

MaxEntSplice ([Yeo and Burge 2004](#))

NNSplice ([Reese et al. 1997](#))

PWM ([Shapiro and Senapathy 1987](#))

Impact sur les ESE/ISE (*Exonic/Intronic Splicing Enhancer*)

Conservation nucléotidique

GERP (<http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html>)

PhyloP (<http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt>)

PhastCons ([Siepel et al. 2005](#))

Conservation des acides aminés

Variation tronquante

% de protéine restante après une mutation tronquante, Données de RNAseq disponibles

# Groupe de travail annotation – plan France Médecine Génomique

## Annotations spécifiques des SV

La taille ou le pourcentage du chevauchement entre les 2 évènements.

Un seuil de 70% est classiquement utilisé

Le génotype ou nombre de copies

La liste des gènes concernés (et les annotations décrites pour les gènes)

Le nombre de variants (SNV/indel) du patient dans la zone du CNV et leur statut (#htz, #hmz).

Ceci afin de par exemple écarter des faux positifs.

La caractérisation des **points de cassure** est encore difficile à déterminer de manière automatique et avec précision. Néanmoins sur la base d'un intervalle ou si le point de cassure précis est déterminé (+/- 100 bases) certaines annotations sont utiles :

Les éléments répétés (ex : track du *genome browser* UCSC “repeatmasker”)

Les LCR (*Low Copy Repeat*) (ex : track du *genome browser* UCSC “Segmental Dups” et “Self Chain”)

La présence de motif PRDM9 (homologous recombination hotspot motif 5'-CCNCCNTNNCCNC-3') ([Myers et al. 2008](#))

Les possibles gènes/transcrits de fusion connus (sens, en phase)

# Groupe de travail annotation – plan France Médecine Génomique

## Autres

Eléments mobiles

Parmi les annotations possibles, une base de données de MEI polymorphiques est disponible (<http://dbrip.brocku.ca/>).

Les TAD (*Topology Associated Domains*)

## Données de ségrégation

*de novo* certain : hétérozygote chez le cas index, absent chez ses parents avec une couverture suffisante (à définir)

*de novo* candidat : hétérozygote chez le cas index, faible profondeur chez les parents (<6x)

Variant hérité du parent 1

Variant hérité du parent 2

*de novo* post-zygotique : faible AB chez l'enfant, absent chez les parents

# Annotation et visualisation : approche pragmatique

1. Déterminer une **liste restreinte d'annotations nécessaires** pour
  - Permettre une filtration facile
  - classer rapidement sans dérouler un fichier trop large

## A. Métadonnées patient

- Code échantillon
- Sexe
- +/- âge ou autres métadonnées spécifiques de la maladie

## B. Information de base sur le variant

- Type de variant (faux sens, perte de fonction etc.)
- Nomenclature coordonnées génomiques, cDNA, protéine
- Fréquences dans gnomAD (liste restreinte choisie ex. popmax WES/WGS/globale)
  - AC, AN, AC hémi/homoz
- Fréquence locale si possible
- Prédications pathogénicité : un minimum de données pour une vision globale rapide
- Qualité du variant : DP, Ref/alt/AB, GQ
- Bases de données de variants : clinvar
- Autres bases de données utiles si possible Ex. DDD, HGMD, denovodb, autres bases de données en pathologie
- Ségrégation / Génotype parents si disponible et données qualité chez les parents

# Annotation et visualisation : approche pragmatique

## 3. Information Gène

- Nom de la pathologie OMIM
- Mode de transmission OMIM
- pLI, DD2GP / mécanisme connu haploinsufficance ou non

## 2. Exploration complémentaire de variants d'intérêt

Par l'accès online direct (liens) ou informations d'annotation supplémentaires  
(affichage à la demande)

Ex.

- OMIM clinical synopsis
- HGMD liste des publications
- Clinvar ensemble des variants pathogènes pour évaluer le spectre / mécanismes / position
- UCSC browser
- Varsome
- Decipher
- GnomAD informations complémentaires
- Logiciels spécifiques ex. Alamut Visual (splice ++)

# Utilisation du NGS pour le diagnostic moléculaire : besoins

1. Déterminer une stratégie et une cible adaptée à la question clinique
2. Déterminer un pipeline bioinformatique adapté aux besoins
3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires
- 4. Déterminer les filtres adaptés**
5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif

# Filtration

## Filtre de qualité

AB, AD, DP, Filter et batch count

Dépend de l'objectif : allelic ratio ?

