



ACMG遗传变异分类标准与指南

朱凤娇、刘博

目录

content

01

背景

02

标准术语

03

命名规则

04

文献及数据库使用

05

预测软件

06

解读基因变异依据的标准

07

注意问题



1. 背景



美国医学遗传学与基因组学学会(The American College of Medical Genetics and Genomics, ACMG)于2015年编写和发布了《ACMG遗传变异分类标准与指南》。

- 为了评估临床实验室的观点,对列入GeneTests.org上位于美国和加拿大的超过100家的测序实验室进行了调研,要求各实验室填写参考术语及变异分类的评估证据。这些实验室有检测包括罕见病、药物基因组学和癌症体细胞突变的经验。
- 第二次调研在GeneTests.org上的相同实验室以及AMP清单上的约2000个单位中进行,同时给各单位提供了分类方案和详细的方案补充说明,要求各实验室使用该分类方案并对以下内容进行反馈,包括各标准的适宜性和每个标准的相对权重、分类体系的易用性以及他们是否会在自己的实验室采用这样的体系。来自超过33个实验室的答复表明多数实验室支持所推荐的分类方案,同时,他们的反馈进一步地指导了标准和指南的完善。
- 其后进行变异分类研讨,提出了修订后的分类标准和两个评分体系。虽然大多数回复更倾向于分数评价体系,但每个标准中具体分数的设置量化了对每个标准的理解,但是这一量化指标目前缺乏科学依据,并且没有考虑遗传证据解读时的复杂性。

适应范围:

该变异分类方法适用于所有孟德尔基因变异,包括单基因、多基因panel、外显子组和基因组测序发现的变异。期望这种变异分类方法会随着技术和知识水平提高而与时俱进。由于不同基因和不同疾病中的应用和加权评估的标准可能不同,特定疾病组的工作应继续,以制定更有针对性的具体基因的变异分类指南。



2. 标准术语

突变 (mutation) 是指核苷酸序列的永久改变；多态性 (polymorphism) 是指频率在 1% 以上的变异 (variants)。但是这两个概念很容易导致混淆，在这里我们建议采用“变异”这个术语，这样分类：

- 1) 致病性的 (pathogenic)
- 2) 可能致病的 (likely pathogenic)
- 3) 意义不明确的 (uncertain significance)
- 4) 可能良性的 (likely benign)
- 5) 良性的 (benign)

在对致病变异进行解读时，建议注明疾病和遗传方式，如

c.1521_1523delCTT (p.Phe508del), pathogenic, cystic fibrosis, autosomal recessive

对应中文：

c.1521_1523delCTT (p.Phe508del)，致病性，囊胞性纤维症，常染色体隐性

注意：

- 意义不确定 (uncertain significance) 在这里和 ii, iv 没有合并在一起，而是单独列出一项。
- 可能 (likely)：这些变异具有很高的可能性是致病或良性，这里的意思是成为致病或良性的可能性大于 90%。
- 本指南推荐的术语与细胞遗传学基因芯片检测的拷贝数变异分类不同。



3.命名规则：统一变异命名

- Human Genome Variation Society (HGVS, 人类基因组变异协会) 关于序列变异描述进行了规范 (见网址 <http://www.hgvs.org/mutnomen>, HGVS), 使用时**注意版本问题**。同时 HGVS 还提供了工具可以查询变异命名 (<https://mutalyzer.nl>)。除非另有说明, 一般推荐该命名法作为确定变异命名的首要准则!
- 临床报告应该包括参考序列以确保不混淆 DNA 命名, 另外也要同时提供编码基因、蛋白命名规则, 以更好的解读, 例如 “g.” 代表基因组序列 (genomic sequence), “c.” 代表编码 DNA 序列 (coding DNA sequence), “p.” 代表蛋白质 (protein), “m.” 代表线粒体 (mitochondria)。
- 编码基因命名原则是用 ATG 起始密码子中的 “A” 代表位置 1。那么, 刚才出现的 c.1521_1523delCTT 代表的**意思就是编码 DNA 序列第1521 到 1523 个位置的 CTT 缺失, 这个属于 deletion**。另外还有替换 (substitutions), 重复 (duplications), 插入 (insertions), 缺失/插入 (deletion/insertions, indels), 倒置 (inversions), 转换 (conversions), 易位 (translocations), 一个基因或/和染色体上两个或更多变异 (two or more changes in one gene or/and chromosome), 重复序列 (repeated sequences), 复杂变异 (complex)。**注意!!! 这些都是 DNA 水平的。**
- HGVS(<https://www.hgvs.org/mutnomen>) 并未覆盖所有类型的变异(如复杂变异), 但已有对复杂变异描述的报道。此外, ACMG 支持 HGVS 命名规则之外的三种特殊例外:
 - (i) 除了 “*” 和 “Ter (termination)” 外, “X” 也可以表示 “无义变异”
 - (ii) 外显子的编号应该根据所选的参考转录本制定 (就像 NM_000492.3:c.1521_1523delCTT, 应该加上 NM_000492.3)
 - (iii) 建议 “致病性” 替代 “功能影响”, 因为临床解释一般是评价致病的情况。




3.命名规则：统一变异命名

- 当描述编码变异时,应该在报告中使用和提供每个基因的一个参考转录本. **该转录本应该是最长的已知转录本或者是最具临床相关性的转录本.** (目前临床产品使用的转录本是否符合该规则?)
- 然而,当这些区域发生临床可解释的已知变异时,实验室应该评估该变异对所有临床相关的转录本的影响,包括含有其他外显子或非翻译区延伸的可变剪切转录本.

建议使用的参考序列、基因组、转录本采用的数据库，如下所示：

| 类型 | 参考数据库及说明 |
|-------|---|
| 参考序列 | National Center for Biotechnology Information RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/), 需带有版本号; 或者 Locus Reference Genomic database (http://www.lrg-sequence.org) |
| 参考基因组 | 标准基因组组装，如 hg19 |
| 参考转录本 | Locus Reference Genomic, Consensus CDS Database, Human Gene Mutation Database (http://www.hgmd.cf.ac.uk), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar), 或者特异基因座数据库。 |



The Human Gene Mutation Database

at the Institute of Medical Genetics in Cardiff

[Home](#) [Search](#) [help](#) [Statistics](#) [New genes](#) [What is new](#) [Background](#) [Publications](#) [Contact](#) [Register](#) [Login](#) [LSDBs](#) [Other links](#)

Gene symbol

Go!

Symbol:

Missense/nonsense

Go!

HGMD® 简介

人类基因突变数据库(The Human Gene Mutation Database,HGMD®)是通用型数据库的代表,它全面收集引起人类遗传疾病或与人类遗传疾病相关的核基因突变。此数据库建立的初衷是用于突变机制的分析,但由于它收录最新的、完整的有关人类疾病突变谱的参考数据,HGMD 至今已获得了更为广泛的应用。

HGMD 系统收录文献报道中的所有致病突变和与疾病相关的功能多态并将这些数据以易于读取的格式提供给感兴趣的各方面,不管他们是来自学术、临床还是商业领域。HGMD 收集的数据包括了单碱基置换(比如编码序列中的错义突变和无义突变以及调控和剪切区域中的点突变)、微缺失(micro-deletions)和微插入(micro-insertions)、缺失/插入(indels)、重复序列扩增以及大的基因损伤(缺失、插入和扩增)和复杂的基因重组。

HGMD 通过互联网(<http://www.hgmd.org>)免费提供给注册的(registered)学术和非盈利用户,但此公共网站只提供那些已收录三年以上的数据。最新的数据可从 HGMD 专业版(HGMD Professional)通过我们的商业合作伙伴 BIOBASE GmbH 以付费的形式获得。除了最新的突变数据,HGMD 专业版还提供公共网站上不具有的功能强大的搜索工具以及与基因和突变有关的特别信息。HGMD 专业版每三个月更新一次。

Stenson *et al.* (2020) The Human Gene Mutation Database (HGMD ®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet* epub. doi: 10.1007/s00439-020-02199-3 [[PubMed](#)]



