·专论·

人类孟德尔遗传性疾病基因组序列变异解析与临床规范

唐北沙 曾胜 李凯

【关键词】 遗传性疾病,先天性; 基因; 突变; 综述
【Key words】 Genetic diseases, inborn; Genes; Mutation; Review

Clinical standards and interpretation of gene sequence variants in human Mendelian disorders

TANG Bei-sha[1,2], ZENG Sheng[1], LI Kai[1]
[1]*Department of Neurology, Xiangya Hospital,* [2]*State Key Laboratory of Medical Genetics, National Clinical Research Center for Geriatric Diseases, Central South University, Changsha 410008, Hu´nan, China*
Corresponding author: TANG Bei-sha (Email: bstang7398@163.com)
This study was supported by Key Project of the National Natural Science Foundation of China (No. 81130021).

随着二代基因测序(NGS)技术的不断完善,其在临床应用和研究逐渐普及,越来越多的科研或医疗机构开始应用该项技术[主要包括全基因组测序(WGS)、全外显子测序(WES)、目标区域捕获测序]进行人类孟德尔遗传性疾病的分子诊断和遗传学研究[1-4]。临床实践中,基因组检测流程需规范化、基因组序列变异判断需标准化、测序技术需严格质控、具体测序技术需合理选择[5-7]。人类基因组全外显子组水平约包含25×10^3个变异(variants)[8],如何精准检测这些变异、筛选出致病性突变,是医学遗传学必须面对的问题。鉴于此,美国医学遗传学和基因组学会(ACMG)、欧洲人类遗传学会(ESHG)分别公布二代基因测序的临床应用指南[4,9-10]。因此,根据我国实际情况制定人类孟德尔遗传性疾病基因组序列变异解析与临床规范势在必行。本文仅针对人类基因组DNA序列,而线粒体DNA序列和表观遗传领域RNA序列、甲基化等不在本文阐述范围。本文拟从临床资料采集、遗传因素判断、二代基因测序选择、质控管理、序列变异检测及公共数据库过滤、序列变异生物信息学分析、遗传学和功能学试验、序列变异解析原则、伦理学和遗传咨询方面进行阐述。

一、翔实的临床资料采集

翔实的临床资料采集是进行分子诊断和遗传学研究的基础。完善的临床资料可以有效降低临床诊断和分子诊断的误诊率,有助于变异解析的后续分析[1,4],主要包括主诉、现病史、家族史、近亲婚配史、体格检查、实验室检验、量表评价和影像学检查等。

二、遗传因素判断

根据临床资料判断疾病是否系遗传因素所致以及是否符合孟德尔遗传规律,包括常染色体显性遗传(AD)、常染色体隐性遗传(AR)和X连锁遗传(X-linked)。随着对某些疾病的深入认识,某些罕见疾病(Joubert综合征等)和先天性疾病(先天性无痛症等)也受到遗传因素的影响,也可以选择二代基因测序技术进行分子诊断[10-11]。

三、合理选择二代基因测序技术

确定遗传因素在疾病发病中发挥主要作用后,制定合理的二代基因测序方案、选择适宜的检测疾病遗传结构变异(genetic architecture)的二代基因测序技术和数据分析方法,是提高分子诊断率的先决条件。基因组序列变异包括以下5种形式[12-13]:单

doi:10.3969/j.issn.1672-6731.2017.07.001
基金项目:国家自然科学基金重点资助项目(项目编号:81130021)
作者单位:410008 长沙,中南大学湘雅医院神经内科(唐北沙、曾胜、李凯),医学遗传学国家重点实验室 国家老年疾病临床医学研究中心(唐北沙)
通讯作者:唐北沙(Email:bstang7398@163.com)