## Filtre de fréquence

Variations rares, fréquence allélique dans toutes les sous-populations inférieure à 0.01

Autres seuils possibles : définition maladie rare 1/10,000 → 0.005

Voire possiblement moins  $<0,001$  comme définit à 0,0045 par [\(Piton, Redin, and Mandel 2013\)](#) pour le mode de transmission autosomique dominant) [\(Whiffin et al. 2017\)](#).

## Utilité de la popmax de gnomAD pour les maladies rares

Complément possible : bases de données nationales (Frex) ou locales

Attention à quelques pièges,

- par ex. variant non rare car souvent somatique dans gnomAD
- Maladies récessives avec AF  $>1\%$  dans certaines populations

# Filtration

## Filtre de “génétique fonctionnelle”

Il est proposé de concentrer l’analyse des SNV et des indels sur

- Les régions géniques **exoniques** : les variations modifiant la séquence codante ou également les synonymes avec une prédiction d’impact sur l’épissage
- Les variations **introniques géniques** avec une annotation intronique disponible
- Les **autres variations non codantes géniques** avec une annotation disponible (ex. 3’UTR, 5’UTR, variant annoté dans une base de donnée de variants type HGMD/ClinVar/gène spécifique)
- Les régions intergéniques avec une annotation disponible avec fort niveau de preuve

Filtres selon le mode de transmission →

# DEFIDIAG : filtration WGS en solo

**Variant filtration protocols (2):**  
**SIMPLEX analysis: Analysis of the extended-OMIM panel**

To be considered in the extended-OMIM panel and exonic  $\pm 20$  bases regions

pathogenic or likely pathogenic variants from the OMIM-extended list	Any type of variant in exonic regions of these genes when the variant is annotated as pathogenic or likely pathogenic in a patient-specific, disease-non-specific variant database
Exonic and possible splice variants from the OMIM-extended list	Variants with < 4 counts in gnomAD (AR genes) or less than 3 hemizygous counts in gnomAD (X linked genes) or AF <0.01% (AD genes) AND <ul style="list-style-type: none"> <li>o Exonic, except synonymous</li> <li>o OR Intronic in the following intervals [-2, -1 ; +1,+2]</li> <li>o OR synonymous or intronic &gt;+2 or &lt;-2 and predicted to create/enhance a cryptic splicing site or significantly reduce the strength of a canonical splice site</li> </ul> CNVs or SVs that are predicted to disrupt one morbid gene, whatever the size
<b>CNVs</b>	(deletions/duplications) encompassing at least one exon of the list

These lists of variants will be loaded into the interpretation interface and interpreted following the three hypotheses:

AD hypothesis	Heterozygous variants and heterozygous and hemizygous variants mapping to the pseudo-autosomal regions of the X chromosome
AR hypothesis	Homozygous variants and at least two heterozygous variants within the same gene from -AR genes.
X-linked hypothesis	Heterozygous and hemizygous variants mapping to the X chromosome, non-pseudo-autosomal regions.

# DEFIDIAG : filtration WGS en trio

## Variant filtration protocols: TRIO Analysis

### TRIO Analysis: AD or X-linked inheritance, de novo hypothesis

The <i>de novo</i> variants analysis is performed genome wide	<ul style="list-style-type: none"><li>-Certain <i>de novo</i> variants, whatever their type (CNV, SV, SNV, indel) and their location in the genome</li><li>-Probable <i>de novo</i> variants (e.g., insufficient coverage/quality in the parents/possible germline mosaic in parents), whatever their type (CNV, SV, SNV, indel) and their location in the genome</li></ul>
---	---

Further prioritization and interpretation will be performed based on annotation data (variant quality: variants with allele count >3 in gnomAD will be excluded (as they are most likely false positives), amongst other quality criteria), variant quality in the parents, type of variation, morbid gene or not, predicted effect...).

NB: this step is unlikely to be sufficient to allow the identification of a novel disease-causing gene (outside known Mendelian genes) in the absence of recurrence data. We propose to prioritize first morbid genes (OMIM-extended list) or *de novo* CNVs with annotation suggesting a regulatory role on a non-encompassed morbid gene.

