



遗传变异分类标准与指南

王秋菊, 沈亦平, 邬玲任, 陈少科, 陈子江, 方向东, 傅松滨, 龚瑶琴, 黄国英, 黄国宁, 黄荷凤, 黄山, 郝晓柯, 冀小平, 李红, 梁波, 廖灿, 乔杰, 苏海翔, 魏军, 王磊, 王树玉, 王晓红, 邢清和, 徐湘民, 袁慧军, 杨正林, 周从容, 周文浩, 曾勇, 张学军, 黄涛生, 郑茜, 秦胜营, 于世辉, 关静, 王洪阳, 王大勇, 赵立东, 王慧君, 孔令印, 宣黎明, 冒燕, 祝铁君, 徐君玲, 王剑青, 王莉, 赵婷, 秦一丁, 夏滢颖, 樊丽霞, 赵丁丁, 邱浩 and 贺林

Citation: [中国科学: 生命科学](#) ; doi: 10.1360/N052017-00099

View online: <http://engine.scichina.com/doi/10.1360/N052017-00099>

Published by the [《中国科学》杂志社](#)

Articles you may be interested in

[《科学通报》投稿指南](#)

科学通报 **45**, 559 (2000);

[《科学通报》投稿指南](#)

科学通报 **45**, 2806 (2000);

[《科学通报》投稿指南](#)

科学通报 **45**, 335 (2000);

[《科学通报》投稿指南](#)

科学通报 **45**, 111 (2000);

[《科学通报》投稿指南](#)

科学通报 **49**, 314 (2004);





标准与指南

遗传变异分类标准与指南

王秋菊¹, 沈亦平², 邬玲仟³, 陈少科⁴, 陈子江⁵, 方向东⁶, 傅松滨⁷, 龚瑶琴⁵, 黄国英⁸, 黄国宁⁹, 黄荷凤¹⁰, 黄山¹¹, 郝晓柯¹², 冀小平¹³, 李红¹⁴, 梁波¹⁵, 廖灿¹⁶, 乔杰¹⁷, 苏海翔¹⁸, 魏军¹⁹, 王磊²⁰, 王树玉²¹, 王晓红²², 邢清和²⁰, 徐湘民²³, 袁慧军²⁴, 杨正林²⁵, 周从容²⁶, 周文浩⁸, 曾勇²⁷, 张学军²⁸, 黄涛生²⁹, 郑茜³⁰, 秦胜营³¹, 于世辉³², 关静¹, 王洪阳¹, 王大勇¹, 赵立东¹, 王慧君⁸, 孔令印¹⁵, 宣黎明¹⁵, 冒燕¹⁵, 祝轶君¹⁵, 徐君玲¹⁵, 王剑青¹⁵, 王莉¹⁵, 赵婷¹⁵, 秦一丁¹⁵, 夏滢颖¹⁵, 樊丽霞¹⁵, 赵丁丁¹⁵, 邱浩¹⁵, 贺林^{31*}

1. 解放军总医院, 北京 100039;
2. Boston Children's Hospital, Boston 02101-02117, USA;
3. 中南大学湘雅医院, 长沙 410008;
4. 广西妇幼保健院, 南宁 530005;
5. 山东大学, 济南 250100;
6. 中国科学院北京基因组研究所, 北京 100101;
7. 哈尔滨医科大学, 哈尔滨 150081;
8. 复旦大学附属儿科医院, 上海 200433;
9. 重庆市妇幼保健院, 重庆 400010;
10. 国际和平妇幼保健院, 上海 200030;
11. 贵州省临床检验中心, 贵阳 550002;
12. 第四军医大学附属西京医院, 西安 710032;
13. 内蒙古自治区妇幼保健院, 呼和浩特 010020;
14. 苏州市立医院, 苏州 215000;
15. 上海交通大学, 上海 200240;
16. 广州市妇女儿童医疗中心, 广州 510000;
17. 北京大学第三医院, 北京 100191;
18. 甘肃省医学科学研究院, 兰州 730030;
19. 宁夏医科大学, 银川 750004;
20. 复旦大学, 上海 200433;
21. 首都医科大学附属北京妇产医院, 北京 100069;
22. 第四军医大学唐都医院, 西安 710038;
23. 南方医科大学, 广州 510515;
24. 重庆第三军医大学, 重庆 400038;
25. 四川省人民医院, 成都 610072;
26. 贵阳医学院附属医院, 贵阳 550000;
27. 深圳中山泌尿外科医院, 深圳 518045;

引用格式: 王秋菊, 沈亦平, 邬玲仟, 等. 遗传变异分类标准与指南. 中国科学: 生命科学, 2017, 47

Wang Q J, Shen Y P, Wu L Q, et al. Standards and guidelines for the interpretation of sequence variants. Sci Sin Vitae, 2017, 47, doi: 10.1360/N052017-00099

28. 安徽医科大学第一附属医院, 合肥 230022;
29. Cincinnati Children's Hospital Medical Center, Cincinnati 45229-3039, USA;
30. Great Ormond Street Hospital, London WC1N 3JH, UK;
31. 上海交通大学 Bio-X 研究院, 上海 200030;
32. 广州金域医学检验中心有限公司, 广州 510000;
* 联系人, E-mail: helinhelin3@vip.163.com

收稿日期: 2017-04-13; 接受日期: 2017-04-20; 网络版发表日期: 2017-06-13

摘要 美国医学遗传学与基因组学学会(The American College of Medical Genetics and Genomics, ACMG)曾制定过序列变异解读指南. 在过去的十年中, 随着新一代高通量测序的出现, 测序技术有了快速发展. 利用新一代测序技术, 临床实验室检测遗传性疾病的产品种类不断增加, 包括基因分型、单基因、基因包、外显子组、基因组、转录组和表观遗传学检测. 随着技术的复杂性日益增加, 基因检测在序列解读方面不断面临着新的挑战. 因此 ACMG 在 2013 年成立了一个工作组来重新审视和修订序列变异解读的标准和指南, 工作组包括 ACMG、分子病理协会(the Association for Molecular Pathology, AMP)和美国病理学家协会(the College of American Pathologists, CAP)的代表. 该工作组由临床实验室主任和临床医生组成. 本报告代表了工作组中来自 ACMG, AMP 和 CAP 的专家意见. 本报告提出的建议可应用于临床实验室的各种基因检测方法, 包括基因分型、单基因、基因包、外显子组和基因组. 本报告建议使用特定标准术语来描述孟德尔疾病相关的基因变异——“致病的”、“可能致病的”、“意义不明的”、“可能良性的”和“良性的”. 此外, 本报告描述了基于典型的数据类型(如人群数据, 计算数据, 功能数据, 共分离数据)对变异进行五级分类的标准过程. 由于临床基因检测分析和解读中不断增加的复杂性, ACMG 强烈建议临床分子基因检测应在符合临床实验室改进修正案(CLIA)认证的实验室中进行, 其检测结果应由通过职业认证的临床分子遗传学家或分子遗传病理学家或相同职能的专业人员解读.

关键词 ACMG 实验室指南, 临床遗传检测, 解读, 报告, 序列变异术语, 变异报告

随着遗传病患者样本中所检测基因数目的快速增加, 临床分子实验室检测到越来越多的新的序列变异. 某些表型仅与单个基因相关, 而多数表型与多个基因相关. 对某个给定序列变异的临床意义进行分级解读, 从某个变异几乎可以肯定是某种疾病的致病性变异到几乎可以肯定是良性变异. 虽然 ACMG 之前的建议提供了序列变异的解读分类及解读的算法, 但并没有提供明确的术语或详细的变异分类指导^[1]. 本研究依据专家意见及经验数据, 阐述了最新的序列变异分类标准和指南.

1 方法

2013 年, ACMG, AMP 和 CAP 的成员, 代表临床实验室主任和临床医生成立了一个工作组, 该工作组依据专家建议、工作组共识和公众反馈开发了一种

可以对现有的证据进行加权的系统, 并应用此系统对序列变异进行标准分类. 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. 第一次调研于 2013 年 2 月开展, 该调研旨在评估参考术语的偏好, 调研结果公布在同年 ACMG 年会公开论坛上, 该年会有超过 75 个与会者参加. 调研结果代表超过 45 个位于北美的实验室. 调研和公开论坛的结果表明: (i) 五级术语系统“致病的”、“可能致病的”、“意义不明的”、“可能良性的”和“良性的”是优选认可的, 且已在多数实验室使用; (ii) 工作组的首要重点应着重于孟德尔疾病和线粒体变异.

在第一次调研中, 参与的实验室被要求提供他们的变异评估方法, 最终有 11 个实验室提供并分享

了他们的变异评估方法. 通过分析所有提交的方法, 工作组制定了一组准则, 包括变异证据评估的加权标准体系和应用这个标准将变异归类为五类的分类准则. 在今后的几周时间里, 工作组成员通过在自己实验室或其他机构已进行分类的变异来验证这个方案. 另外, 还将典型变异的常见证据进行分类, 来测试工作组成员达成一致的现有方法是否可以对这些变异进行分类. 2013 年 8 月, 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.

2013 年 11 月, 工作组在 AMP 会议期间举行了超过 50 人参加的研讨会, 提出了修订后的分类标准和两个评分体系. 一个体系与这里介绍的方法是一致的, 另一个体系则是一个分数体系, 每一项标准都有一个分数, 正分数为致病标准, 负分数为良性标准, 根据总分数进行变异分类. 参与者使用此系统并进行反馈, 回答在评估变异证据过程中他们如何权衡各个标准(如强、中度或支持、或不使用). 参与者的反馈结果再次综合到这里介绍分类体系中. 但要指出的是, 虽然大多数回复更倾向于分数评价体系, 但本工作组认为, 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.

