

编者按

全球罕见病约 7000 多种,有 3 亿多罕见病患者,我国现有各类罕见病患者 2000 余万人。罕见病常累和多系统多器官、临床表现复杂、遗传因素多见,具有诊断难、治疗难和诊疗同质性差的特点。继去年国家《第二批罕见病目录》出台后,今年“两会”指出“加强罕见病研究、诊疗服务和用药保障”,我国罕见病防治与保障工作已成为医药领域发展新质生产力的重要一环。为进一步提高对罕见病的认识、推动罕见病研究,本期立足罕见病学科前沿,聚焦难题热点,由四川省人民医院罕见病医学中心负责人谈颂教授主持,邀请全国深耕罕见病领域多年的知名专家学者撰写专题文章进行研讨,为罕见病的诊疗和科研工作提供参考。

基因意义不明变异的应对策略

马端

复旦大学代谢分子医学教育部重点实验室,复旦大学医学遗传研究院,上海 200032

【摘要】 美国医学遗传学与基因组学学会将基因变异分为致病、可能致病、良性、可能良性和意义不明(VUS)。VUS 在基因检测与分析中甚为常见,为临床遗传病因分析带来困扰。无论是基因位点变异,还是其它的序列变异,凡是遇到 VUS,可以从变异位点的准确性、遗传模式、人群中点变异频率、更多同样疾病和同种基因变异家系、氨基酸改变、基因调控、生物信息学、蛋白质功能、携带突变基因的细胞表型和动物表型等方面进行分析。我们研发的人工智能基因分析系统有助于提升基因分析的效率和准确性。

【关键词】 基因;意义不明变异;应对策略;人工智能。

【中图分类号】 R394

【文献标志码】 A

【文章编号】 1672-6170(2024)05-0001-05

Strategies for dealing with genetic variants uncertain significance MA Duan *Key Laboratory of Metabolism and Molecular Medicine of Ministry of Education, Institute of Medical Genetics & Genomics, Fudan University, Shanghai 200032, China*

【Abstract】 Genetic variants can be classified into pathogenic variants, probably pathogenic variants, benign variants, probably benign variants, and variants uncertain significance (VUS) according to American College of Medical Genetics and Genomics (ACMG). VUS is very common in genetic testing and analysis. It brings problems to clinical genetic etiology analysis. Whether it is gene locus variation or other sequence variation, VUS can be analyzed from the accuracy of the variation locus, the inheritance pattern, and the frequency of point mutations in the population. It can also be analyzed from more families with the same disease and the same gene variation, amino acid changes, gene regulation, bioinformatics, protein function, cell and animal phenotypes carrying mutated genes whenever you encounter VUS. Our AI Genetic Analysis System can improve the efficiency and accuracy of genetic analysis.

【Key words】 Gene; Variants uncertain significance; Analytical strategies; Artificial intelligence

在基因检测应用越来越多的情况下,对基因变异的解读显得越来越重要。根据美国医学遗传学与基因组学学会(American college of medical genetics and genomics, ACMG)标准与指南对基因变异的评级,可将基因变异分为五个级别:致病,可能致病、良性、可能良性、意义不明(variants uncertain significance, VUS)^[1]。前 4 种级别易于理解,唯独 VUS 让人头疼。当一种基因变异无法确定其临床意义时,就将其称为 VUS。根据 ClinVar Miner 数据

最新统计(2024 年 6 月 30 号),VUS 比其它四个级别的总和还要多。见表 1。随之而来的问题是:遇到 VUS 该怎么办? 本文收集整理相关文献,结合本科研团队的实际经验,提出 VUS 的应对策略,供大家参考。

表 1 人类基因组中各种变异的数量^[2]

级别	变异数量	基因数量
VUS	1531969	38476
可能良性	1033394	21718
良性	283534	16298
致病	229030	14925
可能致病	138517	7212

1 基因变异的主要类型

基因变异分为序列变异和修饰变异,本文仅讨论序列变异。根据变异大小的不同,可以分为点变

【基金项目】 国家重点研发计划(编号:2021YFC270101, 2021YFC2700803)