# DEFIDIAG : filtration WGS en trio

## TRIO Analysis: Autosomal recessive inheritance

Variants to be considered:

pathogenic or likely pathogenic or of unknown significance from the OMIM-extended list	Any type of variant in genomic regions of morbid genes which is annotated as pathogenic or likely pathogenic or of unknown significance in a patient variant database whatever their frequency
protein truncating variant from the OMIM-extended list	every protein truncating variant whatever the frequency
AR genes from the OMIM-extended list	Variants with < 4 homozygous in GnomAD AND <ul style="list-style-type: none"><li>o Exonic, except synonymous</li><li>o OR Intronic in the following intervals [-2,-1,+1,+2]</li><li>o OR synonymous or intronic &gt;+2 or &lt;-2 and predicted to create/enhance a cryptic splicing site or significantly reduce the prediction scores of a canonical splice site</li></ul>
CNV and SV	CNVs or SVs that disrupt or dysregulate one morbid gene, whatever the size

Will be loaded into the interpretation interface and then further prioritized/interpreted: homozygous and compound heterozygous variants. Further prioritization and interpretation will be performed based on annotation data, especially the allele counts of homozygous variants carriers in gnomAD: the analysis will first focus on variants with <3 homozygous counts homozygous in gnomAD.

NB: this step is unlikely to be sufficient to allow the identification of a novel disease-causing gene in the absence of recurrence data (genes not considered in morbid genes lists). We propose to skip the analysis of the AR hypothesis in genes not reported in morbid genes lists in this short-term, diagnostic assessment protocol. Further analyses in the context of ancillary studies will be performed

# DEFIDIAG : filtration WGS en trio

## TRIO Analysis: X-linked hypothesis, inherited variants

To be considered, regarding genes mapping to the X chromosome:

pathogenic or likely pathogenic or of unknown significance from the OMIM-extended list	Any type of variant in genomic regions of morbid genes which is annotated as pathogenic or likely pathogenic or of unknown significance in a patient variant database
Variants in morbid genes with < 4 hemizygous carriers in gnomAD from the OMIM-extended list	Variants in morbid genes with < 4 hemizygous carriers in gnomAD AND <ul style="list-style-type: none"> <li>o Exonic, except synonymous</li> <li>o OR Intronic in the following intervals [-2, -1 ; +1,+2]</li> <li>o OR synonymous or intronic &gt;+2 or &lt;-2 and predicted to create/enhance a cryptic splicing site or significantly reduce the strength of a canonical splice site</li> </ul>
CNVs or SVs	CNVs or SVs that disrupt or dysregulate one morbid gene, whatever the size

## TRIO Analysis: Other pathogenic and protein-truncating variants, including AD-inherited variants

This step will be used to recover dominant pathogenic variants that have been missed by the de novo analysis due low sequencing quality, high allelic counts in GnomAD, or because the the variants are inherited (reduced penetrance).

To be considered, regarding genes mapping to the autosomes

pathogenic or likely pathogenic or of	Any type of variant in genomic regions of morbid AD genes which is annotated as pathogenic or likely pathogenic or in a patient- variant database even if they are absent from the list of morbid genes
Protein truncating variants in morbid AD genes with0. from the OMIM-extended list	Exonic, ANDstop gain, start loss or frameshift variants or Intronic in the following intervals [-2, -1 ; +1,+2]
CNVs or SVs	CNVs or SVs that disrupt or dysregulate one morbid AD gene from the OMIM-extended list, whatever the size

# Autre stratégie : Priorisation

## Outils de priorisation

Utilisation de paramètres plus ou moins transparents pour classer les variants

Plutôt que les faire apparaître dans un ordre aléatoire / positionnel

Ex. Polyweb / Imagine - Necker

Ex. Seqone / deep learning

Etc.

- Intérêt pour la rapidité de l'analyse quand positive
- La transparence paraît nécessaire mais difficile avec deep learning
- Risque de négliger des variants mal classés : une procédure diagnostique doit être assez systématique, maladies secondaires (partie du phénotype)
- Préfigure l'interprétation de demain...?