工作组还评估了文献中推荐的其他专业协会和工作组在乳腺癌、结肠癌和囊性纤维化中已制定的变异分类指南, 以及在特定疾病中应用统计分析来进行变异定量评估的方法^[2-5]. 这些变异分析指南在特定条件下是有效的, 但很难将他们推荐的标准应用于所有基因变异及不同的实验室条件. 本文描述的变异分类方法适用于所有孟德尔基因变异, 包括单基因、多基因包、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的提高而与时俱进. 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. 总论

2.1 术语

突变是指核苷酸序列的永久性改变, 而多态性是指频率超过 1% 的变异. 虽然术语“突变”和“多态性”已被广泛使用, 但由于这两个术语已经错误地与致病性和良性结果关联了起来, 所以往往会造成混淆. 因此, 建议使用“变异”加以下修饰词替代上述两个术语: 致病性的、可能致病性的、意义不明的、可能良性的或良性的. 虽然这些修饰词不可能适用所有的人类表型, 但是正如本指南提出的它包含了孟德尔疾病相关的变异分类五级系统. 建议所有致病性(包括可能致病)的结论必须注明疾病及相应的遗传模式(如 c.1521_1523delCTT(p.Phe508del), 致病性, 囊性纤维化, 常染色体隐性遗传).

应当注意的是, 一些实验室可能选择其他等级(如意义不明的变异的子分类, 特别是内部使用时), 这种做法不被认为与指南不一致. 还应当指出的是, 某种程度上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^[6]. 虽然拷贝数变异分类系统也包括五级分类标准, 但是它使用“临床意义不明确-可能致病的”和“临床意义不明确-可能良性的”. 由于本指南提出的“可能的”变异分类标准比拷贝数变异分类指南中用到的“可能的”包含更强的证据, 合并这两个“可能的”分类会使医务工作者和临床报告接收者产生混淆, 因此大多数工作组成员不支持使用“意义不明的”来修饰“可能致病的”或“可能良性的”. 然而, 有人认为“可能的”一词的使用应限于有数据支持其致病性或良性可能性很大的变异. 虽然对“可能的”一词没有量化的定义, 但是在某些变异分类系统中已有指导性意见. 然而, ACMG 开放论坛的一项调查建议“可能的”这一术语具有更广泛的适用性. 认识到这一点, 建议术语“可能致病的”和“可能良性的”用来说明一个具有大于 90% 可能引起致病或者可能良性的变异, 尽管是人为的界定, 但还是给实验室提供了一种共同的定义. 类似地, 国际癌症机构指南^[2]支持致病性的确定水平为 95%, 但是工作组(通过 ACMG 公开论坛期间的反馈确认)认为, 临床医生和患者愿意容忍略高的错误机会, 从而

做出确定为 90% 的决定。还应当指出的是, 考虑到多数疾病具有异质性, 目前大多数变异没有数据能将它们量化性地归于上述五个变异类别之一。希望随着时间的推移, 能够建立实验和统计方法来客观地赋予变异的致病可信度, 并且采用更严格的方法来定义临床专业人员所期望达到的可信度, 从而能更完整地诠释这些术语及可能性。

新术语的使用可能需要专业培训, 鼓励专业团队对所有实验室和医务工作者进行这些术语的培训, 也鼓励实验室直接对其开具检测报告单的医生进行培训教育。

2.2 命名

建议通过一套规范的标准对变异进行统一命名来确保变异的明确定义, 并实现基因组信息的有效共享和下游使用。标准的基因变异命名由人类基因组变异协会(the Human Genome Variation Society, HGVS)维护和版本化(<https://www.hgvs.org/mutnomen>), 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^[6]。实验室应该注意他们在实验方法中所使用的版本。当描述变异时, 可利用这些工具提供正确的 HGVS 命名(<http://mutalyzer.nl>)^[7]。临床报告应该包含参考序列以确保该变异在 DNA 水平上的明确命名, 并提供编码和蛋白质命名法来协助功能注释(如“g”为基因组序列, “c”为编码 DNA 序列, “p”为蛋白质, “m”为线粒体)。

编码命名应该使用翻译起始密码子 ATG 中的“A”作为位置编号 1 来描述。在传统命名已被使用的地方, 当今命名应该对传统命名进行额外注释。参考序列应该是完整的, 并来源于具有版本号美国生物技术信息参考序列数据库(<http://www.ncbi.nlm.nih.gov/Refseq/>)或 LRG 数据库(<http://www.lrg-sequence.org>)。基因组坐标应根据标准基因组版本(如 hg19)或覆盖整个基因(包括 5' 和 3' 非翻译区以及启动子)的基因组参考序列来界定。当描述编码变异时, 应该在报告中使用的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。协会支持的参考转录本通常可以通过 LRG 数据库(<http://www.lrg-sequence.org>)、CDS 共识数据库(<https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi>)、人类基因突变数据库(<http://www.hgmd.cf.ac.uk>)、ClinVar(<http://www.ncbi.nlm.nih.gov/clinvar>)或特异基因座数据库来确定。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。

nih.gov/clinvar)或特异基因座数据库来确定。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。

HGVS(<https://www.hgvs.org/mutnomen>)并未覆盖所有类型的变异(如复杂变异), 但是复杂变异的可能描述已被报道^[8]。此外, ACMG 支持 HGVS 命名规则之外的三种特殊例外: (i) 除了当今 HGVS 推荐的“*”和“Ter”, “X”仍然被认为用于报告无义变异; (ii) 建议根据指定变异选择的参考转录本对外显子进行编号; (iii) 通常因为临床解释直接评估致病性, 所以推荐使用术语“致病性”而不是“影响功能”。

2.3 文献及数据库使用

目前人类基因组中大量变异不断被发现, 且已被许多数据库广泛收录。当临床实验室需要对某一变异进行分类并出具报告时, 可在已有的数据库及发表的文献中寻找有价值的参考信息。如上文提及, 序列数据库还可用于确定合适的参考序列。数据库有助于信息收集, 但需谨慎使用。

人群数据库(表 1)适用于获取某变异在大规模人群中发生频率的相关信息。需要注意的是, 人群数据库中的信息不仅来源于健康个体, 也包含致病性的变异。另外, 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。在使用人群数据库时, 须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; (如能确认)数据库是否收录了同一家庭多名成员的信息以及数据库收录的受试者的年龄范围。

疾病数据库(表 1)主要包含病患中发现的变异以及对其致病性的评估。疾病数据库和特定基因的数据库常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。因此, 在使用疾病数据库时, 考虑患者是如何被确诊的尤为重要, 如下所述:

当使用数据库时, 临床实验室应做到: (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,

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。

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。

变异解读也需要检索科学和医学文献。在参考一些采用旧的命名和分类系统或基于单一观察结果的文献时需要慎重。在参考携带某一变异并伴有相关表型的个体和家系的信息时，考虑患者是如何被确诊尤为重要。在评估这些文献的数据时需要谨慎客观，这是由于受累患者及相关个体在基于不同背景和规模的研究中常常被多次重复报道。重复报道的发生可能是由于作者重叠、实验室间合作或先证者及其家庭成员同时被不同临床系统随访。而这些重复报道可能会导致受累个体被错误地重复计数，进而使变异频率假性增高。**作者或其研究机构互相重叠是发现数据集重复的第一线索。**

临床实验室应建立一个内部系统对已报告的基因序列变异及临床诊断进行记录。这对于分析基因

型-表型之间的相关性，以及该变异在患者和正常人群中的发生频率尤为重要。临床实验室也应该积极提交变异数据到相关数据库，如 ClinVar 数据库，包含提交临床评估信息以及用于变异分类的证据，以帮助人们不断加深对人类遗传变异所产生的效应的理解。在任何时候，提供临床数据应遵循“健康保险携带和责任法案(HIPAA)”对个人隐私保护的规定。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。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。

2.4 生物信息学计算预测程序

各种公共和商业化工具可以辅助解读序列变异. 每种工具使用的算法可能有差异, 但都会包含序列变异在核苷酸及氨基酸水平上作用影响的判断, 包括变异对主要转录本, 可变转录本, 其他基因组元件影响作用的确认, 也包括对蛋白质潜在影响作用的判定. 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^[9].

错义改变的影响作用是由不同的条件决定的, 例如一个氨基酸或核苷酸的进化保守性、其在蛋白质序列中的位置及其上下游序列, 以及氨基酸置换导致的生化结果等. 对各种计算机算法中的一个或几个条件进行评测可以进一步评估错义改变带来的影响. 已经有一些工作在评估预测软件的预测性能, 是通过对这些预测软件之间的相互比较评估他们预测已知致病突变的能力来实现的^[10-13]. 一般情况下, 多数算法预测已知致病的错义突变的准确率能达到 65%~80%^[12]. 但是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^[14]. 目前临床实验室常用的错义变异解读工具有 PolyPhen 2^[15], SIFT^[16]和 MutationTaster^[17]. 用于预测错义变异的生物信息分析工具见表 2.

目前已开发出许多用于预测剪接的软件, 这是基于内含子或外显子水平上剪接位点的丢失或产生原理基础上而完成的^[18]. 一般情况下, 相对于特异性(60%~80%), 预测工具在预测剪接位点异常方面具有较高的敏感性(~90%~100%)^[19,20]. 一些常用的剪接位点预测分析计算工具见表 2.