核苷酸变异(SNV)、插入/缺失变异、拷贝数变异(CNV)、短串联重复序列(STR)和结构变异(SV);以及以下4个部位:基因组外显子区(exonic regions)、基因组基因间区(intergenic regions)、基因组内含子区(intronic regions)及基因组启动子区(promoter regions)和非翻译区(UTR)。在选择二代基因测序技术时,应考虑每种测序技术的特点和局限性:发生于基因组外显子区的单核苷酸变异、插入/缺失变异,可以选择全外显子测序;发生于全基因组的单核苷酸变异、插入/缺失变异、拷贝数变异可以选择全基因组测序;发生于全基因组的短串联重复变异,既不能选择全外显子测序也不能选择全基因组测序[14-15]。随着基因检测技术的发展,三代基因测序技术逐渐广泛应用,有望实现全基因组短串联重复变异和复杂结构变异的检测[16-18]。

四、严格规范质控管理

在基因检测方案和数据分析方法合理的情况下,对整套基因检测流程进行严格质控是进行变异解析后续分析的有力保证[6]。首先,应确保检测样本的DNA质量并准确标记;其次,应保证检测样本的建库质量;再次,应采用合格的目标区域捕获测序试剂和设备,并严格按照操作流程进行,以避免人为操作造成的失误[19];最后,应选择正规的检测机构和实验室。数据分析包括以下步骤:(1)对基因检测所获得的原始数据(raw data)进行基本质检,如测序质量检测软件FastQC、评价测序准确性的碱基质量值(Q30代表质量值为30时错误识别率为0.1%)、鸟嘌呤-胞嘧啶(GC)含量、数据产量等,再通过剔除接头和低质量数据将原始数据转换为有效数据。(2)采用读长(reads)比对率、测序平均覆盖深度、测序深度分布、目标区域覆盖率(如基因组外显子区测序深度>10×的百分比等)评价数据质量。(3)采用比对软件(如BWA软件,https://sourceforge.net/projects/bio-bwa/files/)进行比对,并通过一种或多种检测软件对序列变异进行检测和注释。(4)通过比对检测样本单核苷酸变异与单核苷酸多态性(SNP)数据库(https://www.ncbi.nlm.nih.gov/projects/SNP/)中单核苷酸变异比值以及转换/颠换比值等评价变异提取过程的生物信息学分析质量[19-20]。

五、序列变异检测及公共数据库过滤

二代基因测序技术的生物信息学分析软件主要用于数据质控、参考基因组比对、变异检测、变异注释等。应注意不同生物信息学分析软件各有优缺点[4]:若检测结果中无足够候选变异,应进一步增加候选变异,可考虑采用不同序列变异检测软件,如GATK(https://software.broadinstitute.org/gatk/)、SAMtools(http://www.htslib.org/)、SOAPsnp(http://soap.genomics.org.cn/soapsnp.html)等,或更新变异注释软件,如更新ANNOVAR软件版本(http://www.openbioinformatics.org/annovar/annova_download_form.php)重新提取变异。对于人类孟德尔遗传性疾病,考虑其发病率低,进行公共数据库过滤时多以少数等位基因频率(MAF)<0.1%作为显性遗传性疾病限定值[14],但可能导致假阴性结果[21-22]。随着精准医疗(PM)的开展,临床信息完整并可长期随访的人群队列基因组数据将不断产生,可以有效解决现有数据库临床信息不足的问题。

六、序列变异的生物信息学分析

不同生物信息学分析软件预测致病性突变的方法各不相同,主要包括GERP++(http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html)、PhyloP(http://compgen.bscb.cornell.edu/phast/)、SIFT(http://sift.jcvi.org)、PolyPhen-2(http://genetics.bwh.harvard.edu/pph2)、Mutation Taster(http://www.mutationtaster.org)、CADD(http://cadd.gs.washington.edu)等,其中,GERP++、PhyloP和SIFT软件用于评价序列变异的保守性,PolyPhen-2软件用于评价氨基酸和蛋白质结构改变,Mutation Taster和CADD软件用于评价变异功能[23]。值得注意的是,预测致病性变异位点时,应避免仅采用一种预测方法的结果,亦应避免将多种预测方法的每种结果作为独立支持证据而累加。