# Utilisation du NGS pour le diagnostic moléculaire : besoins

1. Déterminer une stratégie et une cible adaptée à la question clinique
2. Déterminer un pipeline bioinformatique adapté aux besoins
3. Déterminer les annotations indispensables à afficher pour une interprétation rapide et les annotations secondaires
4. Déterminer les filtres adaptés
- 5. Interprétation et établissement du compte rendu : connaître les recommandations pour diminuer le caractère subjectif**

# Assurer l'expertise de l'interprétation des variations génétiques

*American College of Medical Genetics and Genomics (ACMG) –  
Association of Molecular Pathology (AMP)*

Class **5**: Pathogenic

Class **4**: Likely pathogenic

Class **3**: Variant of uncertain significance (VUS)

Class **2**: Likely benign

Class **1**: Benign

Bases de données  
de contrôles

Bases de données  
de malades

Prédictions  
bioinformatiques

Biostatistiques

Données de  
ségrégation

Données  
fonctionnelles

Expertise  
moléculaire

Expertise  
clinique

Confrontations  
clinico-biologiques

# Recommandations ACMG-AMP et ANPGM

**Recommandations d'interprétation de l'ANPGM = ACMG-AMP avec modifications mineures (mais en français)**

**Connaître la signification de chaque argument**

**Rester conscient que ces recommandations ne sont pas applicables à la lettre pour chaque situation**

**Très facile de « tordre » un argument pour faire passer une variation d'une classe à une autre !**

**La conviction moléculaire ET la confrontation clinique doivent être grandes**

**Existence d'outil qui « auto-cochent » : utile mais risqué !**

chr14-73637653-C-T

Link a publication | Classify | Community contributions | Favorites

Short link: [varsome.me/EWh](https://varsome.me/EWh) | API link

Chromosome: chr14 | Position: 73637653 | REF Sequence: C | ALT Sequence: T | Variant type: SNV | Cytoband: 14q24.2 | HGVS: NM\_000021.4:c.236C>T (p.Ala79Val) | RS ID: rs93749824 | [absnp](#)

UCSC genome browser | Mastermind | TraP Score

Gene symbol: PSEN1

This variant has been viewed 19 times on VarSome.

Connect with past and future viewers of this variant.

ACMG Classification - Educational use only | Version 734 | Terms of use | Documentation | Options

Verdict: **Uncertain Significance** | Minimum ClinVar rating: 1

Transcript: NM\_000021.4, canonical, protein length 468, gene PSEN1, missense variant

Rules

<input checked="" type="checkbox"/> PVS1	<input checked="" type="checkbox"/> PS1	<input checked="" type="checkbox"/> PS2	<input checked="" type="checkbox"/> PS3	<input type="checkbox"/> PS4	<input checked="" type="checkbox"/> PM1	<input checked="" type="checkbox"/> PM2	<input type="checkbox"/> PM3
<input checked="" type="checkbox"/> PM4	<input checked="" type="checkbox"/> PM5	<input checked="" type="checkbox"/> PM6	<input checked="" type="checkbox"/> PP1	<input checked="" type="checkbox"/> PP2	<input checked="" type="checkbox"/> PP3	<input type="checkbox"/> PP4	<input checked="" type="checkbox"/> PP5
<input checked="" type="checkbox"/> BA1	<input type="checkbox"/> BS1	<input checked="" type="checkbox"/> BS2	<input checked="" type="checkbox"/> BS3	<input checked="" type="checkbox"/> BS4			
<input checked="" type="checkbox"/> BP1	<input type="checkbox"/> BP2	<input checked="" type="checkbox"/> BP3	<input checked="" type="checkbox"/> BP4	<input type="checkbox"/> BP5	<input checked="" type="checkbox"/> BP6	<input checked="" type="checkbox"/> BP7	

Please tick or untick any rules to switch them on or off - the Verdict will update.

Identified criteria

Rule	Pathogenicity	Explanation
PM2	Pathogenic Moderate	GnomAD exomes allele count = 3 is less than 5 threshold for dominant gene PSEN1 (good GnomAD exomes coverage = 78.6). GnomAD genomes allele count = 1 is less than 5 threshold for dominant gene PSEN1 (good GnomAD genomes coverage = 31.6).
PP3	Pathogenic Supporting	Pathogenic computational verdict because 8 pathogenic predictions from DANN, FATHMM, FATHMM-MKL, LRT, MutationAssessor, MutationTaster, PROVEAN and SIFT with no benign predictions.
PP5	Pathogenic Supporting	ClinVar classifies variant as Pathogenic, 2 stars (above minimum of 1 star) (but no reference to functional or in-vivo studies found to support PS3). UniProt classifies this variant as 'disease' (Alzheimer Disease 3).