【作者简介】 马端,男,博士,教授。中华医学会医学遗传学分会副主任委员,全国卫管协会精准医疗分会副会长,上海市医学会罕见病分会主任委员,上海市健康科技协会基因健康专委会主任委员。主要研究方向:遗传相关疾病的发病机制和防治策略。

异(Point Variant)、插入缺失(Indel)、拷贝数变异(CNV)、染色体结构变异和染色体数量变异(图1)。

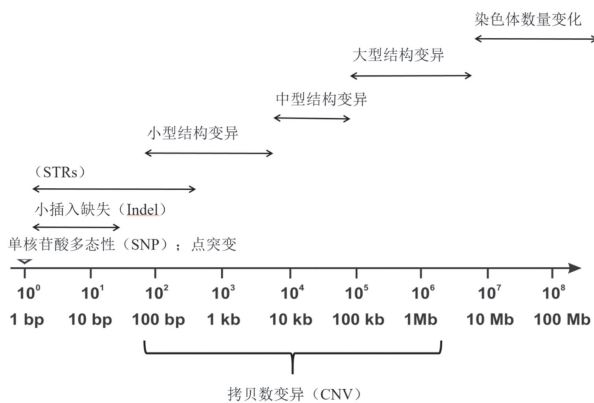


图1 基因在染色体上不同程度的序列变化

1.1 点变异 一个单核苷酸(常称为碱基)变为其它单核苷酸称为点变异。如果变异的频率 $>1\%$,称为单核苷酸多态性(SNP)。如果变异频率 $<1\%$,且证明对基因功能产生影响,则称为点突变(Point Mutation)。点突变如果出现在蛋白质编码基因的编码区,根据是否影响氨基酸及蛋白质序列或活性的变化,分为同义突变、错义突变、无义突变和移码突变。点突变如果出现在非编码区,则要分析是否对基因的转录产生影响。

1.2 插入缺失 指在基因的某个位置上插入或删除1个或多个碱基。在编码序列中出现的插入缺失,可以导致1个或数个氨基酸的变异,也可以造成翻译提前终止,出现截断的多肽或蛋白。出现在调控区的插入缺失,可能影响基因转录水平。

1.3 拷贝数变异(CNV) 编码基因多为单拷贝,少部分为多拷贝。一段序列的拷贝数多于或少于正常,称为CNV。CNV的长度从50 bp到数个Mb^[3]。如果CNV位于基因的编码序列,会影响编码基因的结构及数量;如果CNV位于非编码区,则

可能影响基因的转录水平。

1.4 染色体结构变异 当染色体出现大片的插入、缺失、倒位或易位时,称为染色体结构变异。结构变异所带来的不良结局,与发生变异的部位及大小有关。含有重要基因及调控区的结构变异,往往会使基因数量和功能出现异常。

1.5 染色体数量异常 正常人有22对常染色体和1对性染色体,任何一条染色体出现增多或减少,称为染色体数量异常,如21-三体、18-三体、13-三体等。

2 检测基因变异的主要方法

2.1 Sanger 测序 也称一代测序。可以准确检测600~1000 bp以内的序列是否出现异常,准确率为99.999%,是基因检测的金标准。缺点是检测的序列过短,检测通量低。主要用于目标明确的变异位点检测及筛查发现的可疑变异位点验证。

2.2 二代测序 又称下一代测序。根据检测范围的需求,可将全基因组、全外显子组、任何可捕获的基因组序列打碎成25~400 bp的片段,逐个多重测序后再行拼接,用软件分析序列是否正常。检测的准确率 $>99\%$,但对发现的变异仍然需要进行Sanger测序验证。二代测序通量高,价格逐步下降,已常规用于基因变异的筛查。缺点是测序读长短,难以发现大片段缺失,会出现拼接错误。

2.3 三代测序 三代测序又称长片段单分子实时测序,优点是测序读长往往 >1 kb,可以发现二代测序无法发现的大片段缺失和拼接异常,缺点是价格高,准确性稍差,正在不断改进中。