七、遗传学和功能学试验

二代基因测序技术检出的变异可能存在假阳性结果,应采用Sanger测序验证。同时,对筛选出的候选变异位点,应在家系其他成员中进行共分离验证。对于已知致病基因的新发变异(novel variants),可采用功能学试验补充遗传学和生物信息学分析。功能学试验是否合理主要取决于所选取的功能模型是否适用于该疾病。可以根据具体情况进行自身组织和(或)细胞的功能学试验,或者建立体内或体外模型进行功能学试验[10]。

八、序列变异的解析原则

人类孟德尔遗传性疾病序列变异解析原则主要包括:(1)按照5级分类原则进行变异解析,根据基因组序列变异类型、数据库信息等将序列变异分

表1 序列变异致病性证据分层[4,24]

Table 1. Criteria for classifying pathogenic variants[4,24]

Evidence	Category	Instructions
Very strong	PVS1: null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multiexon, deletion) in a gene where LOF is a known mechanism of disease	Beware of genes where LOF is not a known disease mechanism (eg., GFAP, MYH7) Use caution interpreting LOF variants at the extreme 3′ end of a gene Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact Use caution in the presence of multiple transcripts Do not apply to variants that are near the 3′ end of the gene and escape nonsense-mediated decay
Strong	PS1: same amino acid change as a previously established pathogenic variant regardless of nucleotide change, for example: Val → Leu caused by either G > C or G > T in the same codon	Beware of changes that impact splicing rather than at the amino acid/protein level Does not include the same variant being assessed because it is not yet pathogenic, and the rule is intended for variants with a different nucleotide change
	PS2: de novo (both maternity and paternity confirmed) in a patient with the disease and no family history	Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity Apply this rule as moderate or supporting if the variant is mosaic and its frequency in tissue is consistent with the phenotype
	PS3: well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product	Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established Reduce the strength for assays that are not as well validated or linked to the phenotype
	PS4: the prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls	*RR* or *OR*, as obtained from case-control studies, is > 5.000, and the confidence interval around the estimate of *RR* or *OR* does not include 1.000. See the article for detailed guidance In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidenc
Moderate	PM1: located in a mutational hot spot and/or critical and well-established functional domain (eg., active site of an enzyme) without benign variation	Not meant for truncations; more clarification is needed for applying this rule
	PM2: absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project or Exome Aggregation Consortium	Population data for insertions/deletions may be poorly called by NGS Cannot assume longer indels would be detected by next-generation sequencing Use a published control dataset if its size is at least 1000 individuals Can not be applied for low-quality calls or non-covered regions Must define the condition and inheritance pattern
	PM3: for recessive disorders, detected in trans with a pathogenic variant	This requires testing of parents (or offspring) to determine phase Invoke this rule as supporting if the phase is not established Can upgrade if more than one proband is reported
	PM4: protein length changes as a result of in-frame insertions/deletions in a nonrepeat region or stop-loss variants	Applicable for in-frame insertions/deletions or stop-loss variants, but not frameshifts, nonsense and splice variants
	PM5: novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before	For example: p.Arg156His is pathogenic; now you observe p.Arg156Cys Beware of changes that impact splicing rather than at the amino acid/protein level Ensure pathogenicity of previously reported variant Suggest changing "novel" to "different" because some variants that are not novel might require assessment with this rule
	PM6: assumed de novo, but without confirmation of paternity and maternity	
Supporting	PP1: cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease	May be used as stronger evidence with increasing segregation data
	PP2: missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease	
	PP3: multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)	Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant all lines must agree
	PP4: patient′s phenotype or family history is highly specific for a disease with a single genetic etiology	Not meant to be used for genetically heterogeneous conditions or conditions with unsolved etiology Not typically applied for an analysis of incidental findings, but it could be applied for prior observations
	PP5: reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation	Only applicable when evidence is not available (eg., Sharing Clinical Reports Project)

PVS, pathogenic very strong, 致病性极强; PS, pathogenic strong, 致病性强; PM, pathogenic moderate, 致病性中度; PP, pathogenic supporting, 致病性支持; LOF, loss of function, 功能缺失; GFAP, glial fibrillary acidic protein, 胶质纤维酸性蛋白; MYH7, myosin heavy chain 7, 肌球蛋白重链7; RR, relative risk, 相对危险度; OR, odds ratio, 比值比; NGS, next-generation sequencing, 二代基因测序