Region browser

Transcripts: Show 23 mature transcripts.

UniProt Protein regions

Pathogenicity

Filter:  Synonymous,  Missense,  Nonsense,  Stoploss,  Frameshift,  Inframe indel,  Non-coding

gnomAD genomes | gnomAD exomes

Variant basic info

ACMG classification

Region browser

Community contributions

Publications

Structural Variants

Transcripts

ClinVar

UniProt Variants

Frequencies

Cancer databases

Prediction scores

Beacon network

Gene PSEN1

SAPHETOR

Genome-scale end-to-end bioinformatics analysis and interpretation for single and combined NGS samples.

Variant basic info

ACMG classification

Region browser

Community contributions

Publications

Structural Variants

Transcripts

ClinVar

UniProt Variants

Frequencies

Cancer databases

Prediction scores

Beacon network

Gene PSEN1

SAPHETOR

Genome-scale end-to-end bioinformatics analysis and interpretation for single and combined NGS samples.

Recent VarSome activity

sq [Sapheto] linked the publication 'The Arg499His gain-of-function mutation in the C-terminal domain of PCSK9' to NM\_174936.3(PCSK9):c.1496G>A

sq [Sapheto] linked the publication 'Identification of a Novel C6orf57 Mutation in Iranian Patient with Clericuzio-type Poikiloderma with Neutropenia (CPN): A Case Report' to NM\_024598.3(USB1):c.757T>C

MM classified NM\_015311.3(OBSL1):c.-51A>C as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_005732.3(RAD50):c.430C>T as **Uncertain Significance**

Shaun\_Hug [Fulgent Genetics] classified NM\_00127500.3(MET):c.534C>T as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_00165962.1(ELAC2):c.1440A>G as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_000264.4(PTCH1):c.318C>T as **Likely Benign**

Emily\_S [NGen MD] classified NM\_000249.3(MLH1):c.1039-11\_1039-8delTTTT as **Likely Benign**

Emily\_S [NGen MD] classified NM\_000249.3(MLH1):c.1039-9\_1039-8delTT as **Likely Benign**

mutation in the C-terminal domain of PCSK9' to NM\_174936.3(PCSK9):c.1496G>A

sq [Sapheto] linked the publication 'Identification of a Novel C6orf57 Mutation in Iranian Patient with Clericuzio-type Poikiloderma with Neutropenia (CPN): A Case Report' to NM\_024598.3(USB1):c.757T>C

MM classified NM\_015311.3(OBSL1):c.-51A>C as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_005732.3(RAD50):c.430C>T as **Uncertain Significance**

Shaun\_Hug [Fulgent Genetics] classified NM\_00127500.3(MET):c.534C>T as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_00165962.1(ELAC2):c.1440A>G as **Benign**

Shaun\_Hug [Fulgent Genetics] classified NM\_000264.4(PTCH1):c.318C>T as **Likely Benign**

Emily\_S [NGen MD] classified NM\_000249.3(MLH1):c.1039-11\_1039-8delTTTT as **Likely Benign**

Emily\_S [NGen MD] classified NM\_000249.3(MLH1):c.1039-9\_1039-8delTT as **Likely Benign**

PatriciaCouto [Hermes Pardini] classified NM\_000059.3(BRCA2):c.2350A>G as **Likely Benign**

Emily\_S [NGen MD] classified

# Recommandations ACMG-AMP et ANPGM / NGS diag

## Classe 5 : Variant pathogène

1 **PVS** ET  $\geq 1$  **PS** OU  
 $\geq 2$  **PM** OU  
1 **PM** ET 1 **PP** OU  
 $\geq 2$  **PP** OU  
 $\geq 2$  **PS** OU  
1 **PS** ET  $\geq 3$  **PM** OU  
2 **PM** ET  $\geq 2$  **PP** OU  
1 **PM** ET  $\geq 4$  **PP**

## Classe 4 : Variant probablement pathogène

1 **PVS** ET 1 **PM** OU  
1 **PS** ET 1-2 **PM** OU  
1 **PS** ET  $\geq 2$  **PP** OU  
 $\geq 3$  **PM** OU  
2 **PM** ET  $\geq 2$  **PP** OU  
1 **PM** ET  $\geq 4$  **PP**

→ Variant rendu comme  
« causal », conseil génétique  
possible le plus souvent  
→ Pathogène ne veut pas dire  
100% pénétrant !