2.4 四代测序 四代测序又称纳米孔电流测序,测序长度大于150 kb,设备体积小,但目前处在发展阶段,需大力提升准确度。

不同测序技术的特点见表2。

表2 不同基因测序技术的特点

测序技术	测序原理	序列长度	通量	准确率	缺点
一代测序	Sanger毛细管电泳测序	600~1000 bp	0.0002 Gb/run	99.999%	通量低
二代测序	合成测序、连续测序和焦磷酸测序	50~400 bp	0.45~1500 Gb/run	$>99\%$	读长短
三代测序	DNA单分子实时测序	1000~100000 bp	0.5~9 Gb/run	$>85\%$	准确度稍低
四代测序	纳米孔电流测序	>150 kb	30~400 bp/秒	$>70\%$	准确度较低

3 VUS的主要应对策略

基因变异经过ACMG标准打分后,如果出现VUS,可以从以下几个方面进行判断,最终将VUS升级为致病或可能致病,或者降级为良性变异或可能良性变异。

3.1 确定变异位点的准确性与遗传模式 ①对于

高度怀疑的VUS,必须用Sanger测序确定变异位点的准确性;②确定变异位点是否与疾病表型共分离:在先证者家系中,患者均应携带同样的变异位点,健康亲属则不携带。分析时需要注意迟发型疾病及与性别相关的疾病,即某些“健康”亲属携带了同样变异位点,但未到发病年龄,所以未发病;某些

疾病仅见或更多见于某一性别。

3.2 数据分析 ①在人群数据库中查询位点变异频率:人群数据库会统计已检测的位点变异频率。变异位点的频率越低,致病的可能性越大。表 3 列出了常用的人群数据库及特征。需要注意的是不同国家或民族的变异频率存在差异。②寻找更多同样疾病和同种基因变异家系:当一种变异无法在人群数据库中查询变异频率时,则需通过文献查询或同行间交流,力图发现更多的类似疾病家系。当同一变异出现越来越多的疾病家系时,则可升级 VUS。反之,如果同样的变异在越来越多的正常人群中出现,则需降级该 VUS。③氨基酸改变性质分

析:编码区的位点变异可导致氨基酸改变、氨基酸增加或减少。一般而言,碱性氨基酸和酸性氨基酸的相互改变,极性氨基酸与非极性氨基酸的相互改变,以及性质差别较大的氨基酸相互改变,很有可能影响所在多肽或蛋白质的功能。如果改变的氨基酸出现在酶活性中心或重要的蛋白质结构域,则有可能影响酶的活性及相关蛋白质的功能。④基因调控序列变异分析:人类的每个基因都有调控区,包括基因 5' 的启动子、3' 的 miRNA 结合区域及 PolyA 长度、基因两侧的增强子/沉默子和绝缘子等。当变异发生在这些区域,需分析对基因表达水平的影响。

表 3 常见的人群数据库及特征

数据库名称	数据库描述
Genome Aggregation Database(gnomAD) https://gnomad.broadinstitute.org/	gnomAD 含有全球各种族群的基因组变异信息。gnomAD Exomes 主要保存外显子组的遗传变异数据。在最新的版本中,gnomAD 数据库包含了 730947 个个体的外显子组序列信息,以及 76215 个个体的全基因组测序信息。
1000Genomes, 千人基因组 https://www.internationalgenome.org/	1 千多名基因提供者来自欧洲、美洲、亚洲、非洲的 14 个国家和地区。绘制了一个详尽的基因图谱以供比对。
Exome Aggregation Consortium, ExAC https://gnomad.broadinstitute.org/	ExAC 中包含 60706 个个体的外显子组信息,这些信息已经被整合到了 gnomAD 数据库中。
NHLBI Exome Sequencing Project https://jmorp.megabank.tohoku.ac.jp/	收录了 6503 个个体的外显子组信息
58KJPN https://jmorp.megabank.tohoku.ac.jp/	包含了约 58000 名日本人全基因组范围的等位基因频率信息
UK10K https://www.genomeasia100k.org/	收录了超过 10000 个个体的基因组,其中 4000 名个体拥有全基因组信息,以及神经发育问题样本组(3000 人全外显子组信息),肥胖问题组(2000 人全外显子组信息),罕见疾病组(1000 人全外显子组信息)。
Genome Asia https://www.genomeasia100k.org/	超过 100000 个个体的全基因组信息规模的亚洲基因组信息蓝图。
Arquivo Brasileiro Online de Mutações https://abraom.ib.usp.br/	对来自巴西不同地理和人口群体的个体进行全基因组测序,囊括了 1171 个个体的全基因组测序数据和 609 个个体的外显子组测序数据。
TOPMedBravo https://bravo.sph.umich.edu/freeze8/hg38/	已收录了 55 个心、肺、血液和睡眠障碍领域的全基因组测序数据,包含了 132345 名来自全球的个体。
GME Variome https://www.nature.com/articles/ng.3592	已经收录了 1111 名来自中东地区个体的全外显子组测序信息
Westlake BioBank for Chinese https://wbcc.westlake.edu.cn/	西湖大学中国生物银行已经拥有了超过 4000 名中国人的全基因组测序信息