虽然许多不同的分析软件程序使用不同的算法进行预测, 但其基本原理是相似的; 因此, 在序列解读中, 不同软件工具组合的预测结果被视为单一证据而不是相互独立的证据. 因为每个软件工具基于他们使用的算法都各有优缺点, 所以仍然建议使用多种软件进行序列变异解读; 很多情况下, 预测性可能因为基因和蛋白质序列的不同而有差异. 无论如何, 这些软件分析结果只是预测, 他们在序列变异解读中的应用应该慎重, 不建议仅使用这些预测结果作为唯一证据来源进行临床判断.

3. 序列变异解读的拟定标准

以下评估变异证据的方法是用了解释在临床诊断实验室中具有疑似遗传(主要指孟德尔遗传)疾病患者的变异. 并不适用于解读体细胞变异、药物基因组(PGx)变异、或者是多基因非孟德尔复杂疾病相关的基因变异. 在外显子组或基因组研究中, 对候选基因(意义不明确的基因(GUS))应用这些准则时应当谨慎(见下面注意事项), 因为本指南目的不是满足鉴定新致病基因的研究需求.

虽然这些方法可用于评估在健康个体中发现的变异或与测试指征不相关的变异, 但是正如在指南的几个部分中所述, 对于与指征无关的有较低先验致病性的变异时需更加谨慎. 尽管期望本指南适用于变异分类, 无论是通过分析单基因、基因包、外显子组、基因组或者转录组而鉴定的, 重要的是要关注与疾病有关的致病变异和虽然预测为对蛋白有破坏/损伤但却与疾病无充分关联的变异之间的区别. 这些规则旨在确定在孟德尔遗传病中有明确作用的基因的变异是否对该遗传疾病是致病的. 针对具体的病人, 致病性判定应该独立于对疾病病因的解读. 例如, 某变异在一个案例中被评估为“致病的”, 而在另一个案例中, 由于不能解释该疾病, 就对这个位点不给出“致病的”评价, 这样的情况是绝对不允许的. 确定致病性需要将全部的证据汇集在一起, 包括所有的案例分析, 最终得出一个结论.

此指南的分类方法可能比目前实验室应用的标准更为严格. 这将导致很大一部分的变异被归类为“意义不明确的”. 希望这种方法可以大量减少那些没有足够分类证据支持而报告为致病原因的变异. 需要注意的是, 当临床实验室报告一个变异为“致病的”时, 医疗单位很可能把其当作“可指导临床作为的(actionable)”, 基于这个判断, 从而会改变对患者的治疗、监测^[21], 或去除对基因型为阴性的家庭成员的治疗、监测(参见下面的医务工作者应该如何使用这些指南和建议).

本指南提供了两套标准: 一是用于对致病或可能致病的变异进行分类(表 3), 另一是用于对良性或可能良性的变异进行分类(表 4). 致病变异标准可分为非常强(very strong, PVS1), 强(strong, PS1~4); 中等(moderate, PM1~6), 或辅助证据(supporting, PP1~5). 良性变异证据可分为独立(stand-alone, BA1),

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/fannsdbs/	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
	FSPLICE	http://www.softberry.com/berry.phtml?topic=fsplce&group=progrms&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
核酸保守性预测	PhyloP	http://compugen.bscb.cornell.edu/phast/	保守打分及鉴定保守元件
		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

强(strong, BS1~4), 或辅助证据(BP1~6). 其中, 数字只是作为有助于参考的分类标注, 不具有任何意义. 每个类别中的数字不表示分类的任何差异, 仅用来标记以帮助指代不同的规则. 对于一个给定的变异, 用户基于观察到的证据来选择标准. 根据表 5 的评分规则把标准组合起来进而从 5 级系统中选择一个分类. 这些规则适用于变异上的所有可用数据, 无论是基于调查现有案例获得的数据, 还是来源于先前公布的数据. 未发表的数据也可以通过公共数据库(如 ClinVar 或位点特异数据库)和实验室自有数据库获得. 为了对变异分类具有较好灵活性, 基于收集的证据和专业判断, 可以把某些依据用到不同的证据水平上去. 例如, 如果一个变异多次和已知致病性变异

处于反式位置(位于另一染色体上), PM3 可以上调到强(进一步指导见 PM3 BP2 顺/反式检测). 相反, 在数据并不像描述的那么强的情况下, 可以改判变异到一个较低的水平(见表 3 注 2 PS4). 如果一个变异不符合分类标准(致病的或良性的), 或良性和致病的证据是相互矛盾的, 则默认该变异为“意义不确定的”. 程度判断评价标准如表 6 所示. 请注意, 当考虑所有依据以解读变异证据强度的差异时, 需专家介入进行判断.

下面更详细的变异分类标准(表 3 和 4)中提及了某些概念的解释, 并提供实际使用中的实例和/或误区或易犯错误的地方. 这部分应该与表 3 及 4 一同阅读.

表 3 致病变异分级标准

致病性证据	分类
非常强	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.
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: 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. 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. 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. 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.
中等	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). 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. 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. 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. 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. PM6: 未经父母样本验证的新发变异.
支持证据	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. 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. 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. 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. 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.

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 数据库中位点基因频率>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. 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.

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

3.1 PVS1 极强致病性变异

某些特定类型的变异(如无义突变、移码突变、经典剪接位点 ± 1 或 2 点突变、起始密码子变异、单个或多个外显子缺失)被认为因无转录产物或由无义突变引起的转录子降解, 导致基因产物完全缺失而破坏基因功能. 当将这类变异归类为致病性时, 从业人员需谨慎考虑以下原则:

(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. 例如, 有些基因(如许多肥厚性心肌病基因)只有杂合错义突变时才致病, 而杂合无功能变异却是良性的. 仅基于这一项证据来看, 对显性肥厚性心肌病来说, *MYH7* 基因上出现一个新的杂合无义突变不一定是致病的, 而 *CFTR* 基因上出现一个新的杂合无义突变则有可能是一个隐性致病变异.

(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. 据此所预测的截短蛋白的长度也是致病性评估的因

素, 但这些变异未经功能分析是无法进行判定的.

(iii) 就剪接位点变异而言, 因外显子剪切位点的供体/受体位点改变或产生了新的剪切位点, 从而可能导致外显子丢失、缩短, 也可能会使内含子序列变成外显子部分. 虽然剪切位点变异可能被预测为无功能变异, 然而该变异类型造成的影响需要通过 RNA 或蛋白质功能分析确认. 还必须考虑可读框内缺失/插入的可能性, 其长度变化较小(PM4), 可以保留蛋白质的关键结构域, 因此导致轻微或中性效应, 或功能获得效应.

(iv) 基因会有不同的转录本, 哪一种转录本与生物学功能相关, 在哪些组织会表达哪些转录本, 这些都是需要进行重点考虑的. 如果一个截短变异只限于一个或并非所有转录本, 则必须谨慎考虑到可能存在其他同功型蛋白质, 防止过度解释.

(v) 如果发现一个无功能变异位于某个外显子上, 而该外显子先前无致病变异报道, 那么该外显子可能被选择性剪切了, 此时需谨慎考虑该变异的致病性. 当预测的截短变异是偶然发现时(与检测特征无关)则应特别小心, 在这种情况下该位点致病的可能性非常低.

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	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		多个家系患者中共 分离 PP1	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		隐性疾病中, 反式检 测到致病变异 PM3		
其他数据库		较好信誉的未共享数据 =良性 BP6	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			

3.2 PS1 突变为同一氨基酸

多数情况下, 尤其是当致病机制是蛋白质功能发生改变时, 如已确定某一错义变异是致病变异, 应考虑到与其位于同一变异位点的不同形式的碱基改变也可能产生相同的错义突变结果——氨基酸改变相同(如 c.34G>C(p.Val12Leu)和 c.34G>T(p. Val12 Leu)), 那么, 这些变异也应是致病突变。此外, 还应考虑到, 变异可能不是通过改变氨基酸的水平, 而是通过改变 DNA 的序列来发挥作用, 例如, 破坏剪接位点(可通过软件分析确定), 在这种情况下, 上述的假设是不成立的。

3.3 PS2 PM6 新发变异

当将一个新发变异(父母样本检测结果阴性)归类为强的致病证据时, 需要满足以下条件: (i) 身份检验表明患者的父母是其生物学父母。注意如果父母的身份是假定的而没有被证实, 则判定为 PM6; (ii) 患者的家族史符合新发变异特征。例如, 显性遗传病

患者的父母均未患病。在存在生殖细胞嵌合现象时也可能有 1 个以上同胞患病; (iii) 患者的表型与变异基因异常引起的表型相吻合。例如, 患者具有特殊面容、多毛和上肢缺陷(即 Cornelia de Lange 综合征), 检测到 *NIPBL* 基因的新生突变即为强致病证据, 而患者仅表现为非特异性的发育迟缓, 通过外显子组测序发现的该基因的新发变异, 则判断此变异致病性的证据较弱。

3.4 PS3 BS3 功能研究

功能实验研究是一种研究变异致病性的非常强大的工具, 然而并非所有的功能研究都能有效地预测基因或蛋白的功能。例如, 一些酶学实验利用成熟完善的方法可以用来评估错义变异在代谢途径中对酶活性的影响(如 α -半乳糖苷酶功能实验); 而另一方面, 某些功能实验在评估变异对蛋白质功能的影响时缺乏一致性。评估一个功能检测方法是否有效时, 必须考虑该功能实验在多大程度上反映了其发挥功