4. 文献及数据库使用

表 1 人群数据库、疾病特异性数据库和序列数据库

| 数据库 | 特征 |
|---|--|
| 人群数据库 | |
| Exome Aggregation Consortium (http://exac.broadinstitute.org/) | 本数据库中的变异信息是通过对 61486 个独立个体进行全外显子测序获得。同时也是多种特殊疾病和群体遗传学研究中的一部分。库中不包括儿科疾病患者及其相关人群。 |
| Exome Variant Server (http://evs.gs.washington.edu/EVS) | 本数据库中的变异信息是通过对几个欧洲和非洲裔大规模人群的全外显子测序获得。当缺乏变异信息时，默认该数据已覆盖。 |
| 1000 Genomes Project (http://browser.1000genomes.org) | 本数据库中的变异信息是通过对 26 个种群进行低覆盖度的全基因组测序和高覆盖度的靶序列测序获得。本库所提供的信息比 Exome Variant Server 更具多样性，但也包含有低质量的数据，有些群体中还包含有关联性个体在内。 |
| dbSNP(http://www.ncbi.nlm.nih.gov/snp) | 本数据库由多种来源获得的短片段遗传变异(通常≤50 bp)信息组成。库中可能缺乏溯源性研究的细节，也可能包含致病性突变在内。 |
| dbVar(http://www.ncbi.nlm.nih.gov/dbvar) | 本数据库由多种来源获得的基因结构变异(通常>50 bp)信息组成。 |
| 疾病数据库 | |
| ClinVar(http://www.ncbi.nlm.nih.gov/clinvar) | 对变异与表型和临床表型之间的关联进行确定的数据库。 |
| OMIM(http://www.omim.org) | 本数据库所含人类基因和相关遗传背景，同时具有疾病相关基因遗传变异的代表性样本收录与遗传疾病典型相关的样本变异信息。 |
| Human Gene Mutation Database(http://www.hgmd.org) | 本数据库中的变异注释有文献发表。库中大部分内容需付费订阅。 |
| 其他特殊数据库 | |
| Human Genome Variation Society (http://www.hgvs.org/dblist/dblist.html) | 本数据库由人类基因组变异协会(HGVS)开发，提供数千种专门针对人群中的特殊变异进行的注释。数据库很大一部分是基于 Leiden Open Variation Database system 建立。 |
| Leiden Open Variation Database(http://www.lovd.nl) | |
| DECIPHER(http://decipher.sanger.ac.uk) | 使用 Ensemble 基因组浏览器，将基因芯片数据和临床表型进行关联，便于临床医生和研究人员使用的细胞分子遗传学数据库。 |
| 序列数据库 | |
| NCBI Genome(http://www.ncbi.nlm.nih.gov/genome) | 人类全基因组参考序列的来源。 |
| RefSeqGene(http://www.ncbi.nlm.nih.gov/refseq/rsg) | 医学相关基因参考序列。 |
| Locus Reference Genomic (LRG)(http://www.lrg-sequence.org) | |
| MitoMap(http://www.mitomap.org/MITOMAP/HumanMitoSeq) | 对“剑桥版-人类线粒体 DNA 参考序列”进行修订后形成。 |

EVS (Exome Variant Server) 收录外显子组测序项目，发布的 ESP6500数据由一组2203名非裔美国人和4300名欧洲裔美国人的无关个体组成，共计6503个样本。

dbVar数据库储存了大规模基因组变异相关的信息，包括大插入，缺失，易位和倒位。除了分类突变之外，dbVar还存储已定义突变与表型信息的关联。dbvar与来自EBI的DGVa是互作关系，共同提供了包含健康人类与患病人类的结构变异信息，而**DGV数据库**则只包含了健康人群的染色体结构变异(大于50bp以上)信息。可以作为研究拷贝数变异和表型之间关系的这类课题的对照组数据使用。



4. 文献及数据库使用

人群数据库

人群数据库

Exome Aggregation Consortium
(<http://exac.broadinstitute.org/>)

Exome Variant Server
(<http://evs.gs.washington.edu/EVS/>)

1000 Genomes Project
(<http://browser.1000genomes.org/>)

dbSNP(<http://www.ncbi.nlm.nih.gov/snp>)

dbVar(<http://www.ncbi.nlm.nih.gov/dbvar>)

本数据库中的变异信息是通过 61486 个独立个体进行全外显子测序获得。同时也是多种特殊疾病和群体遗传学研究的一部分。库中不包括儿科疾病患者及其相关人群。

本数据库中的变异信息是通过几个欧洲和非洲裔大规模人群的全外显子测序获得。当缺乏变异信息时，默认该数据已覆盖。

本数据库中的变异信息是通过 26 个种群进行低覆盖度的全基因组测序和高覆盖度的靶序列测序获得。本库所提供的信息比 Exome Variant Server 更具多样性，但也包含有低质量的数据，有些群体中还包含有关联性个体在内。

本数据库由多种来源获得的短片段遗传变异(通常 ≤ 50 bp)信息组成。库中可能缺乏溯源性研究的细节，也可能包含致病性突变在内。

本数据库由多种来源获得的基因结构变异(通常 > 50 bp)信息组成。

特点：获取某变异在大规模人群中发生频率的相关信息，不仅来源于健康个体，也包含致病性的变异。

缺点：人群数据库并不包含变异的功能效应及可能关联的表型等相关信息。（该缺点可以通过疾病数据库来弥补）

注意：明确数据库收录的是健康群体的信息还是患病群体的信息？（人群不同结果不同）

数据库是否收录了同一家庭多名成员的信息？（数据的离散性的保证）

数据库收录的受试者的年龄范围？（成人和儿童疾病是不同的）

不同人群的信息？（人种差异）

数据库更新时间

| rs1425711270 | | Current Build 154 Released April 21, 2020 | |
|----------------|--|--|----------------------------------|
| Organism | Homo sapiens | Clinical Significance | Not Reported in ClinVar |
| Position | chr6:90586883 (GRCh38.p12)  | Gene : Consequence | MAP3K7 : Initiator Codon Variant |
| Alleles | T>C | Publications | 0 citations |
| Variation Type | SNV Single Nucleotide Variation | Genomic View | See rs on genome |
| Frequency | C=0.00003 (1/31352, GnomAD) C=0.0 (0/10680, ALFA Project) | | |

| Study | Population | Group | Sample Size | Ref Allele | Alt Allele |
|----------------------------------|----------------------------------|------------|-------------|------------|------------|
| gnomAD - Genomes | Global | Study-wide | 31352 | T=0.99997 | C=0.00003 |
| gnomAD - Genomes | European | Sub | 18884 | T=1.00000 | C=0.00000 |
| gnomAD - Genomes | African | Sub | 8686 | T=0.9999 | C=0.0001 |
| gnomAD - Genomes | East Asian | Sub | 1558 | T=1.0000 | C=0.0000 |
| gnomAD - Genomes | Other | Sub | 1086 | T=1.0000 | C=0.0000 |
| gnomAD - Genomes | American | Sub | 848 | T=1.000 | C=0.000 |
| gnomAD - Genomes | Ashkenazi Jewish | Sub | 290 | T=1.000 | C=0.000 |



4. 文献及数据库使用

疾病数据库

疾病数据库

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

对变异与表型和临床表型之间的关联进行确定的数据库。

OMIM(<http://www.omim.org>)

本数据库所含人类基因和相关遗传背景，同时具有疾病相关基因遗传变异的代表性样本收录与遗传疾病典型相关的样本变异信息。

Human Gene Mutation Database(<http://www.hgmd.org>)

本数据库中的变异注释有文献发表，库中大部分内容需付费订阅。

特点： 包含病患中发现的变异以及对其致病性的评估。

缺点： 经包含一些分类错误的变异，这些变异在已发表的同行评审的文献中被错误判定，而多数数据库在收录变异相关信息时并未对证据进行基本的审核。（不过 ClinVar 是执行的）



4. 文献及数据库使用

当使用数据库时，临床实验室应注意：

- (i) 确定数据库的更新频率，确定数据库收录相关数据时是否进行了校勘，以及采用什么方法进行数据校勘；
- (ii) 确认采用 HGVs 命名体系，并确定描述变异的基因组版本和转录本参考序列；
- (iii) 确定数据分析准确度的验证程度(如变异是源自于低覆盖的新一代测序,还是通过了 Sanger 测序验证), 并分析用于评估数据准确度的各种指标, 要获得这些信息可能需要阅读 相关的文献；
- (iv) 确定收录对象的来源及其唯一性。

当使用文献检索时，需要注意：

- (i) 注意那些使用较老命名和分类规则的文献，也要注意那些只基于一个观察报告而来的变异。
- (ii) 当鉴定一个个体和家系中的变异和表型时，要注意是如何确定患者的病情的。
- (iii) 因为患者和相关个体可能被多篇文献报道，这些报道可能作者有交叉，或者有相关合作，或者一个家庭成员被不同的临床系统随访，这样会导致变异频率有一定的提高。分辨这些可以看作者是否有交叉，且发表单位是否一样。