为5级，即致病性(pathogenic)、可能致病性(likely pathogenic)、意义不明(uncertain significance)、可能良性(likely benign)和良性(benign)。(2)按照4级分类或3级分类原则进行变异解析，根据序列变异类型、数据库信息等将致病性突变证据分为4级，即非常强、强、中度和支持(表1)[4,24]；将良性突变证据分

表2　序列变异良性证据分层[4,24]

Table 2.　Criteria for classifying benign variants[4,24]

Evidence	Category	Instructions
Stand-alone	BA1: allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project or Exome Aggregation Consortium	
Strong	BS1: allele frequency is greater than expected for disorder	
	BS2: observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous) or X-linked (hemizygous) disorder, with full penetrance expected at an early age	Populations might not have been screened or excluded for the phenotype
	BS3: well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing	
	BS4: lack of segregation in affected members of a family	The presence of phenocopies for common phenotypes (ie., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation
Supporting	BP1: missense variant in a gene for which primarily truncating variants are known to cause disease	Clarify the meaning of "primary"; suggest >90%
	BP2: observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern	Clarify that one should apply BP2 when the pathogenic variant is seen in the same gene as the variant being evaluated and apply BP5 when the pathogenic variant is in a different gene
	BP3: in-frame insertions/deletions in a repetitive region without a known function	
	BP4: multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)	Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant All lines must agree
	BP5: variant found in a case with an alternate molecular basis for disease	Clarify that one should apply BP2 when the pathogenic variant is seen in the same gene as the variant being evaluated and apply BP5 when the pathogenic variant is in a different gene
	BP6: reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation	
	BP7: a synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site and the nucleotide is not highly conserved	

BA，benign stand-alone，良性独立；BS，benign strong，良性强；BP，benign supporting，良性支持

为3级，即独立、强和支持（表2）[4,24]。（3）按照序列变异致病性或良性证据累加作用原则进行变异解析，通过致病性或良性证据累加作用以判断序列变异是致病性、可能致病性、可能良性或良性，若不符合上述标准或致病性证据与良性证据相互矛盾，则判断为意义不明（表3）[4]。值得注意的是，首先变异解析的5级分类原则本质上是致病性概率的判断，“可能（likely）”用于具有90%以上确定的可能致病性或可能良性；其次，旨在鉴定疾病新候选致病基因的情况并不适用于该序列变异解析原则；再次，由于大样本人群变异数据库的发展导致变异证据改变，以前不确定分类的变异可能需要进行再分析；最后，在变异证据分层存在差异时应请该疾病领域专家进行判断。

九、伦理学及遗传咨询

迄今临床实践中全外显子组测序明确诊断率不足30%[2,20,25]，假阴性率仍较高；亦有一些伦理学和遗传学问题尚未解决，例如，是否应告知携带者、检测结果解析出家庭中出现非血缘关系、变异可能不完全外显、评价正常人群或无症状个体或者解释与检测初衷无关的偶然发现、检测出致病性突变但缺乏有效治疗药物等[26]；以及患者检出2种或以上致病基因，如何精准解读临床表型与基因型之间的关系[27]，上述问题的解决，应建立在合理的伦理学和遗传咨询基础上[28]。此外，还应考虑检测结果解析错误可能对患者及其家属的重要影响，如预防性乳腺切除术、心脏除颤器植入术和产前诊断决策，建议参考体格检查、实验室检查、影像学检查和电生理学检查等辅助检查结果，以提供合理检测报告、开展遗传咨询和进行健康管理[1,4,29]。

十、展望

二代基因测序技术在人类孟德尔遗传性疾病分子诊断和遗传学研究中的应用，仍有许多亟待解决的问题。尤其是目前的序列变异解析并非完美，所报道的变异分类并非100%确定，变异分类基于临床数据和经验，随着基因组学数据的不断增加，