能的生物环境。例如,与体外表达蛋白相比,直接在患者或动物模型的活检组织中进行酶的功能实验更有说服力。同样,可以反映蛋白质全部生物学功能(如酶分解底物功能)的实验则比仅反映一部分功能(如一种有附带结合能力的蛋白水解 ATP 的功能)的实验证据性更强。功能实验的有效性、重复性和稳定性应重点考虑,这些参数用来评估功能实验的分析性能以及判定样本诊断信息的完整性,该完整性容易受标本采集的方法及时间、存储及运输的影响。CLIA(临床实验室改进修正案)认证实验室建立的检测方法或商品化试剂盒可减少这些因素对实验的影响。评估变异在剪接位点、编码序列、非翻译区以及更深的内含子区域的影响时,对变异在信使 RNA 水平(如信使 RNA 的稳定性、加工或翻译)进行评估,可以提供丰富的信息。相关的技术方法包括对 RNA 和/或互补 DNA 衍生物进行直接分析,以及体外微小基因剪接分析。

3.5 PS4 PM2 BA1 BS1 BS2 变异频率及对照人群的使用

通过搜索公共人群数据库(如千人基因组数据库, NHLBI 外显子测序数据库, EXAC 数据库; 表 1), 并利用已发表文献中相同种族的对照数据进行基因变异频率分析(译者注: 此条款在指南更新时会有修改), 通过分析变异基因在对照人群或普通人群中的携带频率, 有助于评估该变异的潜在致病性。NHLBI 外显子测序数据库来源于白种人和非裔美国人群, 根据其数据覆盖量能够识别是否存在基因变异。尽管千人基因组数据库缺乏评估基因变异能力, 但它囊括了更多的种族人群, 因此其数据具有更广泛代表性的。EXAC 数据库近期发布了一组来源于不同人群的 6 万多个外显子组的等位基因频率数据, 包括了大约三分之二的 NHLBI 外显子测序数据。一般情况下, 某一等位基因在对照人群的频率大于疾病预期人群(表 7)时, 可认为是罕见孟德尔疾病良性变异的强证据(BS1), 如果频率超过 5% 时, 则可认为是良性变异的独立证据(BA1)。此外, 如果疾病发生在早期, 且变异在健康成人中以隐性(纯合子)、显性(杂合子)或 X-连锁(半合子)的状态存在, 那么这就是良性变异的强证据(BS2)。如果数据库中未能检出变异的存在, 应该确认建立该数据库采用的测序读长深度是否足以检测出该位点上的变异。如果在一个大样本的普

通人群或队列数据的对照人群(>1000 人)中变异不存在(或隐性遗传的突变频率是低频), 并且携带此变异的患者与对照人群为同一种族, 那么可以认为该变异是致病性的中等证据(PM2)。许多良性变异是“个体化的”(即个人或家系独有的), 因此即使在相同种族的人群中缺乏也不能作为致病性的充足甚至强的证据。

当孟德尔遗传病表型显著、频率差异大且是早期发病时, 使用通过“病例-对照”人群研究获得的变异数据库进行变异分析是最有效的。临床实验室检测的患者可能包括“排除”某一疾病的个体, 因此他们可能不能作为表型显著的病例; 当使用普通人群作为对照群体时, 具有亚临床疾病的个体总是可能存在的。在这两种情况下, 认为检测出的变异致病性证据不充分。变异频率有统计学显著差异可以假定为致病性的支持证据。与此相反, 对于统计差异不显著, 特别是极为罕见变异和不明显的表型, 应谨慎解释。

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, 例如, *APOE* 基因 E4/E4 纯合子与 E3/E3 纯合子相比, OR 值为 13(https://www.tgen.org/home/education-outreach/past-summer-interns/2012-summer-interns/erika-kollitz.aspx#.VOSi3C7G_vY)。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。在上面 *APOE* 的例子中, CI 为 10~16。在线可获得简单的 OR 值计算器(<http://www.hutchon.net/ConfidOR.htm/>; <http://easy-calculator.com/statistics/odds-ratio.php/>)。

3.6 PM1 热点突变和/或关键的、得到确认的功能域

某些蛋白结构域对蛋白质的功能起到了关键作用, 如果在这些结构域上发现的所有错义突变均已

疾病	基因	遗传模式	种群	发病率	携带者频率	常见变异	变异分类	ESP6500 AA MAF	ESP6500 EA MAF	ESP6500 All MAF	一致性	分类证据						
囊性纤维化	CFTR	AR	高加索人	0.031%	3.6%	p.F508del	Ex24: p.F508del(致病的)	n/a	n/a	n/a	n/a	多个研究						
							Ex11: c.1523T>G/p.F508C (良性的)	0.070%	0.150%	0.120%	不一致	(变异在 EVS 数据库不存在) Kobayashi(1990) Am J Hum Genet 47, 61						
							Ex23: c.3870A>G/p.(=)(良性的)	15.090%	2.970%	7.070%	部分一致性	AA MAF						
							5' UTR: c.-8G>C(良性的)	1.160%	5.550%	4.060%	部分一致性	EA MAF						
							IVS6: c.743+40A>G(良性的)	0.700%	5.190%	3.670%	部分一致性	EA MAF						
							Ex12: c.1242C>T/p.(=)(良性的)	0.360%	1.310%	0.990%	No	PAH 数据库						
							Ex12: c.1278T>C/p.(=)(良性的)	13.550%	0.090%	4.650%	部分一致性	AA MAF						
苯丙酮尿症	PAH	AR	北欧人	0.010%	2.0%		IVS12: c.1316-35C>T(良性的)	0.320%	2.630%	1.850%	部分一致性	EA MAF						
							Ex9: c.963C>T/p.(=)(良性的)	5.170%	0.000%	1.750%	部分一致性	AA MAF						
							Ex7: c.489T>G/p.(=)(良性的)	7.010%	0.050%	2.410%	部分一致性	AA MAF						
							IVS20: c.1964+17G>T(良性的)	0.200%	0.810%	0.610%	不一致	多个研究						
							Ex61: c.10515C>A/p.S3505R (良性的)	0.230%	1.130%	0.820%	不一致	多个研究						
							Ex66: c.11738G>A/p.R3913H (良性的)	1.270%	0.000%	0.430%	不一致	AA MAF						
							Ex17: c.1587T>C/p.(=)(良性的)	1.380%	6.860%	5.010%	部分一致性	EA MAF						
Ret 综合征	MECP2	X 连锁	非特定的	0.012%	新发		Ex65: c.11525G>T/p.R3842L (良性的)	0.360%	2.430%	1.730%	部分一致性	EA MAF						
							Ex61: c.10585G>C/p.E3529Q (良性的)	3.950%	0.010%	1.350%	部分一致性	AA MAF						
							Ex4: c.1161C>T/p.(=)(良性的)	0.030%	0.000%	0.010%	部分一致性	AA MAF						
							Ex4: c.608C>T/p.T203M(良性的)	0.000%	0.060%	0.040%	部分一致性	多个研究						
							Ex4: c.683C>G/p.T228S(致病的)	0.830%	0.000%	0.300%	部分一致性	RETT 数据库						
							IVS31: c.8047-15C>T(良性的)	0.000%	0.020%	0.020%	部分一致性	EA MAF						
							Ex31: c.6836G>A/p.Gly2279E (良性的)	0.000%	0.120%	0.080%	部分一致性	EA MAF						
Kabuki 综合征	KMT2D (MLL2)	AD	非特定的	0.003%	新发		Ex2: c.6478G>A/p.A2160T (良性的)	1.460%	0.000%	0.490%	部分一致性	AA MAF						
							Ex31: c.6478G>A/p.A2160T (良性的)	1.250%	0.000%	0.390%	部分一致性	AA MAF						
							Ex2: c.856A>G/p.R286Gly (良性的)	0.780%	0.000%	0.250%	部分一致性	AA MAF						
							Ex2: c.35deIG(致病的)	0.090%	1.080%	0.740%	不一致	多个研究						
							Ex4: c.845G>A/p.C282Y(其他的)	1.520%	6.410%	4.750%	不一致	多个研究						
							CHARGE 综合征	CHD7	AD	非特定的	0.010%	新发						
GJB2 相关听力损失	GJB2	AR	非特定的	0.067%	2.5%	c.35deIG	Ex2: c.35deIG(致病的)	0.090%	1.080%	0.740%	不一致	多个研究						
	HFE	AR	混合的	0.040%	8.3%	p.C282Y	Ex4: c.845G>A/p.C282Y(其他的)	1.520%	6.410%	4.750%	不一致	多个研究						

被证实为致病突变, 且这些结构域中一定没有已知的良性突变, 那么这就能作为致病的中等证据. 此外, 基因中某些功能尚未确定的区域已被证实存在许多突变热点, 若突变发生在基因突变热点上, 且一个或多个邻近残基中存在较高频率的已知致病突变, 那么这也能作为致病的中等证据.