3.3 生物信息学分析 随着基因测序应用的不断增加,为了快速判断变异位点的性质,生物信息学分析软件也不断增多。这些软件可能主要针对某一种变异类型进行分析,也可能针对多种变异进行

分析。分析的依据来自于既往的功能研究、进化保守性和大数据整合的结果。表 4 列出了部分常用的基因变异生物信息学分析软件及依据。

表 4 部分基因变异生物信息学分析软件及依据

分类	名称	网站	依据
错义突变预测	FATHMM	http://fathmm.biocompute.org.uk	进化保守性
	SIFI	https://sift.bii.a-star.edu.sg/www/code.html	进化保守性

分类	名称	网站	依据
	CADD	https://github.com/kircherlab/CADD-scripts	对于来自模拟变异的等位基因进行不同的注释
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	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
	AlphaMissense	https://github.com/google-deepmind/alphamissense	基于 AlphaFold 模型的无监督蛋白质语言建模
	VEST4	http://hg19.craivat.us/CRAVAT/	基于随机森林的监督学习分类器
	REVEL	https://sites.google.com/site/revelgenomics/	基于随机森林的多工具集成方法
	MPC	ftp://ftp.broadinstitute.org/pub/ExAC_release/release1/regional_missense_constraint/	基于逻辑回归分析的预测方法
	PrimateAI	https://github.com/Illumina/PrimateAI	基于深度残差网络的预测方法
插入和缺失预测	BayesDel	https://fenglab.chpc.utah.edu/BayesDel/BayesDel.html	基于贝叶斯方法的多工具加权乘积评分
核酸保守性预测	GREP++	http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html	基因组进化速率分析
剪接预测	dbscSNV	http://www.liulab.science/dbscsnv.html	基于集成学习的预测模型
	SpliceAI	https://github.com/Illumina/SpliceAI	卷积神经网络通过序列信息预测剪接位点

3.4 功能分析 如果基因变异发生在已知功能的蛋白质编码基因中,则可分离患者病变组织中相应的细胞,或在病变涉及的细胞模型中构建基因突变体,检测蛋白质功能是否发生变化,如酶活性、转录因子活性、剪接序列变化、蛋白质定位改变和基因表达水平变化等。对于一个具有分子生物学和细胞生物学实验水平的医院而言,这些实验比较简单,可以在短时间内获得实验结果。

3.5 功能研究 对于功能不明确的基因,上述所有的方法都不足以判断基因变异的结果时,则需要花费更多的时间判断基因变异的性质。以下的功能研究可以根据具体情况选择使用。

3.5.1 细胞层面的功能研究 判断 VUS 相对简单的方法是构建基因突变体,转入与疾病相关的细胞,观察细胞的表型及功能变化。基因突变体构建的方法有多种,其中 CRISPR-Cas9 最为方便^[4]。如果 CRISPR-Cas9 无法完成特点位点的突变,也可以使用 PCR 介导的定点突变^[5]、寡核苷酸介导的定点突变^[6]或单碱基编辑^[7]等方法实现突变。基因突变构建成功后,可以通过病毒载体或非病毒载体将其转入细胞。通过观察细胞增殖能力、细胞迁移和侵袭能力、细胞死亡形式(如死亡、凋亡、焦亡和铁死亡等)、细胞分化能力和细胞自噬能力的变化,确定突变体的功能。对于大基因而言,突变的位点甚多,甚至超过千种。对于此类情况,对 VUS 一一进行独立研究,不仅耗时,成本亦高。贾小彦博士等人发明的饱和突变法可以对一个基因的所有变异