表3 根据致病性或良性证据分层进行序列变异解析的规则[4]

Table 3. Rules for combining criteria to classify sequence variants[4]

Classification	Least evidence required
Pathogenic	1 Very strong (PVS1)
	And ≥ 1 strong (PS1–4)
	Or ≥ 2 moderate (PM1–6)
	Or ≥ 1 moderate (PM1–6) And 1 supporting (PP1–5)
	Or ≥ 2 supporting (PP1–5)
	≥ 2 Strong (PS1–4)
Likely pathogenic	1 Strong (PS1–4)
	And ≥ 3 moderate (PM1–6)
	Or 2 moderate (PM1–6) And ≥ 2 supporting (PP1–5)
	Or 1 moderate (PM1–6) And ≥ 4 supporting (PP1–5)
	1 Very strong (PVS1) And 1 moderate (PM1–6)
	1 Strong (PS1–4) And 1–2 moderate (PM1–6)
	1 Strong (PS1–4) And ≥ 2 supporting (PP1–5)
	≥ 3 Moderate (PM1–6)
	2 Moderate (PM1–6) And ≥ 2 supporting (PP1–5)
	1 Moderate (PM1–6) And ≥ 4 supporting (PP1–5)
Benign	1 Stand-alone (BA1)
	≥ 2 Strong (BS1–4)
Likely benign	1 Strong (BS1–4) And 1 supporting (BP1–7)
	≥ 2 Supporting (BP1–7)
Uncertain significance	Other criteria shown above are not met
	The criteria for benign and pathogenic are contradictory

PVS, pathogenic very strong, 致病性极强; PS, pathogenic strong, 致病性强; PM, pathogenic moderate, 致病性中度; PP, pathogenic supporting, 致病性支持; BA, benign stand-alone, 良性独立; BS, benign strong, 良性强; BP, benign supporting, 良性支持

在现有指南基础上,通过不同领域专家共同协作以建立更加精准的"基因-疾病"解读指南是未来发展方向。随着二代基因测序技术的发展和数据分析软件的完善,检测变异和分析变异能力必将逐步提高。同时,随着精准医疗计划的开展,也将为二代基因测序技术积累更多翔实、可靠的临床信息和基因组学数据,为该项技术更好地应用于人类孟德尔遗传性疾病分子诊断、预防干预、药物治疗和药物研发提供有力保证。

参 考 文 献

[1] Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. N Engl J Med, 2014, 371:1170.

[2] Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA, 2014, 312:1880-1887.

[3] Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain, 2010, 133(Pt 12):3510-3518.

[4] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, HegdeM, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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. Genet Med, 2015, 17:405-424.

[5] Frebourg, T. The challenge for the next generation of medical geneticists. Hum Mutat, 2014, 35:909-911.

[6] Altman RB, Prabhu S, Sidow A, Zook JM, Goldfeder R, Litwack D, Ashley E, Asimenos G, Bustamante CD, Donigan K, Giacomini KM, Johansen E, Khuri N, Lee E, Liang XS, Salit M, Serang O, Tezak Z, Wall DP, Mansfield E, Kass-Hout T. A research roadmap for next-generation sequencing informatics. Sci Transl Med, 2016, 8:335.

[7] Endrullat C, Glokler J, Franke P, Frohme M. Standardization and quality management in next-generation sequencing. Appl Transl Genom, 2016, 10:2-9.

[8] Singleton AB. Exome sequencing: a transformative technology. Lancet Neurol, 2011, 10:942-946.

[9] Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, Race V, Sistermans E, Sturm M, Weiss M, Yntema H, Bakker E, Scheffer H, Bauer P; EuroGentest, European Society of Human Genetics. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet, 2016, 24:2-5.

[10] MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. Nature, 2014, 508:469-476.

[11] McKusick VA. Mendelian inheritance in man and its online version, OMIM. Am J Hum Genet, 2007, 80:588-604.

[12] Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L. Challenges and standards in integrating surveys of structural variation. Nat Genet, 2007, 39: S7-15.

[13] den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PE. HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat, 2016, 37:564-569.

