**遗传变异分类标准与指南**

表1 数据库（人群；疾病；序列）

|  |  |
| --- | --- |
| **人群数据库** | **数据库简介** |
| Exome Aggregation Consortium<http://exac.broadinstitute.org/> | 61486个独立个体外显子组测序发现的变异数据库，作为多种疾病特异性和人群基因组研究的一部分。儿童疾病患者和相关个体未纳入。 |
| Exome Variant Server <http://evs.gs.washington.edu/EVS>  | 几个包括欧洲和非裔美国人的大型队列的基因组测序过程中发现的变异数据库。包括覆盖数据来报告变异缺失。 |
| 1000 Genomes Project <http://browser.1000genomes.org> | 来自26个人群低覆盖和高覆盖的基因组和目标与测序过程中发现的变异数据库。与外显子组汇聚联盟 (Exome Aggregation Consortium) 相比更具多样性，但也包含低质量的数据，且一些队列中包含了相关的个体。 |
| dbSNP <http://www.ncbi.nlm.nih.gov/snp>  | 多来源的短的遗传变异数据库 (通常≤50个碱基) 数据库，可能缺乏原始研究数据且可能包含致病变异。 |
| dbVar <http://www.ncbi.nlm.nih.gov/dbvar> | 多来源的结构变异（通常>50个碱基）数据库 |
| **疾病数据库** | **数据库简介** |
| ClinVar<http://www.ncbi.nlm.nih.gov/clinvar> | 人类变异临床意义和表型关系认定的数据库。 |
| OMIM <http://www.omim.org>  | 人类基因和遗传条件（包含疾病相关的遗传变异的典型样本）的数据库。 |
| Human Gene Mutation Database<http://www.hgmd.org> Locus/disease/ethnic/other-specific databases | 文献中公开的变异注释数据库，需付费订阅来访问大部分内容。 |
| Human Genome Variation Society <http://www.hgvs.org/dblist/dblist.html>Leiden Open Variation Database<http://www.lovd.nl> | 人类基因组变异协会网站开发的数以千计的一系列对人类变异特定子集提供变异注释的数据库。数据库很大一部分是建立在莱顿开放变异数据库系统。 |
| DECIPHER http://decipher.sanger.ac.uk  | 一个分子细胞遗传学数据库，临床医生和研究人员通过Ensembl基因组浏览器链接具有表型的基因芯片数据。 |
| **序列数据库** | **数据库简介** |
| NCBI Genome <http://www.ncbi.nlm.nih.gov/genome> | 全人类基因组参考序列来源。 |
| RefSeqGene <http://www.ncbi.nlm.nih.gov/refseq/rsg>Locus Reference Genomic (LRG)<http://www.lrg-sequence.org> | 医学相关的基因参考序列资源。 |
| MitoMap <http://www.mitomap.org/MITOMAP/HumanMitoSeq> | 人类线粒体DNA剑桥参考序列修订版。 |

表2 生物信息预测算法

|  |  |  |  |
| --- | --- | --- | --- |
| **分类** | **名称** | **网站** | **依据** |
| 错义预测 | Consurf | http://consurftest.tau.ac.il | 进化保守性 |
| FATHMM | http://fathmm.biocompute.org.uk | 进化保守性 |
| MutationAssessor | 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.htmlMAPP/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 | 比对和测量变异序列与同源蛋白序列的相似度 |
| nsSNPAnalyzer | http://snpanalyzer.uthsc.edu | 多重序列比对和蛋白结构分析 |
| Condel | http://bg.upf.edu/fannsdb/ | 联合SIFT、PolyPhen-2和MutationAssessor进行综合预测 |
| CADD | http://cadd.gs.washington.edu | 对比人类模拟变异中固定/近固定衍生等位基因的注释 |
| 剪切位点预测 | GeneSplicer | http://www.cbcb.umd.edu/software/GeneSplicer/ gene\_spl.shtmlgene\_spl.shtml | 马尔可夫模型 |
| Human Splicing Finder | http://www.umd.be/HSF/ | 位置相关的逻辑 |
| MaxEntScan | http://genes.mit.edu/burgelab/maxent/Xmaxentscan\_ scoreseq.html scoreseq.html | 最大熵原理 |
| 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 |
| FSPLICE | http://www.softberry.com/berry.phtml?topic=fsplice&group=programs&subgroup=gfind | 基于权重矩阵模型对物种特异性剪切位点进行预测 |
| 核酸保守性预测 | GERP | http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html | 基因组进化速率分析 |
| PhastCons | http://compgen.bscb.cornell.edu/phast/ | 保守评分和保守因素鉴定 |
| PhyloP | http://compgen.bscb.cornell.edu/phast/ |
|  | http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt | 比对和进化树：对特定谱系或所有分支保守或加速进行P值计算。 |