临床实验室的要求

- (i) 实验室应该有自己的的一套系统，当每一个基因和临床主张报道时，都可以跟踪到。
- (ii) 积极提交变异数据到相关数据库,比如 ClinVar, 此数据库包括临床主张和变异分类，帮助大家更好的理解变异。
- (iii) 提供临床数据应遵循“健康保险携带和责任法案(HIP AA)”对个人隐私保护的规定。
- (iv) 临床实验室应与临床医生合作，以获得临床信息，从而更好地理解基因型是如何影响临床表型的，并解决不同实验室对遗传变异解读存在差异的问题。
- (v) 临床变异数据库极大地促进临床实验室工作的开展，因此需对其进行扩展并标准化. 标准化便于临床实验室获取数据库的最新信息，同时有助于提交更新的信息. 例如，ClinVar 数据库允许变异连同临床表型和诊断相关信息一并提交，同时追踪提交变异的审核状态，以便对校勘质量的水平提供一个更加透明的概貌。



5. 预测软件

表 2 生物信息分析工具

| 分类 | 名称 | 网站 | 依据 |
|---------|-----------------------|---|---|
| 错义预测 | Consurf | http://consurf.tau.ac.il | 进化保守性 |
| | FATHMM | http://fathmm.biocompute.org.uk | 进化保守性 |
| | MutationAsses | http://mutationassessor.org | 进化保守性 |
| | PANTHER | http://www.pantherdb.org/tools/csnpscoreForm.jsp | 进化保守性 |
| | PhD-SNP | http://snps.biofold.org/phd-snp/phd-snp.html | 进化保守性 |
| | SIFT | http://sift.jcvi.org | 进化保守性 |
| | SNP&GO | http://snps-and-go.biocomp.unibo.it/snps-and-go | 蛋白结构/功能 |
| | Align GVGD | http://agvgd.iarc.fr/agvgd_input.php | 蛋白结构/功能和进化保守性 |
| | MAPP | http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html | 蛋白结构/功能和进化保守性 |
| | MutationTaster | http://www.mutationtaster.org | 蛋白结构/功能和进化保守性 |
| | MutPred | http://mutpred.mutdb.org | 蛋白结构/功能和进化保守性 |
| | PolyPhen-2 | http://genetics.bwh.harvard.edu/pph2 | 蛋白结构/功能和进化保守性 |
| | PROVEAN | http://provean.jcvi.org/index.php | 变异序列和蛋白序列同源性之间的相似性比对和测量 |
| | nsSNPAnalyzer | http://snpanalyzer.uthsc.edu | 多序列比对和蛋白结构分析 |
| 剪接位点预测 | Condel | http://bg.upf.edu/fannsdb/ | 综合 SIFT, PolyPhen-2 和 MutationAssessor 进行综合预测 |
| | CADD | http://cadd.gs.washington.edu | 对于来自模拟变异的等位基因进行不同的注释 |
| | GeneSplicer | http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml | 马尔可夫模型 |
| | Human Splicing Finder | http://www.umd.be/HSF/ | 位置依赖的逻辑 |
| | MaxEntScan | http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html | 最大熵原则 |
| | NetGene2 | http://www.cbs.dtu.dk/services/NetGene2 | 神经网络 |
| | NNSplice | http://www.fruitfly.org/seq_tools/splice.html | 神经网络 |
| 核酸保守性预测 | FSPLICE | http://www.softberry.com/berry.phtml?topic=fsplce&group=programs&subgroup=gfind | 基于权重矩阵模型进行种特异性预测 |
| | GERP | http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html | 基因组进化速率分析 |
| | PhastCons | http://compugen.bscb.cornell.edu/phast/ | 保守打分及鉴定保守元件 |
| | PhyloP | http://compugen.bscb.cornell.edu/phast/ | |
| | | http://compugen.bscb.cornell.edu/phast/help-pages/phyloP.txt | 比对和分子进化树: 在家系特异或者所有分支中, 计算保守或者加速的 <i>P</i> 值 |

各种公共和商业计算机工具可以辅助解读序列变异. 这些工具主要分为两类: 一类可以预测错义变异是否会破坏蛋白质的功能或结构; 另一种可以预测是否影响剪接(表 2). 新的工具已可以处理额外的非编码序列。



5. 预测软件

基因错义改变对基因功能的影响与一些因素有关，比如这个氨基酸或碱基序列的进化保守性，突变处于蛋白序列的位置和环境，以及氨基酸替代导致的结果。在这些算法中就是综合了这些因素来评价错义改变的影响力。

这些算法是否准确呢？它们都是经过评价的，比如会对那些已知的可以影响疾病的基因变异进行评价。对于错义算法的预测，准确率大致在 65%-80%。但是也存在一些问题，大多数工具的特异性较低，导致有些错义改变被过度预测为有害突变，而且对于影响较小的错义变异的预测也不可靠。目前临床实验室常用的错义变异解读工具有 PolyPhen 2, SIFT 和 MutationTaster.

SIFT 软件是基于保守性的一个算法，简单的说就是将自己的序列和其他同源蛋白进行比对，会得到一个数值，用以评价是否 deleterious，是否中性。

剪接位点的预测工具可以预测外显子水平和内含子水平。相对于特异性(60%-80%)，其具有较高敏感性(90%-100)。建议采用多个软件对序列变异进行解读，因为不同的软件有他们的优点和缺点。

注意：这些只是预测，所以使用的时候尤其要注意，我们不建议仅通过这一来源而做临床论证。



5. 预测软件

错义改变

原始
序列

DNA: 5' – AAC AGC CTG C**G**T ACG GCT CTC – 3'
3' – TTG TCG GAC GCA TGC CGA GAG – 5'
mRNA: 5' – AAC AGC CUG CGU GCG ACG CUC – 3'
蛋白质: Asn Ser Leu Arg Thr Ala Leu

变异
序列

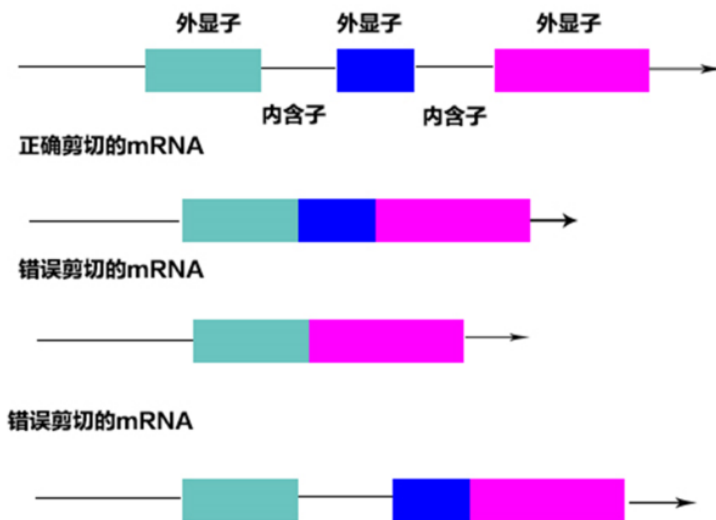
DNA: 5' – AAC AGC CTG C**T**T ACG GCT CTC – 3'
3' – TTG TCG GAC GAA TGC CGA GAG – 5'
mRNA: 5' – AAC AGC CUG C**U**U GCG ACG CUC – 3'
蛋白质: Asn Ser Leu **Leu** Thr Ala Leu

取代一个单核苷酸

不正确的氨基酸

剪接位点改变是指前体RNA变为成熟RNA的过程中发生可变剪切，在剪接识别位点发生的插入、缺失和碱基改变可能会导致剪接位点的改变。

例如，遗传性痴呆与17号染色体的exon10（外显子）处的可变剪接突变有关。





5. 预测软件

基因错义改变对基因功能的影响与一些因素有关，比如这个氨基酸或碱基序列的进化保守性，突变处于蛋白序列的位置和环境，以及氨基酸替代导致的结果。在这些算法中就是综合了这些因素来评价错义改变的影响力。

这些算法是否准确呢？它们都是经过评价的，比如会对那些已知的可以影响疾病的基因变异进行评价。对于错义算法的预测，准确率大致在 65%-80%。但是也存在一些问题，大多数工具的特异性较低，导致有些错义改变被过度预测为有害突变，而且对于影响较小的错义变异的预测也不可靠。目前临床实验室常用的错义变异解读工具有 PolyPhen 2, SIFT 和 MutationTaster.

SIFT 软件是基于保守性的一个算法，简单的说就是将自己的序列和其他同源蛋白进行比对，会得到一个数值，用以评价是否 deleterious，是否中性。

剪接位点的预测工具可以预测外显子水平和内含子水平。相对于 特异性(60%-80%)，其具有较高敏感性(90%-100)。建议采用多个软件对序列变异进行解读，因为不同的软件有他们的优点和缺点。

注意：这些只是预测，所以使用的时候尤其要注意，我们不建议仅通过这一来源而做临床论证。



5. 预测软件

基因错义改变对基因功能的影响与一些因素有关，比如这个氨基酸或碱基序列的进化保守性，突变处于蛋白序列的位置和环境，以及氨基酸替代导致的结果。在这些算法中就是综合了这些因素来评价错义改变的影响力。

这些算法是否准确呢？它们都是经过评价的，比如会对那些已知的可以影响疾病的基因变异进行评价。对于错义算法的预测，准确率大致在 65%-80%。但是也存在一些问题，大多数工具的特异性较低，导致有些错义改变被过度预测为有害突变，而且对于影响较小的错义变异的预测也不可靠。目前临床实验室常用的错义变异解读工具有 PolyPhen 2, SIFT 和 MutationTaster.