3.7 PM3 BP2 顺式/反式检测

检测双亲样本以确定变异在基因上以顺式(*in cis*)(位于基因的同一拷贝)或是反式(*in trans*)(位于基因的不同拷贝)方式排列, 这对评估变异的致病性非常重要. 例如, 当两个杂合变异发生在隐性遗传病的致病基因上时, 如果已知其中一个变异为致病变异, 那么当另一个待分类变异与其呈反式排列时, 这可以作为待分类变异的中等致病证据(PM3). 另外, 若待分类变异与多个已知致病变异均呈反式排列, 则该证据可升级为强致病证据. 但是, 若待分类变异在普通人群中存在, 则需要用统计学方法判断该现象是否为随机共发生事件. 相反, 当已知致病变异与另一个待分类变异呈顺式排列时, 这可以作为待分类变异的良性支持证据(BP2). 如果发生在隐性遗传病致病基因上的两个杂合变异的致病性均未知, 那么确定它们以顺式或是反式排列, 并不能为判断其中任一变异的致病性提供更多信息. 但是, 如果两者以顺式排列, 则该基因两个拷贝均受影响的可能性将会降低.

对于显性遗传病而言, 若待分类变异与致病变异呈反式排列, 则可作为该变异的良性支持证据(BP2); 对于特定研究成熟的疾病模型, 甚至可以考虑将其作为独立良性证据(如 *CFTR* 相关变异的评价)^[3].

3.8 PM4 BP3 由于框内缺失/插入和终止密码子丧失导致的蛋白长度改变

相较于单一的错义突变所导致的蛋白质长度变化, 一个或多个氨基酸的缺失或插入、以及由终止密码子变为翻译氨基酸的密码子(如终止密码子丢失)而导致的蛋白质延长更可能破坏蛋白质功能. 因此, 框内缺失/插入以及终止密码子丢失可作为中等致病证据. 缺失、插入或延伸范围越大, 缺失区域的氨基酸越保守, 则支持致病的证据越强. 相反, 在重复区域或在进化中不是很保守的区域中小的框内缺失/插

入致病的可能性较小.

3.9 PM5 同一位置新的错义变异

如果一个新发错义突变发生在一已知致病突变导致相同氨基酸改变的位置上(如 *Trp38Ser* 和 *Trp38Leu*), 那么可作为中等致病证据(但不能假定一定是致病的), 尤其当新的突变比已知致病错义突变更保守时. 此外, 不同的氨基酸变化可能导致不同的表型. 例如, *FGFR3* 基因编码的 *Lys650* 残基的不同变化与不同的临床表型相关: *p.Lys650Gln* 或 *p.Lys650Asn* 会导致轻度软骨发育不良; *p.Lys650Met* 会导致严重的软骨发育不全伴发育迟缓和黑棘皮病; *p.Lys650Glu* 会导致 2 型发育异常及致命的骨骼发育不良.

3.10 PP1 BS4 共分离分析

在使用家系中变异的共分离现象作为致病性证据时需谨慎. 事实上, 一个与某种表型相关的特定变异在某一家系中的共分离现象是位点与疾病连锁的证据, 而不是变异本身致病性的证据. 一个已经发表的统计方法^[23,24]显示, 在某个家系中鉴定的变异可能与真正的致病变异是连锁不平衡的. 统计模型考虑到了年龄相关的外显率和拟表型率, 一些新的方法也将生物信息分析预测以及与已知致病突变共存作为致病性的单独定量指标^[25]. 将远亲纳入统计之中是很重要的, 因为与核心家系成员相比, 他们不太可能同时有该疾病和变异. 对整个基因进行测序(包括整个内含子和 5'和 3'非编码区)可排查其他致病变异或另一个可能致病的变异的存在. 除非仔细评估基因位点, 否则非致病变异可能被错误地认为是致病变异.

当目标基因的特定变异在多个患病的家系成员中以及不同种族背景的多个家系中与表型或疾病共分离时, 则其作为致病的证据不太会受到连锁不平衡和确认偏倚的影响. 在这种情况下, 该标准可以作为中等或强致病证据而不是支持性证据, 其强度取决于共分离的程度.

另一方面, 一个变异与表型并不共分离时, 为其非致病的强证据. 需要进行仔细的临床评估来排除正常个体的轻度症状和可能的拟表型(患者表型由非遗传或不同的遗传原因引起). 此外, 需确认生物学

家庭关系来排除收养、非生父、精子和卵子捐献以及其他非生物学关系。同时, 外显率下降和年龄依赖性的外显率也必须考虑, 以确保无症状家系成员是真正的无症状。

在临床实验室进行共分离的统计评估可能并不容易, 当鉴定了合适的家系时, 为了确保建模合适, 并避免得出变异与疾病相关性的错误结论, 鼓励临床实验室与统计或群体遗传学专家合作。

3.11 PP2 BP1 变异谱

许多基因具有明确的致病变异和良性变异谱。在某些基因中, 错义突变是导致疾病的常见原因, 且该基因上的良性突变非常少, 那么这种基因上的新发错义突变可作为致病变异的支持证据(PP2)。相反, 有些基因致病的唯一已知变异是截短突变, 该基因上的新发错义突变可作为良性的支持证据(BP1)。例如, *ASPM* 基因的截短变异是该基因引起常染色体隐性遗传小头畸形的主要致病变异类型, 且该基因发生错义多态性突变的频率高, 因此 *ASPM* 基因上的错义变异可认为是良性影响的支持证据。

3.12 PP3 BP4 生物信息分析数据

不能过分相信生物信息分析所得到的结果, 特别是不同的生物信息算法依赖于相同或相近的数据进行预测, 并且大多数生物信息算法未被已知致病变异验证过。此外, 相同算法对不同的基因的预测结果可能完全不同。如果不同种类算法的分析预测结果一致, 那么生物信息分析结果可以作为支持的证据。如果绝大多数算法的预测结果不一致, 则这些预测的结果不能用于对变异进行分类。若某一变异引起的氨基酸改变, 在多个非人哺乳动物物种不太保守的区域中出现, 说明该变异可能不会损害功能, 可以作为良性解读的强的证据。然而, 如果某基因已在人类中发生进化(如参与免疫功能的基因), 那么在判定该基因在非保守区域中发生的变异为良性时必须小心。

3.13 PP4 表型支持

考虑到几乎所有接受疾病针对性测试的患者都有某种表型, 通常, 不将患者表型与某个基因临床特征谱匹配作为判断致病的证据。但是, 如果满足以下条件, 患者的表型可作为支持证据: (i) 临床检测的

灵敏度高, 大多数带有该基因致病突变的患者都被检测为阳性; (ii) 患者有某种明确的综合征的症状, 与其他临床表现几乎无重叠(如戈尔林综合征包括基底细胞癌、掌跖坑和牙源性角化); (iii) 该基因通常不存在太多的良性变异(可通过外显子组等人群测序确定的良性变异); (iv) 家族史与疾病遗传方式一致。

3.14 PP5 BP6 可靠的来源

现在有越来越多可靠来源(如长期专注于某一疾病领域的临床实验室)的致病性分类信息被分享在数据库中, 但分类判断所依据的证据往往并未提供或者很难获取。在这种情况下, 如果分类信息是近期提交的, 那它就可以作为一个单独的支持证据。然而, 还是鼓励实验室共享分类的判断依据, 并与提交者进行沟通以评估和创建分类证据。如果能获得证据, 则不应使用这一条款, 而是应该使用相关的证据。

3.15 BP5 对共发变异的观察

一般情况下, 当某一变异是在一个有明确的遗传病因的疾病患者中被观察到时, 可作为将该变异解读为良性的证据。不过, 也有例外。某一个体可以是某一不相关隐性遗传疾病致病变异的携带者, 因此本证据与隐性遗传性疾病相比, 更支持显性遗传性疾病基因良性变异的分类。此外, 有些疾病当具有多个变异可以导致更严重的疾病。例如, 在一个具有严重表型的显性遗传患者中鉴定了两个变异, 一个是致病的, 一个是新的变异, 父母中的一个也有轻微的疾病, 这种情况下, 必须考虑新的变异致病的可能性, 且新的变异使先证者表型加重。在这种临床情况下, 观察到的第二个新的变异不应分类为良性变异, (尽管在无进一步证据的前提下也不认为该变异是致病的)。最后, 有些疾病已知为多基因遗传模式, 如 *Bardet-Beidel* 综合征, 在第二个基因座位上的额外变异也有可能是致病的, 但应谨慎进行报告。

3.16 BP7 同义变异

人们逐渐认识到经典的剪接序列以外的剪接错误是一类重要的致病机制, 特别是对那些功能丧失为其常见致病机制的基因。因此, 在假设同义核苷酸改变没有影响时应持谨慎态度。然而如果核苷酸位置进化不保守, 且剪接评估算法预测其对剪接一致序列没有影响, 也不会产生新的经典剪接序列, 那么

剪接影响的可能性就比较小. 因此, 如果生物信息分析证据支持(BP4), 可将新发同义变异分类为可能良性. 然而, 如果生物信息分析证据表明剪接可能有影响或怀疑有影响(例如, 发生在隐性遗传病致病基因上, 且与已知致病突变呈反式排列的变异), 那么在有功能评估可以提供更确切的对影响的评估, 或者得到其他可排除该变异致病作用的证据之前, 该类变异应该归类为意义不明确.