**表3 致病变异分级标准**

|  |  |
| --- | --- |
| **致病性****证据** | **分类** |
| **很强** | **PVS1**：LOF是已知疾病致病机制的基因上的无效变异（无义、框移、典型的±1、2个碱基的剪切位点、起始密码子、一个或多个外显子缺失）**注意事项**：* 注意LOF不是已知的疾病致病机制的基因（如*GFAP*、*MYH7*）
* 谨慎解读基因的3'端末端的LOF的变异
* 谨慎对待被预测导致外显子跳读，但剩余的蛋白质完整的剪切变异
* 注意多个转录本的存在
 |
| **强** | **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：相同密码子中的G>C或G> T改变均可导致缬氨酸→亮氨酸的改变注意事项：相比于氨基酸/蛋白水平的改变，更需注意剪切影响的改变**PS2**：与疾病吻合且无家族史的患者中检出的新生变异（经双亲验证）注：仅父系验证是不够的。捐卵、代孕、胚胎移植的差错等都可能是非产妇遗传的因素。**PS3**：完善的体外或体内功能研究支持基因或基因产物的损伤效应注：已经过验证并被证明在临床诊断实验室环境中是可重复的和健全的功能研究被认为是最完善的。**PS4**：相比于对照组，变异在患者中的频率显著增高注1：病例-对照研究中获得的相对风险或OR> 5.0，且估计的置信区间不包括1.0（详细指南见正文）。注2：极罕见的变异在病例对照研究可能无统计学意义，在多个具有相同表型的患者中优先观察到该变异且在对照中未观察到可作为中等水平证据。 |
| **中等** | **PM1**：位于热点突变和/或关键的完善的无良性变异的功能域（如酶的活性位点）**PM2**：外显子组测序计划、千人基因组计划或外显子组聚合联盟的对照组中频率为0（或隐性遗传中极低的频率）（表6）注意事项：新一代测序中插入/缺失的人口数据通常很差。**PM3**：对于隐性遗传疾病，在反式中检测到致病变异注意：本证据需父母（或子女）的验证来确定阶段。**PM4**：在非重复区域或末端缺失变异中由于框内缺失/插入导致蛋白质长度改变**PM5**：新的错义改变是位于先前已确定为致病性的氨基酸残基的一个不同的错义改变例如：现在观察到的是Arg156Cys，而Arg156His是已知致病的注意事项：注意影响剪切，而不是影响氨基酸/蛋白水平的改变。**PM6**：假定为新生变异，但未经双亲验证 |
| **辅助证据** | **PP1**：一个确定的已知致病基因在家系多个患者中与疾病共分离注：越来越多的共分离数据可作为更强的证据**PP2**：在一个错义变异时常见致病机制的基因中，某基因的错义变异为良性错义变异的概率较低**PP3**：多条计算证据支持基因或基因产物（保守性、进化、剪切影响等）的有害影响注意事项：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。 PP3在一个任何变异的评估中只能使用一次。**PP4**：患者的表型或家族史对具有单一遗传病因的疾病具有高度特异性**PP5**：近期有信誉的来源的报告报导某变异为致病的，但实验室不能获得该证据来实施独立评估 |

LOF，loss of function，功能缺失；OR，odds ration，比值比

表4 良性变异分类标准

|  |  |
| --- | --- |
| 良性影响证据 | 分类 |
| 独立证据 | **BA1**：基因组测序项目、千人基因组计划或外显子组汇聚联盟 等位基因频率>5% |
| 强 | **BS1**：等位基因频率比疾病预期更大（表6）**BS2**：在一个健康的成年个体中观察到隐性（纯合子），显性（杂合子）或X-连锁（半合子）障碍，预计在幼年完全外显**BS3**：完善的体外或体内功能研究表明对蛋白质的功能或剪切没有破坏性影响**BS4**：在一个家系患者中缺乏共分离注意事项：常见的表型（即，癌症、癫痫）中拟表型的存在可以模仿患病个体的共分离。此外，常染色体显性遗传疾病家系中可能有一个以上的致病变异，进一步混淆明显的分离缺失。 |
| 辅助证据 | **BP1**：既往已知截短变异致病的基因中鉴定的错义变异**BP2**：全外显的显性基因/疾病具有致病变异的反式中观察到的变异，或者任何遗传方式基因/疾病具有致病变异的顺式中观察到的变异**BP3**：无已知功能的重复区域内的框内缺失/插入变异**BP4**：多个计算证据表明基因或基因产物是没有影响的（保守性、进化、剪切影响等）注意事项：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。 PP3在一个任何变异的评估中只能使用一次。**BP5**：发现的变异在疾病中具有可替代的分子基础**BP6**：近期正规来源的报告报导某变异为良性的，但实验室不能获得该证据来实施独立评估**BP7**：剪切预测算法预测对剪切一致性序列无影响，且新产生的剪切位点和核苷酸并不高度保守的一个同义（沉默）变异 |