SIFT 软件是基于保守性的一个算法，简单的说就是将自己的序列和其他同源蛋白进行比对，会得到一个数值，用以评价是否 deleterious，是否中性。

剪接位点的预测工具可以预测外显子水平和内含子水平。相对于特异性(60%-80%)，其具有较高敏感性(90%-100)。建议采用多个软件对序列变异进行解读，因为不同的软件有他们的优点和缺点。

注意：这些只是预测，所以使用的时候尤其要注意，我们不建议仅通过这一来源而做临床论证。



6. 解读基因变异依据的标准

- 指南针对的是遗传性变异（孟德尔遗传疾病相关变异），而并非体细胞变异、药物基因组变异、或者是针对非孟德尔复杂疾病。
- 指南可能比实验室所使用的严格，这样会有很多基因成为“意义不明确”。有些变异被认为会引起疾病，但是却缺乏足够的证据，我们希望此指南可以减少这些变异的数量。需要谨记一点：当一个变异被认为是致病，那么医疗工作者会依据临床实验室提供的建议进行干预，或者改变治疗方法，或者会对非致病基因型携带者不再干预。**因此应当注意。**
- 指南提供了两套标准，一套标准是对致病和可能致病变异证据的分类，一套是对良性或可能良性的变异证据进行分类。致病变异的证据可分为**非常强 (very strong, PVS1)**，**强 (strong, PS1-4)**；**中等 (moderate, PM1-6)**，或**支持 (supporting, PP1-5)**。良性变异证据可分为**独立 (stand-alone, BA1)**，**强 (strong, BS1-4)**，或**支持 (BP1-6)**。**注意：数字只是作为分类标注，并不具有任何意义。**

这些规则适合各类变异，无论是来自目前的案例，还是已被审核发表的数据及未发表的数据。未发表的数据可仍一些公共数据库（ClinVar或者位点特异性数据库）或实验室自有数据库中获得。



6. 解读基因变异依据的标准

表 5 遗传变异分类联合标准规则

| 标准 | 描述 |
|--------|---|
| 致病的 | (i) 1 个非常强(PVS1)和 (a) ≥ 1 个强(PS1~PS4)或 (b) ≥ 2 个中等(PM1~PM6)或 (c) 1 个中等(PM1~PM6)和 1 个支持(PP1~PP5)或 (d) ≥ 2 个支持(PP1~PP5) (ii) ≥ 2 个强(PS1~PS4)或 (iii) 1 个强(PS1)和 (a) ≥ 3 个中等(PM1~PM6)或 (b) 2 个中等(PM1~PM6)和 ≥ 2 个支持(PP1~PP5)或 (c) 1 个中等(PM1~PM6)和 ≥ 4 个支持(PP1~PP5) |
| 可能致病的 | (i) 1 个非常强(PVS1)和 1 个中等(PM1~PM6)或 (ii) 1 个强(PS1~PS4)和 1~2 个中等(PM1~PM6)或 (iii) 1 个强(PS1~PS4)和 ≥ 2 个支持(PP1~PP5)或 (iv) ≥ 3 个中等(PM1~PM6)或 (v) 2 个中等(PM1~PM6)和 ≥ 2 个支持(PP1~PP5)或 (vi) 1 个中等(PM1~PM6)和 ≥ 4 个支持(PP1~PP5) |
| 良性的 | (i) 1 个独立(BA1)或 (ii) ≥ 2 个强(BS1~BS4) |
| 可能良性的 | (i) 1 个强(BS1~BS4)和 1 个支持(BP1~BP7)或 (ii) ≥ 2 个支持(BP1~BP7) |
| 意义不明确的 | (i) 不满足上述标准或 (ii) 良性和致病标准相互矛盾 |



6.1 致病变异分级标准

表 3 致病变异分级标准

| 致病性证据 | 分类 |
|-------|--|
| 非常强 | <p>PVS1: 当一个疾病的致病机制为功能丧失(LOF)时, 无功能变异(无义突变、移码突变、经典±1 或 2 的剪接突变、起始密码子变异、单个或多个外显子缺失). 注: 1, 该基因的 LOF 是否是导致该疾病的明确致病机制(如 GFAP, MYH7); 2, 3'端末端的功能缺失变异需谨慎解读; 3, 需注意外显子选择性缺失是否影响到蛋白质的完整性; 4, 考虑一个基因存在多种转录本的情况.</p> |
| 强 | <p>PS1: 与先前已确定为致病性的变异有相同的氨基酸改变. 例如: 同一密码子, G>C 或 G>T 改变均可导致缬氨酸→亮氨酸的改变. 注意剪切影响的改变.</p> <p>PS2: 患者的新发变异, 且无家族史(经双亲验证). 注: 仅仅确认父母还不够, 还需注意捐卵、代孕、胚胎移植的差错等情况.</p> <p>PS3: 体内、体外功能实验已明确会导致基因功能受损的变异. 注: 功能实验需要验证是有效的, 且具有重复性与稳定性.</p> <p>PS4: 变异出现在患病群体中的频率显著高于对照群体. 注: 1, 可选择使用相对风险值或者 OR 值来评估, 建议位点 OR 大于 5.0 且置信区间不包括 1.0 的可列入此项(详细见指南正文); 2, 极罕见的变异在病例对照研究可能无统计学意义, 原先在多个具有相同表型的患者中观察到该变异且在对照中未观察到可作为中等水平证据.</p> |
| 中等 | <p>PM1: 位于热点突变区域, 和/或位于已知无良性变异的关键功能域(如酶的活性位点).</p> <p>PM2: ESP 数据库、千人数据库、EXAC 数据库中正常对照人群中未发现的变异(或隐性遗传病中极低频位点)(表 7). 注: 高通量测序得到的插入/缺失人群数据质量较差.</p> <p>PM3: 在隐性遗传病中, 在反式位置上检测到致病变异. 注: 这种情况必须通过患者父母或后代验证.</p> <p>PM4: 非重复区框内插入/缺失或终止密码子丧失导致的蛋白质长度变化.</p> <p>PM5: 新的错义突变导致氨基酸变化, 此变异之前未曾报道, 但是在同一位点, 导致另外一种氨基酸的变异已经确认是致病性的, 如: 现在观察到的是 Arg156Cys, 而 Arg156His 是已知致病的. 注意剪切影响的改变.</p> <p>PM6: 未经父母样本验证的新发变异.</p> |
| 支持证据 | <p>PP1: 突变与疾病在家系中共分离(在家系多个患者中检测到此变异). 注: 如有更多的证据, 可作为更强的证据.</p> <p>PP2: 对某个基因来说, 如果这个基因的错义变异是造成某种疾病的原因, 并且这个基因中良性变异所占的比例很小, 在这样的基因中所发现的新的错义变异.</p> <p>PP3: 多种统计方法预测出该变异会对基因或基因产物造成有害的影响, 包括保守性预测、进化预测、剪接位点影响等. 注: 由于做预测时许多生物信息学算法使用相同或非常相似的输入, 每个算法不应该算作一个独立的标准. PP3 在一个任何变异的评估中只能使用一次.</p> <p>PP4: 变异携带者的表型或家族史高度符合某种单基因遗传疾病.</p> <p>PP5: 有可靠信誉来源的报告认为该变异为致病的, 但证据尚不足以支持进行实验室独立评估.</p> |

(i) 当将该类变异归类为致病性时, 需确认无功能变异(null variants)是已知的致病机理, 且与该疾病的遗传模式相一致.

(ii) 当文献中将3' 远端下游截短变异注释成致病突变时, 要特别小心. 特别是当所预测的终止密码子出现在最后一个外显子, 或者出现在倒数第二个外显子的最后50个碱基对时, 这种无义突变介导的转录降解^[22]可能不会发生, 这个蛋白很可能会表达.

(iii) 剪接位点变异有可能会引起外显子跳跃、缩短、或者包含内含子. 虽然这一类变异是被预测为无效变异 (null variants) 的, 但是其效应的确认需要通过RNA或蛋白功能分析来确定. 还应考虑阅读框内缺失和插入的可能性, 其有可能重新获得蛋白的关键区域, 仍而导致中性.