4. 序列变异报告

编写简明而内容丰富的临床报告不是一件容易的事情, 因为从检测单个基因, 到多基因包, 再到外显子组和基因组, 变异情况的报告内容复杂程度会大大增加. 为规范报告内容已出台了一些指南文件, 包括符合 ACMG 临床实验室标准的新一代测序检测完整报告示例^[26-29]. 临床报告是实验室检测结果的最终体现, 通常会放入到患者的电子健康档案中. 因此, 有效的报告应该是简明扼要且易于理解的. 报告应该使用清晰的语言书写, 避免使用医学遗传学术语, 当必须要使用时需指明所用术语的定义. 报告应包含所有的检测基本要素, 包括结构化的结果、解释、参考文献、检测方法和适当的免责声明. 《临床实验室改进法案》(CLIA)以及美国病理学家学会在针对新一代测序临床实验标准^[30]中, 也强调了上述基本要素.

4.1 结果

结果部分应根据 HGVS 命名规则(见命名部分)列出变异. 考虑到在基因检测中发现的变异数目越来越多, 以包含基本内容的表格呈现变异结果可能是传达信息的最好方法. 这些基本内容包括在核苷酸(基因组和 cDNA)和蛋白质水平的命名、基因名称、疾病、遗传模式、外显子、合子性及变异的分类. 若亲本来源明确, 也可包括在内. 此外, 如果变异是通过基因分型检测的, 实验室应特别注明受检变异的完整描述及曾用名. 当报告外显子组或全基因组测序结果, 或偶尔报告包含基因数目较多的疾病基因包检测结果时, 将变异按“与表型明确相关的疾病基因的变异”、“与表型可能相关的疾病基因的变异”及(在适当情况下)“附带(次要)发现”进行分类可能有益.

4.2 解读

解读应包含对变异检测结果进行分类的证据, 包括编码蛋白的功能影响预测, 以及检测所发现的变异是否可能全部或部分地解释患者的临床表型. 报告也应包括对临床医生的建议, 这些建议包括一些需补充的临床检测, 如对患者进行细胞酶学/功能的检测, 以及对患者家系其他成员进行的变异检测, 以便为进一步解读变异检测结果提供支持. 解读应当包括检测结果部分描述的全部变异, 以及其他附加信息. 对于各个变异需要注明是否已经在先前的文献、疾病病例或对照数据库中有过报道. 在报告结尾处需要列出对变异检测结果分类时所引用的全部参考文献和信息. 解读部分其他的附加信息可以包括对变异位点进行进化保守性分析的结果总结. 由于医务工作者可能不熟悉预测算法的局限性(详见上文“3.4 生物信息学计算预测程序”小节), 因此, 应该避免报告对个体进行生物信息学预测的计算结果(如分数, 诸如“破坏性”之类的术语), 以免造成医务工作者对报告产生误解. 如果存在疾病的外显率下降和表现度差异, 也需要将有关的讨论包含在最终的报告中.

4.3 方法学

报告中应说明使用的实验方法、检测所涉及的变异类型、检测过程的难点, 以及检测变异所使用的方法的局限性. 需要说明的实验方法应包括核酸的获取方法(如聚合酶链式反应、捕获、全基因组扩增等)以及核酸的检测方法(如双向 Sanger 测序、下一代测序、染色体基因芯片、基因分型技术等), 这些信息可以为医务工作者提供必要的信息, 以帮助其决定是否需要追加实验来跟进这些检测结果. 方法部分还应包括人类基因组组织基因命名委员会批准的正式基因名称、转录本的 RefSeq 登录号和所参考的基因组版本. 对于大的基因包, 基因水平的信息可以通过引用 URL 来加以说明. 实验室还可以选择增加对检测过程中常见问题(如样本质量问题、样品混合污染等)的免责声明.

4.4 患者维权团体、临床实验和研究的获取

尽管不提倡在实验室报告中对患者提供具体临床指导, 但是在报告中提供对于检测结果分类的总

体信息(如全部阳性检测结果)是恰当且有益的。大量病人群体和临床试验现在可用于多种疾病的支持和治疗。实验室可以选择将此信息添加到报告的正文或附加信息,并且与报告一起发送给医务工作者。在遵守医疗保险便携性和责任法案(HIPAA)保护患者隐私的前提下,当某一变异检测结果被归为意义不明确时,实验室可尝试帮助医务工作者和特定的疾病研究小组建立联系。

4.5 变异再分析

随着新的变异证据增加,现有的分类标准需要修订。例如,当大样本的有效的人群变异频率被报道后,许多原本意义不明确的变异,可以因为明确意义而进行重新分类,而检测家系中其他成员的结果也可以导致重新分类。

随着检测变异数量的增加及检测范围的扩大,无论是全外显子检测还是全基因组测序,都可以得到数以百万的变异信息量。如果实验室缺乏有效的分析方法和足够的文献数据库支撑,将无法进行变异再分析。为了满足医务人员和患者的实际需求,实验室应该开展基因检测数据再分析,并明确再分析是否产生额外费用。应该鼓励实验室为帮助医务人员和患者而不断开发更新信息的新途径^[31,32]。

当报告中有针对主要指征的基因中存在临床意义不明的变异,在实验室又无法及时提供更新的数据时,建议医务人员定期查询其不明意义的变异结果是否被更改。另一方面,鼓励实验室在对变异的分类有重要变化时(如致病性或良性的变异被修改)必须主动及时地更新报告。关于医生对病人报告更新方面的责任,可详见 ACMG 有关指南^[33]。

4.6 变异的验证

关于变异验证的建议已在其他地方阐述了^[29,30]。再次重申,对于孟德尔疾病的致病或可能致病变异需进行正交法验证。具体方法包括但不限于以下几种:重新取样和检测、检测父母的变异情况、限制性内切酶消化、对于目标区域重新测序或使用另一种基因分型技术。

5. 特殊考虑

5.1 对临床意义不明确的基因(GUS)中的变异的评估和报告

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。这种情况可出现在当一个基因从未与任何病人表型相关联时,或者与此基因相关联的表型不同于正在被考虑的表型。当推荐的指南应用于 GUS 时必须特别注意。在这种情况下,由于目前指南中的变异分类规则适用于已经明确的基因型-表型关联,但并不适合未知的情况。例如,纵观外显子组或基因组,考虑到所有个体的外显子组中预计约有 1 个新发变异或基因组中约有 100 个新发变异,新发变异的发现不再是致病性的强有力证据。同样地,整个基因组中成千上万个变异可与显著的 LOD 值共分离。此外,许多明显破坏基因或其蛋白的有害变异(无义、移码、典型 $\pm 1, 2$ 剪接位点、外显子水平缺失)可能被检测出来,然而,在对任何疾病表型的解释中,这些变异都不是充分的致病证据。

GUS 中发现的变异可作为候选,并可报告为“意义不明确的基因变异”。如果报道这些变异,应该一直被分类为意义不明确。在任何基因变异可被考虑为疾病的致病原因之前,都需要附加的证据支持基因与疾病的关联^[5]。例如,与罕见表型匹配和存在相同基因上存在有害变异的其他病例使得可以根据此指南对某一变异进行分类。

5.2 在健康个体中评估变异或作为偶然发现

当评估在健康或无症状个体中检测到的变异或者解释与主要检测指征无关的偶然发现的变异时,必须谨慎使用此指南。在这些情况下,所识别变异为致病变异的概率可能远低于疾病靶向性检测。正因为如此,当判定这些变异为致病变异时,不仅需要更强的证据支持,而且需要额外谨慎。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。

5.3 线粒体变异

除了明确的致病变异,线粒体变异的解读是复杂且依旧充满挑战的,此处提出了一些特殊的考虑。

线粒体变异的命名法与核基因的标准命名法不同,使用基因名和 m.编号(如 m.8993T>C)和 p.编号,而不是标准的 c.编号(见命名法)。目前公认的参考序列是人类线粒体 DNA 修订版剑桥参考序列: 基因库序列 NC_012920 gi: 251831106(<http://www.mitomap.org>).

org/MITOMAP/HumanMitoSeq)^[34].

如果已通过检测对异质性水平进行确定, 应该对异质性或同质性, 以及变异异质性的评估进行报道. 不同组织类型的异质性百分比因检测样本的不同而有所改变. 因此, 低异质性水平也必须结合所检测组织进行解读, 且它们可能仅在受累及的组织中才是有意义的, 如肌肉组织. 超过 275 个与疾病相关的线粒体 DNA 变异已被记录(<http://mitomap.org/bin/view.pl/MITOMAP/WebHome>). MitoMap 是线粒体变异及单倍型相关信息的主要来源. 其他资源, 如频率信息(<http://www.mtodb.igp.uu.se/>)、二级结构、序列和线粒体转运 RNA 的比对(<http://mamittrna.ustrasbg.fr/>)、线粒体单倍群(<http://www.phylotree.org/>)^[35]和其他信息(<http://www.mtcommunity.org/default.aspx>), 可能在解读线粒体变异时是有用的.

鉴于线粒体变异评估的难度, 本指南并未包括单独的证据清单. 然而, 任何证据的应用均需要格外谨慎. 线粒体基因组中的基因编码转运 RNA 和蛋白质, 因此, 评估氨基酸的变化仅与蛋白质的编码基因有关. 同样地, 因为很多线粒体变异是错义突变, 截短突变的证据标准可能并不适用. 由于截短突变并不符合多数线粒体基因的已知变异谱, 其意义可能是不确定的. 尽管线粒体变异是典型的母系遗传, 它们也可以散发. 然而由于异质性可能低于试验检测水平或组织间的差异, 新发变异是难以评估的. 异质性水平可能是家族内表达差异和外显率降低的原因. 尽管如此, 异质性百分比和疾病严重程度之间仍缺乏相关性^[36]. 肌肉、肝脏或尿液可以作为附加样本类型用于临床评估. 未检测到的异质性也可能影响病例、病例对照和家系一致性研究的结果. 此外, 没有现成的功能研究方法, 尽管评估肌肉形态可能会有所帮助(即破碎红纤维的存在). 频率数据和已发表的证明因果关系的研究往往是检测报告上唯一的评估标准. 单倍群分析可以作为线粒体疾病的另一个工具, 但可能不是临床实验室已使用的常规方法, 而且临床相关性难以解释.