[14] Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet, 2011, 12:745-755.

[15] van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet, 2014, 30: 418-426.

[16] Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-

accuracy detection of low-frequency variation. Genome Res, 2013, 23:843-854.
[17] Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, Sudmant PH, Graves TA, Alkan C, Dennis MY, Wilson RK, Turner SW, Korlach J, Eichler EE. Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res, 2014, 24:688-696.
[18] Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, Kuk J, Park GH, Kim J, Ryu H, Kim J, Roh M, Baek J, Hunkapiller MW, Korlach J, Shin JY, Kim C. De novo assembly and phasing of a Korean human genome. Nature, 2016, 538:243-247.
[19] Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Lehner T, Shugart YY, Price AL, de Bakker PI, Purcell SM, Sunyaev SR. Exome sequencing and the genetic basis of complex traits. Nat Genet, 2012, 44:623-630.
[20] Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. N Engl J Med, 2013, 369:1502-1511.
[21] Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell, 2013, 155:27-38.
[22] Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, Zhang J, Zhou H, Tian L, Prakash O, Lemire M, Sleiman P, Cheng WY, Chen W, Shah H, Shen Y, Fromer M, Omberg L, Deardorff MA, Zackai E, Bobe JR, Levin E, Hudson TJ, Groop L, Wang J, Hakonarson H, Wojcicki A, Diaz GA, Edelmann L, Schadt EE, Friend SH. Analysis of 589 306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nat Biotechnol, 2016, 34:531-538.
[23] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet, 2014, 46:310-315.
[24] Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, Cooper GM, Dorschner MO, Dulik MC, Ghazani AA, Ghosh R, Green RC, Hart R, Horton C, Johnston JJ, Lebo MS, Milosavljevic A, Ou J, Pak CM, Patel RY, Punj S, Richards CS, Salama J, Strande NT, Yang Y, Plon SE, Biesecker LG, Rehm HL. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research consortium. Am J Hum Genet, 2016, 99:247.
[25] Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA, 2014, 312:1870-1879.
[26] Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med, 2013, 15:565-574.
[27] Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, Walkiewicz M, Bi W, Xiao R, Ding Y, Xia F, Beaudet AL, Muzny DM, Gibbs RA, Boerwinkle E, Eng CM, Sutton VR, Shaw CA, Plon SE, Yang Y, Lupski JR. Resolution of disease phenotypes resulting from multilocus genomic variation. N Engl J Med, 2017, 376:21-31.
[28] Biesecker LG. Exome sequencing makes medical genomics a reality. Nat Genet, 2010, 42:13-14.
[29] Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med, 2013, 15:733-747.

（收稿日期：2017-07-01）

·小词典·

中英文对照名词词汇（一）

半乳糖苷结合凝集素3
lectin-galactose binding-soluble 3（LGALS3）
表观扩散系数 apparent diffusion coefficient（ADC）
表皮生长因子 epidermal growth factor（EGF）
丙酰肉碱 propinoylcarnitine（C3）
Fahr病 Fahr's disease（FD）
Kufor-Rakeb病 Kufor-Rakeb disease（KRD）
常染色体显性遗传 autosomal dominant（AD）
常染色体隐性遗传 autosomal recessive（AR）
垂体肿瘤转化基因
pituitary tumor-transforming gene（PTTG）
雌激素受体α estrogen receptor α（ERα）
促肾上腺皮质激素 adrenocorticotropic hormone（ACTH）
单核苷酸变异 single nucleotide variation（SNV）
单核苷酸多态性 single nucleotide polymorphism（SNP）
单核苷酸多态性微阵列
single nucleotide polymorphism array（SNP array）
低血钾型周期性麻痹
hypokalemic periodic paralysis（HypoPP）
低β脂蛋白血症-棘红细胞增多症-视网膜色素变性-苍白球变性综合征
hypobetalipoproteinemia, acanthocytosis, retinitis pigmentosa and pallidal degeneration syndrome（HARP）
短串联重复序列 short tandem repeat（STR）
断裂转导蛋白样增强子
transducin-like enhancer of split（TLE）