(iv) 考虑可替代基因转录本的存在, 理解其具有什么样的生物相关性以及蛋白在什么组织表达, 这些都是非常重要的. 如果一个截短变异只是在一个而不是所有的转录本, 应该防止对其过度解释, 因为有可能会存在其他蛋白质亚型.

(v) 如果在一个以前没有发现致病变异的外显子上发现一个无效变异, 即使这个外显子有可能是可变剪接, 对于其是否为致病性变异也应该谨慎.



6.1 致病变异分级标准

表 3 致病变异分级标准

| 致病性证据 | 分类 |
|-------|--|
| 非常强 | <p>PVS1: 当一个疾病的致病机制为功能丧失(LOF)时, 无功能变异(无义突变、移码突变、经典±1 或 2 的剪接突变、起始密码子变异、单个或多个外显子缺失). 注: 1, 该基因的 LOF 是否是导致该疾病的明确致病机制(如 GFAP, MYH7); 2, 3'末端的功能缺失变异需谨慎解读; 3, 需注意外显子选择性缺失是否影响到蛋白质的完整性; 4, 考虑一个基因存在多种转录本的情况.</p> |
| 强 | <p>PS1: 与先前已确定为致病性的变异有相同的氨基酸改变. 例如: 同一密码子, G>C 或 G>T 改变均可导致缬氨酸→亮氨酸的改变. 注意剪切影响的改变.</p> <p>PS2: 患者的新发变异, 且无家族史(经双亲验证). 注: 仅仅确认父母还不够, 还需注意捐卵、代孕、胚胎移植的差错等情况.</p> <p>PS3: 体内、体外功能实验已明确会导致基因功能受损的变异. 注: 功能实验需要验证是有效的, 且具有重复性与稳定性.</p> <p>PS4: 变异出现在患病群体中的频率显著高于对照群体. 注: 1, 可选择使用相对风险值或者 OR 值来评估, 建议位点 OR 大于 5.0 且置信区间不包括 1.0 的可列入此项(详细见指南正文); 2, 极罕见的变异在病例对照研究可能无统计学意义, 原先在多个具有相同表型的患者中观察到该变异且在对照中未观察到可作为中等水平证据.</p> |
| 中等 | <p>PM1: 位于热点突变区域, 和/或位于已知无良性变异的关键功能域(如酶的活性位点).</p> <p>PM2: ESP 数据库、千人数据库、EXAC 数据库中正常对照人群中未发现的变异(或隐性遗传病中极低频位点)(表 7). 注: 高通量测序得到的插入/缺失人群数据质量较差.</p> <p>PM3: 在隐性遗传病中, 在反式位置上检测到致病变异. 注: 这种情况必须通过患者父母或后代验证.</p> <p>PM4: 非重复区框内插入/缺失或终止密码子丧失导致的蛋白质长度变化.</p> <p>PM5: 新的错义突变导致氨基酸变化, 此变异之前未曾报道, 但是在同一位点, 导致另外一种氨基酸的变异已经确认是致病性的, 如: 现在观察到的是 Arg156Cys, 而 Arg156His 是已知致病的. 注意剪切影响的改变.</p> <p>PM6: 未经父母样本验证的新发变异.</p> |
| 支持证据 | <p>PP1: 突变与疾病在家系中共分离(在家系多个患者中检测到此变异). 注: 如有更多的证据, 可作为更强的证据.</p> <p>PP2: 对某个基因来说, 如果这个基因的错义变异是造成某种疾病的原因, 并且这个基因中良性变异所占的比例很小, 在这样的基因中所发现的新的错义变异.</p> <p>PP3: 多种统计方法预测出该变异会对基因或基因产物造成有害的影响, 包括保守性预测、进化预测、剪接位点影响等. 注: 由于做预测时许多生物信息学算法使用相同或非常相似的输入, 每个算法不应该算作一个独立的标准. PP3 在一个任何变异的评估中只能使用一次.</p> <p>PP4: 变异携带者的表型或家族史高度符合某种单基因遗传疾病.</p> <p>PP5: 有可靠信誉来源的报告认为该变异为致病的, 但证据尚不足以支持进行实验室独立评估.</p> |

比值比(OR)或相对风险用于衡量基因型(即存在于基因组中的变异)和表型(即所患疾病/结果)之间的关联, 适用于任何孟德尔疾病或复杂疾病. 本指南只涉及其在孟德尔疾病中的使用. 相对风险与OR不同, 但概率较小时相对风险近似等于OR. OR 值为1.0 意味着该变异与疾病风险不相关, 大于1.0 意味着变异与疾病风险正相关, 小于1.0 意味着变异与疾病风险负相关.

一般情况下, 具有孟德尔中等效应的变异, 其OR 值为3 或者更大, 高度外显的变异具有非常高的OR 值.

OR值的置信区间(confidence interval, CI)也是一个重要的衡量工具. 如果CI 中包括1.0(如OR=2.5, CI=0.9~7.4), 则关联的可信度很小.

在线可获得简单的OR 值计算器
(<http://www.hutchon.net/ConfidOR.htm/>; <http://easycalculateon.com/statistics/odds-ratio.php/>).



6.1 致病变异分级标准

表 9 在自然人群中评价变异频率

| 疾病 | 基因 | 遗传方式 | 人群 | 发病率 | 携带频率 | 常见突变 | 变异分类 | ESP6500 AA MAF | ESP6500 EA MAF | ESP6500 All MAF | 一致性 | 支持分类的标准 |
|-----------------------|-----------------|------|------|--------|---------|---------------------------|----------------------------------|-------------------|-------------------|--------------------|------|---|
| 囊胞性纤维症 | CFTR | AR | 高加索人 | 0.031% | 3.6% | p.F508del | Ex24:p.F508del (致病) | N/A | N/A | N/A | N/A | 注意不在 EVS 的变异 ^[57] |
| | | | | | | | Ex11:c.1523T>G / p.F508C (良性) | 0.070% | 0.150% | 0.120% | 不一致 | Ref. ^[58] |
| | | | | | | | Ex23:c.3870A>G / p.(=) (良性) | 15.090% | 2.970% | 7.070% | 部分一致 | AA MAF |
| | | | | | | | 5' UTR:c.-8G>C (良性) | 1.160% | 5.550% | 4.060% | 部分一致 | EA MAF |
| | | | | | | | IVS6:c.743+40A>G (良性) | 0.700% | 5.190% | 3.670% | 部分一致 | EA MAF |
| 苯丙酮尿症 | PAH | AR | 北欧人 | 0.010% | 2.0% | | Ex12:c.1242C>T / p.(=) (良性) | 0.360% | 1.310% | 0.990% | 不一致 | PAH 数据库 (http://www.pahdb.mcgill.ca/) |
| | | | | | | | Ex12:c.1278T>C / p.(=) (良性) | 13.550% | 0.090% | 4.650% | 部分一致 | AA MAF |
| | | | | | | | IVS12:c.1316-35C>T (良性) | 0.320% | 2.630% | 1.850% | 部分一致 | EA MAF |
| | | | | | | | Ex9:c.963C>T / p.(=) (良性) | 5.170% | 0.000% | 1.750% | 部分一致 | AA MAF |
| | | | | | | | Ex7:c.489T>G / p.(=) (良性) | 7.010% | 0.050% | 2.410% | 部分一致 | AA MAF |
| MCADD (中链酰基辅酶A脱氢酶缺乏症) | ACADM | AR | 无特异性 | 0.006% | 1.5% | p.K329E aka p.K304E | | | | | | |
| ARPKD (常染色体隐性多囊肾病) | PKHD1 | AR | 无特异性 | 0.005% | 1.4% | | IVS20:c.1964+17G>T (良性) | 0.200% | 0.810% | 0.610% | 不一致 | AA MAF |
| | | | | | | | Ex61:c.10515C>A / p.S3505R (良性) | 0.230% | 1.130% | 0.820% | 不一致 | EA MAF |
| | | | | | | | Ex66:c.11738G>A / p.R3913H (良性) | 1.270% | 0.000% | 0.430% | 不一致 | AA MAF |
| | | | | | | | Ex17:c.1587T>C / p.(=) (良性) | 1.380% | 6.860% | 5.010% | 部分一致 | EA MAF |
| | | | | | | | Ex65:c.11525G>T / p.R3842L (良性) | 0.360% | 2.430% | 1.730% | 部分一致 | EA MAF |
| 雷特氏综合征 | MECP2 | X 连锁 | 无特异性 | 0.012% | De novo | | Ex61:c.10585G>C / p.E3529Q (良性) | 3.950% | 0.010% | 1.350% | 部分一致 | AA MAF |
| | | | | | | | Ex4:c.1161C>T / p.(=) (良性) | 0.030% | 0.000% | 0.010% | 部分一致 | AA MAF |
| | | | | | | | Ex4:c.608C>T / p.T203M (良性) | 0.000% | 0.060% | 0.040% | 部分一致 | http://www.ncbi.nlm.nih.gov/clinvar/variation/95198/ |
| | | | | | | | Ex4:c.683C>G / p.T228S (致病) | 0.830% | 0.000% | 0.300% | 部分一致 | RETT 数据库 |
| | | | | | | | IVS31:c.8047-15C>T (良性) | 0.000% | 0.020% | 0.020% | 部分一致 | EA MAF |
| 歌舞伎综合征 | KMT2D (MLL2) | AD | 无特异性 | 0.003% | De novo | | Ex31:c.6836G>A / p.Gly2279E (良性) | 0.000% | 0.120% | 0.080% | 部分一致 | EA MAF |
| | | | | | | | Ex2:c.309G>A / p.(=) (良性) | 1.460% | 0.000% | 0.490% | 部分一致 | AA MAF |
| | | | | | | | Ex31:c.6478G>A / p.A2160T (良性) | 1.250% | 0.000% | 0.390% | 部分一致 | AA MAF |
| | | | | | | | Ex2:c.856A>G / p.R286Gly (良性) | 0.780% | 0.000% | 0.250% | 部分一致 | AA MAF |
| | | | | | | | Ex2:c.35delG (致病) | 0.090% | 1.080% | 0.740% | 不一致 | ref. ^[59] |
| CHARGE 联合畸形 | CHD7 | AD | 无特异性 | 0.010% | De novo | | | | | | | |
| GJB2 有关的听力丧失 | GJB2 | AR | 无特异性 | 0.067% | 2.5% | c.35delG | | | | | | |
| 血色病 | HFE | AR | 所有人 | 0.040% | 8.3% | p.C282Y | Ex4:c.845G>A / p.C282Y (其他报道) | 1.520% | 6.410% | 4.750% | 不一致 | http://www.ncbi.nlm.nih.gov/sites/GeneTests/review/gene/HFE |