因为核基因变异也可能是氧化疾病的致病原因或起着调节线粒体变异的作用, 因此应考虑检测与线粒体疾病相关的核基因.

5.4 药物基因组学

确认基因变异在药物代谢中的作用具有挑战性,

部分原因在于其表型只有在接触药物后才得以显现. 不过, 临床上现已报告了各种与药物疗效和副作用风险相关的基因变异, 且其数量仍然在不断增加. 相关基因的汇总及其有临床意义的变异可查询药物基因组学知识库网站(<http://www.pharmgkb.org/>). 有关细胞色素 P450 基因家族等位基因及其命名可查询网站 <http://www.cypalleles.ki.se/>. 尽管解读药物基因组变异已超出了本文的范围, 还是对与解读及报告药物基因组结果相关的挑战和鉴别进行了讨论.

传统的药物基因组等位基因命名使用星号(*)标记等位基因, 通常用于表示单倍型或同一等位上基因变异的组合. 依据旧的参考序列的传统核苷酸编号规则仍然在使用. 而将传统命名转换为使用新参考序列的标准命名会是一项艰巨的任务, 但这对于下一代测序的信息学分析应用是必要的.

在药物基因组相关基因上已经鉴定了多种变异类型, 如截短、错义、缺失、重复(含功能及非功能等位基因)及基因转换, 它们可导致等位基因功能性或部分功能性丧失(功能减退或降低)以及无功能性(无效)等位基因. 解读序列变异常需依据由各种遗传变异组合成的单倍型信息. 单倍型一般根据人群频率和已知变异关联分析信息进行推定, 而非通过直接检测染色体片段(分子单倍型)来实现.

此外, 对于许多药物基因组基因(特别是酶编码基因变异), 其整体表型取决于二倍型, 即两个等位基因上的变异或单倍型组合. 由于药物基因组变异并不直接导致疾病, 使用代谢(快速、中等及减弱)、疗效(耐药、响应及敏感)或“风险”而非“致病”应更为恰当. 需要建立起在本领域保持一致性的专业术语和解读指南.

5.5 常见复杂疾病

与孟德尔疾病以家系为基础的研究不同, 常见复杂疾病(如 2 型糖尿病、冠心病和高血压)相关基因的鉴定, 在很大程度上依赖于以人群为基础的方法(如全基因组关联分析)^[37,38]. 目前, 大量的全基因组关联研究报告已对 1200 余种常见复杂疾病和性状的风险等位基因进行了编目. 然而, 这些变异大多数位于基因之间的区域, 尚需要进一步的研究来确定这些变异是否通过影响调控因子而直接导致疾病, 或者与致病变异处于连锁不平衡状态^[39].

常见复杂风险等位基因通常被赋予较低的相对

风险, 且预测能力薄弱^[40]。迄今为止, 常见复杂风险等位基因检测对于患者治疗的效用^[41]尚不清楚, 将多个指标组合起来进行累计风险评估的模型往往是有缺陷的, 通常并不优于家族史、人口统计资料和非遗传性临床表型等传统风险因素^[42,43]。另外, 在几乎所有的常见疾病中, 风险等位基因仅可解释至多10%的群体变异, 即使当疾病有高度遗传度时也是如此。考虑到问题的复杂性, 本建议并不涉及复杂性状的等位基因的解读和报告。然而我们认识到, 在对孟德尔基因进行测序时可以识别这些等位基因中的一部分, 因此需要有偶然发现这些等位基因时如何进行报告的指南。这种情况下, 术语“致病的”和“可能致病的”并不适用, 即使关联在统计学上是有效的。在建立更好的指南之前, 临时的解决办法是将这些变异报告为“风险等位基因”, 或在诊断报告中设立一个单独的“其他报告”类别。同病例对照/全基因组关联研究鉴定一样, 风险证据可以通过修改术语来表达, 如“确定的风险等位基因”, “可能的风险等位基因”或“不确定的风险等位基因”。

5.6 体细胞变异

体细胞变异主要见于癌细胞, 因为其等位基因比值高度可变, 且肿瘤异质性也可导致取样差异。在描述其变异时, 具有原发性变异所没有的复杂性。变异的解读有助于选择治疗方案和预测治疗效果、也应用于评估整体生存率或肿瘤无进展生存期, 因而体细胞变异的分类更加复杂。在对阴性结果解读时, 了解测序分析的检测方法局限性(变异可在何种等位基因频率时被检测到)至关重要, 此外也需要了解样本中肿瘤含量的特定信息。与胚系变异相比, 体细胞变异的分类类别也不同, 通常使用“敏感”、“拮抗”、“驱动”和“伴随”等术语。一个变异是否是体细胞变异需要通过患者胚系 DNA 的序列分析来证实。体细胞变异还需要另外的解读指南, 除了参考原发性突变的数据库以外, 还需要肿瘤特异性数据库作为参考。为了解决这个问题, 最近 AMP 已经成立了一个工作组。

6. 医务工作者如何使用这些指南和建议

临床实验室检测的主要目的是为医疗决策提供依据。在临床上, 基因检测一般用于识别或确认疾病的原因, 并帮助医务工作者做出个性化的治疗决策,

包括用药的选择。鉴于基因检测的复杂性, 检测过程中需相关医务工作者和临床实验室协作才能得到最佳结果。

当医务工作者提出基因检测需求时, 需将患者的临床信息提供给实验室。由于医务工作者越来越多地使用基因组(全外显子组或全基因组)测序, 而详细的临床信息有助于对检测结果的解读, 因此向实验室提供临床信息就变得越来越重要。例如, 当一个实验室在基因组测序样品中发现一个罕见或新发的变异时, 实验室负责人不能仅因为该变异是罕见的、新发现的或者新发的来确定它的致病性。该实验室必须通过患者的病史、家族史、体格检查和前期实验室检查对变异和基因进行评估, 进而区分致病变异和其他偶然(次要)发现或良性变异。事实上, 准确和完整的临床信息对于基因组水平 DNA 序列检测结果的解读是不可或缺的, 若待测样品不能提供此类信息, 实验室可以合理拒绝继续进行检测。

利用如高通量的靶向测序、全外显子组和全基因组测序等覆盖广泛表型的方法进行检测, 实验室可能会发现候选的致病变异。对医务工作者和患者后续的随访可能会发现更多的证据来支持某一变异的致病性。这些补充的表型信息可能是亚临床症状, 需要进一步完善相关的临床检测(例如, 一个在 *SLC26A4* 基因(与 Pendred 综合征相关的基因)上有不确定变异的听力受损患者, 需要进行 CT 检查判断其有无颞骨异常)。此外, 当发现一个变异可能是新发变异, 或者当一个变异在家系中与表型共分离, 或者在隐性遗传致病基因中一个变异与另一个致病变异处于反式位置时, 必须在其他家系成员中进行验证。在显性遗传性疾病的情况下, 在健康亲属中观察到的绝大部分变异可以被过滤或删减, 这样可使解读更加有效和准确。为此, 强烈建议在开展外显子组或基因组测序时, 尽力做到“核心家系”检测(即母亲、父亲、患病儿童), 尤其是对怀疑有隐性遗传或新发变异的患者。与成人患者相比, 这显然在儿科患者中更易实现。在没有父母一方或双方时, 纳入患病和正常的兄弟姐妹也是有意义的。

许多遗传变异会导致一系列表型(不同程度的表现度), 疾病发生的机率也可能不是 100%(外显率降低), 这些均进一步强调了向临床实验室提供全面的临床数据来帮助解读变异的重要性。在理想的情况下, 建议应依据医疗保险可携性和责任法案(HIPAA)

和机构审查委员会条例, 将临床数据存入并通过集中存储库共享. 重要的是, 当家庭成员的信息对于解读结果是必需的时候, 相关医务工作者可以进一步帮助临床实验室收集家庭成员的 DNA(例如, 当评估家系患者与疾病共分离时, 父母的基因型分析可用来评估新发变异的发生, 一级亲属可用来确定隐性遗传疾病变异的同线或异线性).

医务工作者如何使用基因检测提供的证据来进行医疗管理决策是一个关键问题. 目前变异分析是不完善的, 报道的变异分类也并不是 100% 确定的. 一般来说, 根据推荐的分类方法划分为致病性的变异符合经验数据形成的标准, 所以医务工作者可以在临床决策时采用分子检测信息. 应尽力避免使用此类信息作为孟德尔疾病的唯一证据, 在可能的情况下应与其他临床资料相结合. 通常情况下, 一个有足够的证据被划分为可能致病的变异, 当与可疑疾病的其他证据相结合时, 医务工作者可以使用分子检测信息进行临床决策的制定. 例如, 产前超声可能显示关键的证据, 对于产后的病例, 其他数据, 如酶检测、体格检查, 或影像学研究可能最终支持临床决策. 然而, 推荐进行所有如上所述的可能的后续检

测, 追踪可能致病变异相关的附加证据的产生, 因为这有可能将可能的致病性变异重新归类为致病变异. 意义不明确的变异不宜应用于临床决策. 应努力将变异分类为致病性或良性. 当变异的重新分类正在进行中时, 对可疑致病的患者进行额外的监测应审慎. 一个有足够证据被考虑为可能良性的变异, 医务工作者可以结合其他信息, 推断此变异不是该患者致病的原因, 例如, 变异并不与家族中的某位患病成员共分离, 而且也不太可能是复杂遗传模式. 一个有足够证据被考虑为良性的变异, 医务工作者可以得出此变异不是该患者致病原因的结论.