表5 遗传变异分类联合标准规则

|  |  |  |
| --- | --- | --- |
| 1 | 致病 | (i) 1个很强（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） |
| 2 | 可能致病 | (i) 1个很强（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） |
| 3 | 良性 | (i) 1个独立（BA1）或 |
| (ii) ≥2个强（BS1-BS4） |
| 4 | 可能良性 | (i) 1个强（BS1-BS4）和1个支持（BP1-BP7）或 |
| (ii) ≥2个支持（BP1-BP7） |
| 5 | 意义不明 | (i) 不满足上述标准或 |
| (ii) 良性和致病标准相互矛盾 |

表6 评估人群中变异频率来策划变异分类

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 疾病  | 基因 | 遗传方式 | 人群 | 频率（%） | 携带率（%） | 常见变异 | 变异分类 | ESP6500AA MAF | ESP6500EA MAF | ESP6500All MAF | 一致性 | 分类支持标准 |
| 囊性纤维化 | *CFTR* | AR | 高加索 | 0.031 | 3.6 | p.F508del | Ex24:p.F508del (致病) | N/A | N/A | N/A | N/A | 参考文献57 |
| Ex11:c.1523T>G / p.F508C (良性)  | 0.070% | 0.150% | 0.120% | 不一致 | 参考文献58 |
| Ex23:c.3870A>G / p.(=) (良性)  | 15.090% | 2.970% | 7.070% | 部分 | AA MAF |
| 5’UTR:c.-8G>C (良性)  | 1.160% | 5.550% | 4.060% | 部分 | EA MAF |
| IVS6:c.743 + 40A>G (良性) | 0.700% | 5.190% | 3.6705% | 部分 | EA MAF |
| 苯丙酮尿症（PKU） | *PAH* | AR | 南欧 | 0.010 | 2.0 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| MCADD | *ACADM* | AR | 不明确 | 0.006 | 1.5 | p.K329Eaka p.K304E |  |  |  |  |  |  |
| ARPKD | *PKHD1* | AR | 不明确 | 0.005 | 1.4 |  |  |  |  |  |  |  |
| Rett综合征 | *MECP2* | X连锁 | 不明确 | 0.012 | 新生变异 |  |  |  |  |  |  |  |
| 歌舞伎综合征 | *KMT2D (MLL2)* | AD | 不明确 | 0.003 | 新生变异 |  |  |  |  |  |  |  |
| CHARGE综合征 | *CHD7* | AD | 不明确 | 0.010 | 新生变异 |  |  |  |  |  |  |  |
| GJB2相关的听力损失 | *GJB2* | AR | 不明确 | 0.067 | 2.5 | c.35delG |  |  |  |  |  |  |
| 血色素沉着症 | *HFE* | AR | 全部 | 0.040 | 8.3 | p.C282Y |  |  |  |  |  |  |

一些常见遗传疾病的变异中，显性遗传（如歌舞伎综合征和CHARGE综合征）和X连锁疾病（如Rett综合征）有已知的发生率，隐性遗传疾病（如囊性纤维化、苯丙酮尿症（PKU）、中间共同酰基脱氢酶缺乏症（MCADD）、常染色体隐性遗传多囊肾病（ARPKD）、GJB2相关的听力损失和血色素沉着症）有已知的携带率。这些基因中在外显子组测序计划-GO数据库中发现的一些变异列在变异分类中，是基于非裔美国人（African American, AA）和欧裔美国人（European American, EA）亚群等位基因频率的证据和一致性。变异的等位基因频率在至少一个亚群不具有疾病的个体中（AA或EA）高于疾病预期，可以考虑作为良性影响的强证据（BS1）。显性遗传疾病的已知致病变异的杂合子频率英语疾病发生率一致。当等位基因频率低于良性并高于致病变异发生率/携带率时，需要其他数据来对这些变异进行分类。当亚群评论不符合以上所述或无法得到一致性时，一致性一栏中的变异定义为“部分的”。当某一个亚群频率与分类不一致时，变异定义为部分一致。