本表列出一些已知发病率的常见遗传疾病中的基因变异。本表罗列出在 Exome Sequencing Project-GO 数据库中找到的变异，对这些变异依据 African American (AA) and European American (EA) 亚群中等位基因变异频率进行分类。如果变异的等位基因频率高于在 AA 或 EA 人群中疾病频率，且在此人群中并无患病者，则此变异可以被认为是 BSI。显性疾病的致病变异的等位基因频率低于自然人群中的疾病发病率，并且隐性疾病的致病变异所具有的异质性频率应该与疾病发病率一致，因此，需要其他数据重新对变异进行分类。在一致性一栏中，部分一致表示亚群频率与上述不一致或未达成一致。也可以被认为具有部分一致性与亚群频率与分类不一致。



6.2 良性变异分级标准

表 4 良性变异分类标准

| 良性影响的证据 | | 分类 |
|---------|--|---|
| 独立证据 | | BA1: ESP 数据库、千人数据库、EXAC 数据库中等位基因频率>5%的变异。 |
| 强 | | BS1: 等位基因频率大于疾病发病率。 BS2: 对于早期完全外显的疾病, 在健康成年人中发现该变异(隐性遗传病发现纯合、显性遗传病发现杂合, 或者 X 连锁半合子)。 BS3: 在体内外实验中确认对蛋白质功能和剪接没有影响的变异。 BS4: 在一个家系成员中缺乏共分离。 注: 这部分需要考虑复杂疾病和外显率问题。 |
| 支持证据 | | BP1: 已知一个疾病的致病原因是由于某基因的截短变异, 在此基因中所发现的错义变异。 BP2: 在显性遗传病中又发现了另一条染色体上同一基因的一个已知致病变异, 或者是任意遗传模式遗传病中又发现了同一条染色体上同一基因的一个已知致病变异。 BP3: 功能未知重复区域内的缺失/插入, 同时没有导致基因编码框改变。 BP4: 多种统计方法预测出该变异会对基因或基因产物无影响, 包括保守性预测、进化预测、剪接位点影响等。 注: 由于做预测时许多生物信息算法使用相同或非常相似的输入, 每个算法不应该算作一个独立的标准。BP4 在任何一个变异的评估中只能使用一次。 BP5: 在已经有另一分子致病原因的病例中发现的变异。 BP6: 有可靠信誉来源的报告认为该变异为良性的, 但证据尚不足以支持进行实验室独立评估。 BP7: 同义变异且预测不影响剪接。 |



6.3 证据框架

表 6 程度判断评价标准

| | ← 良性 → | | ← 致病 → | | | |
|--------|----------------------------|--|--|--|-----------------------|--------------------------|
| | 强 | 支持 | 支持 | 中等 | 强 | 很强 |
| 人群数据 | 疾病 MAF 太高 或者对照组与疾病外显不一致 | | | 人群数据库中缺失 PM2 | 患者中频率显著高于对照 PS4 | |
| 计算预测数据 | | 多个计算证据表明基因/基因产物无作用 BP4 仅截短变异致病基因的错义变异 BP1 预测无剪切作用的沉默变异 BP7 重复未知功能区域的框内插入/缺失 BP3 | 多个计算证据支持基因/基因产物的有害影响 PP3 | 新的错义改变位于先前已确定为致病性的氨基酸残基 PM5 蛋白长度改变的变异 PM4 | 与已鉴定的致病变异有相同氨基酸改变 PS1 | 在已知致病机制为 LOF 的基因中预测为无效变异 |
| 功能数据 | 完善的功能研究表明无有害影响 BS3 | | 致病错义变异常见、良性错义变异罕见的基因上的错义变异 PP2 多个家系患者中共分离 PP1 | 热点突变或深入研究的无良性变异的功能域 PM1 | 完善的功能研究表明有害作用 PS3 | |
| 共分离数据 | 与疾病不共分离 BS4 | | | 逐渐增加的共分离数据 | | |
| 新发数据 | | | | 未经双亲验证的新发变异 PM6 | 经双亲验证的新发变异 PS2 | |
| 等位基因数据 | | 在反式观察到的显性变异 BP2 在顺式观察到的致病变异 BP2 较好信誉的未共享数据 = 良性 BP6 | | 隐性疾病中，反式检测到致病变异 PM3 | | |
| 其他数据库 | | | 较好信誉的来源 = 致病的 PP5 | | | |
| 其他数据 | | 在可替代病因的患者中发现 BP5 | 患者表型或家族史具有高度基因特异性 PP4 | | | |



7. 注意问题

■ 对临床意义不明的基因(GUS)中的变异评价?

基因组和外显子组测序正在不断鉴定出新的基因型-表型关联. 当实验室发现某个基因的变异, 但尚未证实此基因与病人的表型有关联, 该变异称为**GUS 变异**. 情况可能是以前没有研究过这个基因与表型之间的关系, 或者多个表型都有关系. Now, 我们无法用本指南再对变异进行分类. 假如我们观察到一个 de novo 变异, 但是我们不能认为它是致病的, 因为对于任何一个体而言, 我们都可以发现1个甚至100个这样的变异. 现在, 我们可以通过 LOD 值 (logarithm of the) 找出上千个这样基因.

一些有害变异确实能破坏一个基因或者其编码的蛋白, 如无义变异、移码变异, 经典的 $\pm 1, 2$ 剪接, 外显子缺失等, 即使如此也不能认为其是致病变异, 仍然需要一些其他证据才能证明其与表型之间的关系。

■ 健康个体或偶然发现的变异的评价?

当评估在健康个体中或无症状个体中或者偶然发现与症状无关的变异时, 必须谨慎使用此指南. 在这些情况下, 发现的致病变异的概率可能远低于疾病靶向性检测发现的可能性. 因此, 当判定这些变异为致病变异时, 需要更强的证据支持. 此外, 和基于确诊患者预测的外显率相比, 在无相关表型或家族史个体中发现的致病变异的预测外显率可能要低很多.

■ 线粒体变异的评价?