基因检测的证据如何使用也依赖于临床背景和检测指征. 在产前诊断的病例中, 如果该家庭正在考虑的决定将导致不可逆的后果时, 如宫内治疗或终止妊娠等, 需要在采取行动之前慎重考虑报告中证据的份量和胎儿超声等其他信息. 当基因检测结果是产前检查的唯一证据时, 需要向受检家庭慎重解释可能致病的变异. 关键是相关的医务工作者应与临床实验室深入沟通, 以了解所检测到的变异是如何被分类的, 以期为患者提供准确的遗传咨询和临床决策.

翻译声明 该指南是基于英文原版文章“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”翻译而来, 该英文文章已在 *Genetics in Medicine* 杂志发表, 版权归 *Genetics in Medicine* 所有. 该中文版本已被授权. 未经美国医学遗传学与基因组学学会(ACMG)书面许可, 此中、英文版本禁止复制和使用.

免责声明 ACMG 制定的标准与指南作为教育资源旨在帮助临床遗传学家提供优质的临床检验服务. 遵循该标准和指南属于自愿行为并且不一定能够确保一个成功的医疗结局. 该标准和指南并不囊括所有合适的流程和检测, 也不排斥其他可以获得相同结果的流程和检测. 临床实验室遗传学家应该利用自己的专业知识, 依据病人或样本的具体情况来判断某一具体的流程或检测的合理性. 我们鼓励临床实验室遗传学家记录对病人使用的某一具体流程或检测的原理, 不管这个原理与这些标准与指南是否符合. 同时建议临床实验室遗传学家关注指南的采用时间, 应考虑到此后更新的一些相关医疗和科学信息. 还需谨慎考虑到知识产权可能会限制某些检测或流程的使用.

致谢 感谢中国遗传学会遗传咨询分会的其他委员及参会专家对本指南中文版修订的大力支持(按姓氏拼音排序): 安锡培(中国遗传学会)、边旭明(北京协和医院)、陈嘉妮(美国 Oklahoma 大学医学中心)、陈万涛(上海交通大学附属第九人民医院)、段涛(上海市第一妇婴保健院)、傅启华(上海儿童医学中心)、顾东风(阜外医院)、关

明(复旦大学附属华山医院)、胡娅莉(南京大学医学院附属鼓楼医院)、黄卫东(新疆佳音医院)、黄宥芯(台北马偕纪念医院)、康熙雄(首都医科大学附属北京天坛医院)、李敏(上海交通大学医学院附属仁济医院)、李秋(重庆医科大学附属儿童医院)、李亦学(上海生物信息技术研究中心)、廖敏华(台湾瑞林生物科技)、凌秀凤(南京市妇幼保健院)、刘彩霞(中国医科大学附属盛京医院)、刘嘉茵(江苏省人民医院)、刘杰(复旦大学附属华山医院)、刘丽梅(上海交通大学附属第六医院)、卢大儒(复旦大学)、卢煜明(香港中文大学)、陆国辉(美国 Texas 大学 MD Anderson 肿瘤中心)、马端(复旦大学)、茅彩萍(苏州大学附属第一医院)、尚红(中国医科大学)、孙路明(上海市第一妇婴保健院)、孙筱放(广州医科大学附属第三医院)、王红艳(复旦大学)、王艺(复旦大学附属儿科医院)、王永进(国家卫生计生委能力建设和继续教育中心)、徐晨明(国际和平妇幼保健院)、徐丛剑(复旦大学附属妇产科医院)、杨爱平(国家卫生计生委能力建设和继续教育中心)、杨慧霞(北京大学第一医院)、张世琨(中国妇幼保健研究会)、张巍(美国 Baylor 医学院)、张学红(兰州大学第一医院)、张艳丽(临沂市妇幼保健院)、赵欣之(复旦大学附属儿科医院)、郑波(邢台不孕不育专科医院)、朱丽萍(上海市妇幼保健院)。

参考文献

- Richards C S, Bale S, Bellissimo D B, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*, 2008, 10: 294–300
- Plon S E, Eccles D M, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*, 2008, 29: 1282–1291
- Sosnay P R, Siklosi K R, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet*, 2013, 45: 1160–1167
- Thompson B A, Spurdle A B, Plazzer J P, et al. Application of a 5-tiered scheme for standardized classification of 2, 360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*, 2014, 46: 107–115
- MacArthur D G, Manolio T A, Dimmock D P, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*, 2014, 508: 469–476
- Kearney H M, Thorland E C, Brown K K, et al. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med*, 2011, 13: 680–685
- Wildeman M, van Ophuizen E, den Dunnen J T, et al. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat*, 2008, 29: 6–13
- Taschner P E, den Dunnen J T. Describing structural changes by extending HGVS sequence variation nomenclature. *Hum Mutat*, 2011, 32: 507–511
- Kircher M, Witten D M, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 2014, 46: 310–315
- Hicks S, Wheeler D A, Plon S E, et al. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat*, 2011, 32: 661–668
- Tavtigian S V, Greenblatt M S, Lesueur F, et al. *In silico* analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat*, 2008, 29: 1327–1336
- Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat*, 2011, 32: 358–368
- Thompson B A, Greenblatt M S, Vallee M P, et al. Calibration of multiple *in silico* tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat*, 2013, 34: 255–265
- Choi Y, Sims G E, Murphy S, et al. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, 2012, 7: e46688
- Adzhubei I A, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*, 2010, 7: 248–249
- Kumar P, Henikoff S, Ng P C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 2009, 4: 1073–1081
- Schwarz J M, Rodelsperger C, Schuelke M, et al. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, 2010, 7: 575–576
- Jian X, Boerwinkle E, Liu X. *In silico* tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med*, 2014, 16: 497–503

- 19 Houdayer C, Caux-Montcoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined *in silico/in vitro* studies on BRCA1 and BRCA2 variants. *Hum Mutat*, 2012, 33: 1228–1238
- 20 Vreeswijk M P, Kraan J N, van der Klift H M, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat*, 2009, 30: 107–114
- 21 Henderson L B, Applegate C D, Wohler E, et al. The impact of chromosomal microarray on clinical management: a retrospective analysis. *Genet Med*, 2014, 16: 657–664
- 22 Popp M W, Maquat L E. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet*, 2013, 47: 139–165
- 23 Bayrak-Toydemir P, McDonald J, Mao R, et al. Likelihood ratios to assess genetic evidence for clinical significance of uncertain variants: hereditary hemorrhagic telangiectasia as a model. *Exp Mol Pathol*, 2008, 85: 45–49
- 24 Thompson D, Easton D F, Goldgar D E. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet*, 2003, 73: 652–655
- 25 Goldgar D E, Easton D F, Deffenbaugh A M, et al. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet*, 2004, 75: 535–544
- 26 Scheuner M T, Hilborne L, Brown J, et al. A report template for molecular genetic tests designed to improve communication between the clinician and laboratory. *Genet Test Mol Biomarkers*, 2012, 16: 761–769
- 27 Lubin I M, Caggana M, Constantin C, et al. Ordering molecular genetic tests and reporting results: practices in laboratory and clinical settings. *J Mol Diagn*, 2008, 10: 459–468
- 28 Lubin I M, McGovern M M, Gibson Z, et al. Clinician perspectives about molecular genetic testing for heritable conditions and development of a clinician-friendly laboratory report. *J Mol Diagn*, 2009, 11: 162–171
- 29 Rehm H L, Bale S J, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*, 2013, 15: 733–747
- 30 Aziz N, Zhao Q, Bry L, et al. College of American Pathologists' Laboratory Standards for Next-Generation Sequencing Clinical Tests. *Arch Pathol Lab Med*, 2014
- 31 Aronson S J, Clark E H, Varugheese M, et al. Communicating new knowledge on previously reported genetic variants. *Genet Med*, 2012
- 32 Bean L J, Tinker S W, da Silva C, et al. Free the data: one laboratory's approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. *Hum Mutat*, 2013, 34: 1183–1188
- 33 Hirschhorn K, Fleischer L D, Godmilow L, et al. Duty to re-contact. *Genet Med*, 1999, 1: 171–172
- 34 Behar D M, Oven M V, Rosset S, et al. A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet*, 2012, 90: 675–684
- 35 van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*, 2009, 30: E386–E394
- 36 Wong L J. Diagnostic challenges of mitochondrial DNA disorders *Mitochondrion*. 2007, 7: 45–52
- 37 Manolio T A. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 2010, 363: 166–176
- 38 Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med*, 2009, 360: 1759–1768
- 39 Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*, 1999, 22: 139–144
- 40 McClellan J, King M C. Genetic heterogeneity in human disease. *Cell*, 2010, 141: 210–217
- 41 Topol E J, Damani S B. The KIF6 collapse. *J Am Coll Cardiol*. 2010, 56: 1564–1566
- 42 Meigs J B, Shrader P, Sullivan L M, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med*, 2008, 359: 2208–2219
- 43 Paynter N P, Pare G, Buring J E, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA*, 2010, 303: 631–637