进行高通量判断^[8],由此构建的基因突变功能图谱不仅可以用于现有 VUS 的判断,也可用于未来发现的 VUS 评估。

3.5.2 动物层面的功能研究 当 VUS 出现在重要基因中且机制不明时,仅做细胞层面的研究不足以得出结论,此时构建基因突变动物模型,观察动物表型及相应器官、组织和细胞的形态及功能改变,与患者表型进行对照,由此判断 VUS 的性质。目前常用的动物模型有线虫、果蝇、斑马鱼、小鼠、大鼠和猴。通过在干细胞和生殖细胞层面构建突变体,可以实现突变基因的整体或局部表达。如果子代可以存活,通过观察整体的表型及各个脏器的结构及功能,判断 VUS 的性质。如果子代无法存活,则可以通过观察胚胎的发育及各脏器的结构,判断 VUS 的性质。

4 基因变异分析的智能化解决方案

由于基因的数量多,各类变异更是浩如烟海,即使借助分析软件,出具一份全外显子组基因检测(WES)报告往往要花费数天,所以从取样到拿到结果少则一周,多则一个月以上。即使拿到了报告,许多临床医生对报告的解读能力仍然有限,使得基因检测结果只能在少数专家手中才能发挥防治疾病作用。

基于这种情况,我们团队研发了人工智能“医学遗传分析与咨询系统”,简称 AIGene(图 1)。该系统可以实现自动表型录入、家系图谱绘制、Fastq 文件分析、VCF 文件分析、ACMG 自动评分、基因变

异与疾病关系的自动排序、个性化药物选择和人工智能遗传咨询,既可以分析 WGS 和 WES,也可以分析各种 PANEL 和单基因测序数据,包括对 CNV 和线粒体基因组的分析。系统可以针对 2 万多条蛋白质编码基因、1 万多种遗传性疾病、近 500 种药物基因组数据、1 万 7 千多种表型和 38 亿多个遗传变异位点进行全解读,并拥有所有上述数据的知识

库。使用者在完成表型输入和检测文件的上传后,可以在 2~10 分钟获取分析的结果,并可以根据自己的判断进行选择 and 修正。该系统目前已经在全国 10 多家三级甲等医院进行试用,不久将全面推广至临床和第三方基因检测机构,面对普通大众的 APP 版本也即将上线。



图 1 人工智能医学遗传分析与咨询系统

总之,随着基因检测应用的急剧增多,对基因变异的性质判断尤其重要。基因检测不仅可以用于遗传性疾病的病因诊断,也可以为个性化用药、疾病预警和预后判断、辅助生殖等提供有价值的参考。随着解读能力的增强,VUS 也不再成为临床应用过程中不可逾越的障碍。

【参考文献】

- [1] Richards S, Aziz N, Bale S, et al. 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 [J]. *Genet Med*, 2015, 17 (5):405-424.
- [2] <https://clinvarminer.genetics.etch.edu>
- [3] Pös O, Radvanszky J, Buglyó G, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects [J]. *Biomed J*, 2021, 44 (5):548-559.
- [4] Irfan M, Majeed H, Ifikhar T, et al. A review on molecular scissors with CRISPR/Cas9 genome editing technology [J]. *Toxicol Res (Camb)*, 2024, 13 (4):tfae105.
- [5] Carey MF, Peterson CL, Smale ST. PCR-mediated site-directed mutagenesis [J]. *Cold Spring Harb Protoc*, 2013, 2013 (8):738-742.
- [6] Luan J, Song C, Liu Y, et al. Seamless site-directed mutagenesis in complex cloned DNA sequences using the RedEx method [J]. *Nat Protoc*, 2024, Online ahead of print.
- [7] Liang Y, Chen F, Wang K, et al. Base editors: development and applications in biomedicine [J]. *Front Med*, 2023, 17 (3):359-387.
- [8] Jia X, Burugula BB, Chen V, et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk [J]. *Am J Hum Genet*, 2021, 108 (1):163-175.

(收稿日期:2024-07-31;修回日期:2024-08-02)

(本文编辑:林 赞)