常见线粒体数据库列表

| 数据库 | 网址 |
|--------------------------------|---|
| MitoMap:单倍型 | http://mitomap.org/bin/view.pl/MITOMAP/WebHome |
| mtDB:频率信息 | http://www.mtodb.igp.uu.se/ |
| Mamit-tRNA二级结构、序列和线粒体转运RNA 的比对 | http://mamittrna.ustrasbg.fr/ |
| Phylotree:线粒体单倍群 | http://www.phylotree.org |
| mtDNA community: 其他信息 | http://www.mtDNAcommunity.org/default.aspx |



Table 1 CNV interpretation scoring metric: copy-number loss

| Section 1: Initial assessment of genomic content | | | |
|---|---|----------------------------|-----------|
| Evidence type | Evidence | Suggested points/case | Max score |
| Copy-number loss content | 1A. Contains protein-coding or other known functionally important elements. | 0 (Continue evaluation) | 0 |
| | 1B. Does NOT contain protein-coding or any known functionally important elements. | -0.60 | -0.60 |
| Section 2: Overlap with established/predicted haploinsufficiency (HI) or established benign genes/genomic regions (<i>Skip to section 3 if your copy-number loss DOES NOT overlap these types of genes/regions</i>) | | | |
| Overlap with ESTABLISHED HI genes or genomic regions and consideration of reason for referral | 2A. Complete overlap of an established HI gene/genomic region. | 1.00 | 1.00 |
| | 2B. Partial overlap of an established HI genomic region • The observed CNV does NOT contain the known causative gene or critical region for this established HI genomic region OR • Unclear if known causative gene or critical region is affected OR • No specific causative gene or critical region has been established for this HI genomic region | 0 (Continue evaluation) | 0 |
| | 2C. Partial overlap with the 5' end of an established HI gene (3' end of the gene not involved)... | See categories below | |
| | 2C-1. ...and coding sequence is involved | 0.90 (range: 0.45 to 1.00) | 1.00 |
| | 2C-2. ...and only the 5' UTR is involved | 0 (range: 0 to 0.45) | 0.45 |
| | 2D. Partial overlap with the 3' end of an established HI gene (5' end of the gene not involved)... | See categories below | |
| | 2D-1. ...and only the 3' untranslated region is involved. | 0 (Continue evaluation) | 0 |
| | 2D-2. ...and only the last exon is involved. Other established pathogenic variants have been reported in this exon. | 0.90 (range: 0.45 to 0.90) | 0.90 |
| | 2D-3. ...and only the last exon is involved. No other established pathogenic variants have been reported in this exon. | 0.30 (range: 0 to 0.45) | 0.45 |
| | 2D-4. ...and it includes other exons in addition to the last exon. Nonsense-mediated decay is expected to occur. | 0.90 (range: 0.45 to 1.00) | 1.00 |



ACMG+ClinGen→CNV变异评级指南

| | | | |
|---|---|--|------------------------|
| | 2E. Both breakpoints are within the same gene (intragenic CNV; gene-level sequence variant). | See ClinGen SVI working group PVS1 specifications • PVS1 = 0.90 (Range: 0.45 to 0.90) • PVS1_Strong = 0.45 (Range: 0.30 to 0.90) • PVS1_Moderate or PM4 (in-frame indels) = 0.30 (Range: 0.15 to 0.45) • PVS1_Supporting = 0.15 (Range: 0 to 0.30) • N/A = No points, but continue evaluation | See categories at left |
| Overlap with ESTABLISHED benign genes or genomic regions | 2F. Completely contained within an established benign CNV region. | -1 | -1 |
| | 2G. Overlaps an established benign CNV, but includes additional genomic material. | 0 (Continue evaluation) | 0 |
| Haploinsufficiency predictors | 2H. Two or more HI predictors suggest that AT LEAST ONE gene in the interval is HI. | 0.15 | 0.15 |
| Section 3: Evaluation of gene number | | | |
| Number of protein-coding RefSeq genes wholly or partially included in the copy-number loss | 3A. 0–24 genes | 0 | 0 |
| | 3B. 25–34 genes | 0.45 | 0.45 |
| | 3C. 35+ genes | 0.90 | 0.90 |
| Section 4: Detailed evaluation of genomic content using cases from published literature, public databases, and/or internal lab data (Skip to section 5 if either your CNV overlapped with an established HI gene/region in section 2, OR there have been no reports associating either the CNV or any genes within the CNV with human phenotypes caused by loss of function [LOF] or copy-number loss) | | | |
| Individual case evidence—de novo occurrences | Reported proband (from literature, public databases, or internal lab data) has either: • A complete deletion of or a LOF variant within gene encompassed by the observed copy-number loss OR • An overlapping copy-number loss similar in genomic content to the observed copy-number loss AND... | See categories below | |
| | 4A. ...the reported phenotype is highly specific and relatively unique to the gene or genomic region, | Confirmed de novo: 0.45 points each Assumed de novo: 0.30 points each (range: 0.15 to 0.45) | 0.90 (total) |



ACMG+ClinGen→CNV变异评级指南

| | | | |
|---|---|---|---------------|
| | 4B. ...the reported phenotype is consistent with the gene/genomic region, is highly specific, but not necessarily unique to the gene/genomic region. | Confirmed de novo: 0.30 points each Assumed de novo: 0.15 point each (range: 0 to 0.45) | |
| | 4C. ...the reported phenotype is consistent with the gene/genomic region, but not highly specific and/or with high genetic heterogeneity. | Confirmed de novo: 0.15 point each Assumed de novo: 0.10 point each (range: 0 to 0.30) | |
| Individual case evidence—inconsistent phenotype | 4D. ...the reported phenotype is NOT consistent with what is expected for the gene/genomic region or not consistent in general. | 0 points each (range: 0 to -0.30) | -0.30 (total) |
| Individual case evidence—unknown inheritance | 4E. Reported proband has a highly specific phenotype consistent with the gene/genomic region, but the inheritance of the variant is unknown. | 0.10 points each (range: 0 to 0.15) | 0.30 (total) |
| Individual case evidence—segregation among similarly affected family members | 4F. 3–4 observed segregations | 0.15 | 0.45 |
| | 4G. 5–6 observed segregations | 0.30 | |
| | 4H. 7 or more observed segregations | 0.45 | |
| Individual case evidence—nonsegregations | 4I. Variant is NOT found in another individual in the proband's family AFFECTED with a consistent, specific, well-defined phenotype (no known phenocopies). | -0.45 points per family (range: 0 to -0.45) | -0.90 (total) |
| | 4J. Variant IS found in another individual in the proband's family UNAFFECTED with the specific, well-defined phenotype observed in the proband. | -0.30 points per family (range: 0 to -0.30) | -0.90 (total) |
| | 4K. Variant IS found in another individual in the proband's family UNAFFECTED with the nonspecific phenotype observed in the proband. | -0.15 points per family (range: 0 to -0.15) | -0.30 (total) |
| Case-control and population evidence | 4L. Statistically significant increase amongst observations in cases (with a consistent, specific, well-defined phenotype) compared with controls. | 0.45 per study (range: 0 to 0.45 per study) | 0.45 (total) |
| | 4M. Statistically significant increase amongst observations in cases (without a consistent, nonspecific phenotype OR unknown phenotype) compared with controls. | 0.30 per study (range: 0 to 0.30 per study) | 0.45 (total) |
| | 4N. No statistically significant difference between observations in cases and controls. | -0.90 (per study) (range: 0 to -0.90 per study) | -0.90 (total) |
| | 4O. Overlap with common population variation. | -1 (range: 0 to -1) | -1 |
| Section 5: Evaluation of inheritance pattern/family history for patient being studied | | | |
| Observed copy-number loss is de novo | 5A. Use appropriate category from de novo scoring section in section 4. | Use de novo scoring categories from section 4 (4A–4D) to determine score | 0.45 |
| Observed copy-number loss is inherited | 5B. Patient with specific, well-defined phenotype and no family history. CNV is inherited from an apparently unaffected parent. | -0.30 (range: 0 to -0.45) | -0.45 |
| | 5C. Patient with nonspecific phenotype and no family history. CNV is inherited from an apparently unaffected parent. | -0.15 (range: 0 to -0.30) | -0.30 |
| | 5D. CNV segregates with a consistent phenotype observed in the patient's family. | Use segregation scoring categories from section 4 (4F–4H) to determine score | 0.45 |
| Observed copy-number loss—nonsegregations | 5E. Use appropriate category from nonsegregation section in section 4. | Use nonsegregation scoring categories from section 4 (4I–4K) to determine score | -0.45 |
| Other | 5F. Inheritance information is unavailable or uninformative. | 0 | 0 |
| | 5G. Inheritance information is unavailable or uninformative. The patient phenotype is nonspecific, but is consistent with what has been described in similar cases. | 0.10 (range: 0 to 0.15) | 0.15 |
| | 5H. Inheritance information is unavailable or uninformative. The patient phenotype is highly specific and consistent with what has been described in similar cases. | 0.30 (range: 0 to 0.30) | 0.30 |

ved.



