

# 加快建设高通量测序标准数据库，实现全民“精准医疗”的标准化

杨振<sup>\*\*</sup>，张孝明<sup>#</sup>，黄杰，胡泽斌，李丽莉，孙彬裕，李颖（中国食品药品检定研究院，北京 100050）

**摘要** 目的：为我国高通量测序标准数据库的建设提供参考。方法：总结分析国内外高通量测序产品及数据库建设现状，剖析当前我国高通量测序标准数据库建设中亟待解决的问题。结果与结论：高通量测序是精准医疗重要的组成部分，国内外已批准多个高通量测序产品用于癌症等的临床诊治。但是，高通量测序能否在精准医疗中发挥更好的作用，还取决于是否有标准数据库用于准确地指导测序数据解读。随着高通量测序技术的发展，海量基因测序数据不断产生，如何建立中国人群基因变异解读标准数据库，用于评估和指导相关检测机构对致病基因的解读具有重大的现实意义。目前急需引入新的数据库建设及数据解读机制，为我国高通量测序行业的标准化健康发展及科学监管提供依据和保障。

**关键词**：高通量测序；标准数据库；数据库建设；数据解读；区块链；科学监管

中图分类号：R95 文献标识码：A 文章编号：1002-7777(2019)09-1051-07

doi:10.16153/j.1002-7777.2019.09.015

## Accelerating the Construction of High-Throughput Sequencing Standard Database to Achieve Standardization of National "Precision Medicine"

Yang Zhen<sup>\*\*</sup>, Zhang Xiaoming<sup>#</sup>, Huang Jie, Hu Zebin, Li Lili, Sun Binyu, Li Ying (National Institutes for Food and Drug Control, Beijing 100050, China)

**Abstract Objective:** To provide suggestions for the construction of high-throughput sequencing standard database in China. **Methods:** The current status of high-throughput sequencing products and database construction at home and abroad were summarized and analyzed. The urgent problems needed to be solved in the construction of high-throughput sequencing standard database in China were analyzed. **Results and Conclusion:** Since high-throughput sequencing is an important part of precision medicine, many high-throughput sequencing products have been approved for clinical diagnosis and treatment of tumors at home and abroad. However, whether high-throughput sequencing can play a better role in precision medicine depends on whether there is a standard database to accurately guide the sequencing data interpretation. With the development of high-throughput sequencing technology, massive gene sequencing data are continuously generated. It is of great practical significance to establish a standard database for the interpretation of gene variations of Chinese population, which can be used to evaluate and guide the interpretation of pathogenic genes by relevant detection institutions. At present, it is urgent to introduce a new database construction and data interpretation mechanism in order to provide basis and guarantee

for the healthy development and scientific regulation of high-throughput sequencing industry in China.

**Keywords:** high-throughput sequencing; standard database; database construction; data interpretation; block chains; scientific regulation

高通量测序 (High-throughput Sequencing) 又称下一代测序 (Next Generation Sequencing, NGS) 或大规模平行测序 (MPS), 包含多种可以一次性产生大量数字化基因序列的测序技术。高通量测序采用平行测序的理念, 能够同时对上百万甚至数十亿个DNA分子进行测序, 相较于以Sanger测序为代表的第一代测序技术, 具有通量高、灵敏度高、速度快的优势, 是继Sanger测序后的革命性进步<sup>[1-2]</sup>。不同厂家的高通量测序产品的测序原理不同, 主要包括边合成边测序 (Sequencing by Synthesis, SBS)、基于“DNA簇”和“可逆性末端终结” (Reversible Terminator) 大规模平行测序、四色荧光标记寡核苷酸的连续连接反应测序和半导体芯片测序等。与常规的分子检测相比, 高通量测序技术能够一次性对多个靶基因进行准确检测, 具有所需样本量小、敏感性高、检测成本低、耗时短等优点。近年来, 高通量测序技术不断优化升级、测序数据解读更智能化、检测费用也大大降低。高通量测序技术的应用已从科研领域拓展到遗传性疾病、生殖健康、妇幼保健、肿瘤靶向用药指导、病原微生物的快速鉴定等多个临床分子诊断领域。高通量测序应用于临床疾病诊断, 可以满足传统的分子检测方法 (例如一代Sanger测序、荧光原位杂交等) 不能实现的同时检测多基因的需求<sup>[3]</sup>, 已逐渐成为现代临床分子诊断的重要工具。