## Classe 3 : Variant de signification inconnue

- Les arguments pondérés ne permettent pas de classer le variant dans une autre classe
- Ne doit pas figurer sur le CR ou sur la même page du CR selon les recommandations ANPGM (page 2, courrier séparé)
- Pas de conseil génétique

## Classe 2 : Variant probablement bénin

1 **BS** ET 1 **BP** OU  
 $\geq 2$  **BP**

## Classe 1 : Variant bénin

**BA** OU  
 $\geq 2$  **BS**

→ Non rendus

	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 faux-sens prédit bénin par l'ensemble des logiciels de prédiction de pathogénicité interrogés (BP4)  Variant faux-sens dans un gène où seules les variants tronquants sont décrits comme associées 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)  Indel en phase dans une région répétée sans fonction connue (BP3)	Effet d'un variant faux-sens 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)  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 <i>in vivo</i> ou <i>in vitro</i> 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 variants 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 <i>in vivo</i> ou <i>in vitro</i> 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 ségrégeant pas avec la pathologie chez des apparentés atteints (BS4)		Données issues d'une seule famille : $N \leq 1/8$ : argument faible (PP)*  Données issues de plusieurs familles : $N \leq 1/4$ : argument faible (PP)*	Données issues d'une seule famille : $N \leq 1/16$ : argument moyen (PM)*  Données issues de plusieurs familles $N \leq 1/8$ : argument moyen (PM)*	Données issues d'une seule famille : $N \leq 1/32$ : argument fort (PS)*  Données issues de plusieurs familles : $N \leq 1/16$ : 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-occurrence du variant avec un variant pathogène en <i>trans</i> (dans le même gène): poids de l'argument laissé à l'appréciation du laboratoire (classé BP2 dans les critères ACMG initiaux)  Co-occurrence du variant avec un variant pathogène en <i>cis</i> (dans le même gène): poids de l'argument laissé à l'appréciation du laboratoire (classé BP2 dans les critères ACMG initiaux)		Variant observé en <i>trans</i> 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-occurrence du variant avec un variant pathogène dans un autre gène impliqué dans la même pathologie : poids de l'argument laissé à l'appréciation du laboratoire (classé BP5 dans les critères ACMG initiaux)	Le phénotype du patient ou les antécédents familiaux sont très spécifiques pour le gène (PP4)			

# Interprétation des données pangénomiques diagnostiques

**Choix local : phenotype-first**

**Intérêt de la bonne description phénotypique / accès à l'information :**

- **doubles diagnostics !**
- **Résultats partiels**

**Outils prenant en compte les entrées phénotypiques (codes HPO, mots clé) :  
une aide à la priorisation – à optimiser**

**Intérêt d'une double lecture biologique**

**Confrontations clinico-biologiques ++**

**Implication du clinicien dans l'interprétation = avant le compte rendu**

# Autres besoins

✓ QC

## Qualité et traçabilité globale

- Données « brutes » : couverture, profondeur ROI
- Nombre moyen de variants
- Traçabilité versions du pipeline et bases de données → réinterprétation des négatifs
- Traçabilité de l'interprétation
- Base de donnée locale
- Partage de l'interprétation interlaboratoires

# Conclusions

**Pas de stratégie universelle :**

**Nécessité d'une adéquation Indication clinique – choix moléculaires – pipelines – annotation – filtration – interprétation**

**Choix du phenotype first + confrontation clinico-biologique**

**Des règles et recommandations communes**

**Vers l'IA permettant l'aide à la priorisation / classification**

**En attendant : des données bien organisées et limitées en nombre permettent d'éviter de se noyer, en autorisant un accès secondaire à d'autres données à la demande**

# Remerciements

Aux bio-informaticiens



Sophie  
Coutant



Olivier  
Quenez



Raphaël  
Lanos



Myriam  
Vezain-Mouchard

Aux biologistes



Pascale  
Saugier-veber



François  
Lecoquierre



Kévin  
Cassinari



Stéphanie  
Baert-Desurmont



Edwige  
Kasper



Claude  
Houdayer



Thierry  
Frebourg

Aux membres de la plateforme de génomique

Isabelle Tournier,  
Myriam Vezain,  
Françoise Charbonnier

Aux techniciennes