ACMG+ClinGen→CNV变异评级指南

Table 2 CNV interpretation scoring metric: copy-number gain

| Section 1: Initial assessment of genomic content | | | |
|--|---|---|-----------|
| Evidence type | Evidence | Suggested points/case | Max score |
| Copy-number gain content | 1A. Contains protein-coding or other known functionally important elements. | 0 (Continue evaluation) | 0 |
| | 1B. Does NOT contain protein-coding or any known functionally important elements. | -0.60 | -0.60 |
| Section 2: Overlap with established triplosensitive (TS), haploinsufficient (HI), or benign genes or genomic regions (Skip to section 3 if the copy-number gain DOES NOT overlap these types of genes/regions) | | | |
| Overlap with ESTABLISHED TS genes or genomic regions | 2A. Complete overlap; the TS gene or minimal critical region is fully contained within the observed copy-number gain. | 1 | 1 |
| | 2B. Partial overlap of an established TS region • The observed CNV does NOT contain the known causative gene or critical region for this established TS genomic region OR • Unclear if the known causative gene or critical region is affected OR • No specific causative gene or critical region has been established for this TS genomic region. | 0 (Continue evaluation) | 0 |
| Overlap with ESTABLISHED benign copy-number gain genes or genomic regions | 2C. Identical in gene content to the established benign copy-number gain. | -1 | -1 |
| | 2D. Smaller than established benign copy-number gain, breakpoint(s) does not interrupt protein-coding genes. | -1 | -1 |
| | 2E. Smaller than established benign copy-number gain, breakpoint(s) potentially interrupts protein-coding gene. | 0 (Continue evaluation) | 0 |
| | 2F. Larger than known benign copy-number gain, does not include additional protein-coding genes. | -1 (range: 0 to -1.00) | -1 |
| | 2G. Overlaps a benign copy-number gain but includes additional genomic material. | 0 (Continue evaluation) | 0 |
| | 2H. HI gene fully contained within observed copy-number gain. | 0 (Continue evaluation) | 0 |
| Breakpoint(s) within ESTABLISHED HI genes | 2I. Both breakpoints are within the same gene (gene-level sequence variant, possibly resulting in loss of function [LOF]). | See ClinGen SVI working group PVS1 specifications • PVS1 = 0.90 (Range: 0.45 to 0.90) • PVS1_Strong = 0.45 (Range: 0.30 to 0.90) • N/A = 0 (Continue evaluation) | |
| | 2J. One breakpoint is within an established HI gene, patient's phenotype is either inconsistent with what is expected for LOF of that gene OR unknown. | 0 (Continue evaluation) | 0 |
| | 2K. One breakpoint is within an established HI gene, patient's phenotype is highly specific and consistent with what is expected for LOF of that gene. | 0.45 | 0.45 |
| Breakpoints within other gene(s) | 2L. One or both breakpoints are within gene(s) of no established clinical significance. | 0 (Continue evaluation) | 0 |



ACMG+ClinGen→CNV变异评级指南

Section 4: Detailed evaluation of genomic content using cases from published literature, public databases, and/or internal lab data (Note: If there have been no reports associating either the copy-number gain or any of the genes therein with human phenotypes caused by triplosensitivity, skip to section 5)

| | | | |
|--|---|--|--------------|
| Individual case evidence—de novo occurrences | Reported proband (from literature, public databases, or internal lab data) has either: • complete duplication of one or more genes within the observed copy-number gain OR • an overlapping copy-number gain similar in genomic content to the observed copy-number gain AND... | See categories below | |
| | 4A. ...the reported phenotype is highly specific and relatively unique to the gene or genomic region. | Confirmed de novo: 0.45 points each Assumed de novo: 0.30 points each (range: 0 to 0.45) | 0.90 (total) |

Section 5: Evaluation of inheritance patterns/family history for patient being studied

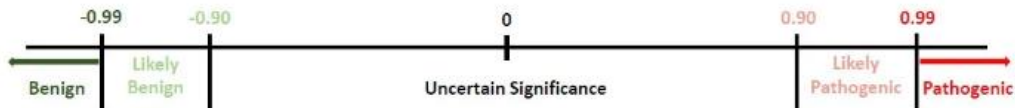
| | | | | | |
|--|---|---|--|---|---------------|
| | 4B. ...the reported phenotype is highly specific and relatively unique to the gene or genomic region. | Observed copy-number gain is de novo | 5A. Use appropriate category from de novo scoring section in section 4. | Use de novo scoring categories from section 4 (4A–4D) to determine score | 0.45 |
| | 4C. ...the reported phenotype is not highly specific and relatively unique to the gene or genomic region. | Observed copy-number gain is inherited | 5B. Patient with a specific, well-defined phenotype and no family history. Copy-number gain is inherited from an apparently unaffected parent. | –0.30 (range: 0 to –0.45) | –0.45 |
| Individual case evidence—inconsistent phenotype | 4D. ...the reported phenotype is not consistent with the family history. | | 5C. Patient with nonspecific phenotype and no family history. Copy-number gain is inherited from an apparently unaffected parent. | –0.15 (range: 0 to –0.30) | –0.30 |
| Individual case evidence—unknown inheritance | 4E. Reports of gene/genomic region in family history. | | 5D. CNV segregates with consistent phenotype observed in the patient's family. | Use segregation scoring categories from in section 4 (4F–4H) to determine score | 0.45 |
| Individual case evidence—segregation among similarly affected family members | 4F. 3–4 observations of similar phenotype in family. | | | | |
| | 4G. 5–6 observations of similar phenotype in family. | Observed copy-number gain—nonsegregations | 5E. Use appropriate category from nonsegregation section in section 4. | Use nonsegregation scoring categories from section 4 (4I–4K) to determine score | –0.45 |
| | 4H. 7 or more observations of similar phenotype in family. | | 5F. Inheritance information is unavailable or uninformative. | 0 | 0 |
| Individual case evidence—nonsegregations | 4I. Variant observed in family with the same phenotype. | | 5G. Inheritance information is unavailable or uninformative. The patient phenotype is nonspecific, but is consistent with what has been described in similar cases. | 0.10 (range: 0 to 0.15) | 0.15 |
| | 4J. Variant with the same phenotype observed in family. | | 5H. Inheritance information is unavailable or uninformative. The patient phenotype is highly specific and consistent with what has been described in similar cases. | 0.15 (range: 0 to 0.30) | 0.30 |
| | 4K. Variant with the same phenotype observed in family. | | | | |
| Case-control and population evidence | 4L. Statistically significant increase among observations in cases (with a consistent, specific, well-defined phenotype) compared with controls. | | | | |
| | 4M. Statistically significant increase among observations in cases (with a consistent, nonspecific phenotype or unknown phenotype) compared with controls. | | 0.30 per study (range: 0 to 0.30 per study) | | 0.45 (total) |
| | 4N. No statistically significant difference between observations in cases and controls. | | –0.90 per study (range: 0 to –0.90 per study) | | –0.90 (total) |
| | 4O. Overlap with common population variation. | | –1 (range: 0 to –1) | | –1 |



ACMG+ClinGen→CNV变异评级指南

| 拷贝数缺失证据 | | 拷贝数扩增证据 | |
|---------|-------|---------|-------|
| 1 | 1A/1B | 1 | 1A/1B |
| 2 | 2A-2H | 2 | 2A-2L |
| 3 | 3A-3C | 3 | 3A-3C |
| 4 | 4A-4O | 4 | 4A-4O |
| 5 | 5A-5H | 5 | 5A-5H |

每个步骤根据相应的评分标准对当前待评估的CNV的相关证据进行评分。支持致病性证据计正分数，反驳致病性证据（良性证据）给负分数；证据强度越高，相应分值也越大。各个步骤的评分累加计算出总评分，根据总评分给出变异评级，评分与变异评级的对应关系如下。



| 分数 | 评级 |
|----------------|---------------|
| ≥ 0.99 分 | 致病变异 (P) |
| 0.98~0.90分 | 可能致病变异 (LP) |
| 0.89~-0.89分 | 不确定意义变异 (VUS) |
| -0.90~-0.98分 | 可能良性 (LB) |
| ≤ -0.99 分 | 良性 (B) |

CNV loss评分准则一共包含5个部分（Section1 - Seciton5），评分流程从Section 1到Section 5一共5个步骤依次进行，如下：

Section 1: 基因组内容初始评分

Section 2: 与已知单倍剂量不足CNV/基因重叠评分，与良性

CNV重叠评分

Section 3: CNV区间内编码基因数量评分

Section 4: CNV区间内基因组内容详细评分

Section 5: 当前病例遗传模式/家族史评分

CNV Gain评分准则一共包含5个部分（Section1 - Seciton5），评分流程从Section 1到Section 5一共5个步骤依次进行，如下：

Section 1: 基因组内容初始评分

Section 2: 与已知三倍剂量敏感CNV/基因重叠评分，与良性

CNV重叠评分，与已知单倍剂量不足基因重叠评分

Section 3: CNV区间内编码基因数量评分

Section 4: CNV区间内基因组内容详细评分

Section 5: 当前病例遗传模式/家族史评分



- *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.*
- 遗传变异分类标准与指南
- *Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen).*

THANK YOU
感谢观看