精准医疗 (Precision Medicine), 又称为个体化医疗, 是通过基因组、蛋白质组等组学技术和医学前沿技术, 对于大样本人群与特定疾病类型进行生物标志物的分析与鉴定、验证与应用, 从而精确寻找到疾病的原因和治疗的靶点, 并对一种疾病不同状态和过程进行精确分类, 最终实现对于疾病和特定患者进行个性化精准治疗的目的<sup>[4]</sup>。高通量测序是精准医疗的高驱动因素, 因为精准医疗需要深入了解细胞基因变异信息, 综合分析基因多态性, 才能找出准确的致病基因, 并确定对应的靶向治疗。目前, 国内外已经研发出多种高通量测序试剂盒, 应用于单基因疾病、癌症等的诊疗中<sup>[5-6]</sup>。

同时, 还应看到, 高通量测序要在精准医疗中发挥更好的作用, 还需建立标准的数据解读机制。一方面, 由于肿瘤发生时与之相关的基因突变位点成千上万, 而起决定性作用的基因突变仅占少数, 如何从海量的突变中找出引发疾病的关键基因并非易事。另一方面, 随着高通量测序临床应用的突飞猛进, 临床机构掌握了越来越多的个人基因组测序信息, 其中既有临床意义明确的基因信息, 又包括大量临床意义不明确的基因信息, 这为数据的解读和有效利用提出了新的挑战<sup>[7-8]</sup>。如何整合海量的临床测序数据, 建立标准的数据库, 以进行有效数据解读, 并为监管提供技术支撑, 是高通量测序更好地服务于精准医疗的关键, 也是一个极具挑战并关系到精准医疗能否蓬勃发展的重要问题。

建立标准数据库, 首先需要收集更多的个体基因数据及基因变异信息。为此, 国内外许多机构纷纷建立了基因大数据库。但是, 由于数据采集机构的数据生成过程和采用的方法往往存在差异, 导致得到的基因变异数据质量参差不齐, 甚至包含许多错误数据, 这极大影响了数据解读和有效利用<sup>[9-10]</sup>。在我国, 为促进高通量测序产业的健康发展, 监管部门已制订了一系列临床应用标准、行业标准、检测指南和监管措施<sup>[11-12]</sup>, 但在基因测序数据库建设及数据解读方面, 与国外还存在较大差距。目前急需建立基于中国人群基因信息的标准数据库, 用于测序数据解读, 为实现全民精准医疗的标准化提供技术支撑。本文概述了我国及美国批准的高通量测序诊断产品及标准测序数据库建设和数据解读的进展, 对我国高通量测序标准数据库建设提出了思考建议。

## 1 美国高通量测序诊断产品、数据解读及数据库发展概况

### 1.1 美国高通量测序诊断产品发展概况

截至2019年6月13日, 美国食品药品监督管理局 (FDA) 共批准了7个基于高通量测序法的诊断产品, 按照批准时间梳理如下:

2013年11月19日, Illumina 公司的2个MiSeq Dx 测序平台与配套试剂盒成功获得FDA批准, 成为

美国首批基于高通量测序的仪器及试剂。MiSeqDx Cystic Fibrosis 139测序平台可同时检测CFTR基因的139个临床相关的致病突变及变异位点（包括美国和加拿大医学遗传学会ACMG和CCMG建议筛查的位点）。MiSeqDx Cystic Fibrosis Clinical测序平台则可准确检测CFTR基因的蛋白编码区域以及内含子/外显子边界的基因信息。

2016年12月19日，美国Foundation Medicine公司的FoundationFocus™ CDxBRCA高通量测序诊断试剂和新药Rubraca同时获得美国FDA批准分别用于BRCA基因突变检测及携带BRCA致病突变（同时包括胚系和体系突变）卵巢癌患者的三线及以上治疗，这是美国历史上第一例药物与诊断试剂同步研发、同步提交、同步获批。由于BRCA 1/2基因的突变没有明显的热点突变而且还有多种突变类型，除了常见点突变、小片段插入/缺失外，也包括大片重排这样的外显子水平的结构变异，常规一代测序方法并不能涵盖以上多种突变类型。高通量测序却可以检测BRCA基因的全部外显子，并通过生物信息算法完成对各种突变类型的分析。

2017年6月23日，Thermo Fisher赛默飞公司的Oncomine DX Target Test获准，成为FDA批准的第一个可筛查多个肿瘤标志物的高通量测序产品。此产品将23个有明确治疗意义的基因标志物整合到一个CDx中，其中BRAF、ROS1、EGFR基因的检测分别获批用于非小细胞肺癌患者对达拉菲尼与曲美替尼联合用药（Dabrafenib with Trametinib）、克里唑蒂尼（Crizotinib）、易瑞沙（Gefitinib）药物治疗方案的选择。

2017年6月29日，Illumina公司开发Extended RAS Panel获得FDA批准，用于基因检测和Amgen公司帕尼单抗药物（Panitumumab）在转移性结肠癌治疗中药物的选择，此试剂盒可检测KRAS/NRAS基因明确的56个突变，发挥了高通量测序可以检测更多明确位点的优势。

2017年11月15日，FDA批准了凯特琳癌症中心（Memorial Sloan Kettering Cancer Center, MSK）的基于高通量测序技术的多基因检测试剂MSK-IMPACT。该产品可对实体瘤患者组织样本的468个基因全部外显子和特定内含子进行深度测序，通过生物信息流程进行突变分析和微卫星不稳定MSI分析，结合MSKCC开发的OncoKB数据库对检测数据

进行解读。除了已知的临床意义明确的基因，该产品提供的信息还包括大量相对尚无明确临床意义，但可以增加对患者肿瘤认识的基因及其突变。该产品在数据解读中对基因突变按照临床意义及证据等级进行划分，以便保证测序过程及数据解读的准确可信。

2017年11月30日，Foundation Medicine公司的旗舰产品FoundationOne CDx获FDA批准，用于非小细胞肺癌、黑色素瘤、乳腺癌、结直肠癌以及卵巢癌的临床诊疗。该产品可检测分析实体瘤患者福尔马林固定石蜡包埋样本的324个基因的多种突变及特定基因融合，并进一步分析两个基因组学特征—微卫星不稳定性（MSI）和肿瘤突变负荷（TMB），检测结果可用于指导15种靶向药物的临床使用。这是通过FDA最新颁布的Breakthrough Device法案获得上市许可的第一个体外诊断产品，也是FDA第一次基于整个基因组及相应的技术平台进行的审批。

FDA对于肿瘤诊断产品的审批都具体到与药物疗效相关的变异位点。但与传统方法相比，高通量测序技术平台的特点在于可以平行检测多基因、多位点、多种变异形式。而且，在同一技术平台检测同类型变异（突变、插入缺失、扩增、或融合重排等）中的不同基因或位点时，高通量检测性能和参数具有高度的一致性，在审核方法学和质控后，即使不对所有位点进行一一验证，也可以保证结果的精准、可重复。随着与用药敏感性、耐药性、预后等多种临床价值密切相关的基因变异不断被发现。临床的迫切需求和科学的发展推动了此次美国FDA监管方式的改革与创新。

## 1.2 美国高通量测序数据解读及数据库建设发展概况

为满足不断增长的高通量测序数据存储需求，美国国家生物技术信息中心（National Center of Biotechnology Information, NCBI）最早在2007年底推出了SRA数据库（Sequence Read Archive），用于存储、显示、提取和分析高通量测序数据<sup>[13]</sup>。为有效开发利用基因测序信息，NCBI于2013年设立了ClinVar数据库和ClinGen项目。ClinVar是一个公开数据库，整合自十多个不同类型的数据库，是将变异、临床表型、实证数据以及功能注解与分析四方面的信息，通过专家评审形成的一个遗传变异—

临床表型相关的基因信息数据库<sup>[14]</sup>。ClinVar使用标准的命名法描述疾病，支持公众免费下载数据。ClinGen的目标是整合临床与基础研究专家的技术力量和资源，为分析基因变异数据与健康与疾病之间的关系开发标准化流程。ClinVar和ClinGen为确定基因和基因变异在疾病发生中的作用而相互协作<sup>[15]</sup>。

尽管ClinVar收集的数据量也在逐年增加，但却存在来自低质量实验室的错误数据的问题，使得公众在ClinVar上查找疾病相关的基因变异时无所适从。而如果临床人员采纳了不准确的信息，可能会导致错误的诊疗路线。为解决上述问题，ClinGen项目的研究人员推出了多种评估方法和制度对ClinVar进行改进。在数据解读方面，一是设立“基因-疾病”有效性框架对基因变异在特定疾病中的作用进行评估，由癌症、心血管疾病等领域的知名研究人员和有建树的临床医生组成多个基因数据管理专家小组对这种评估方法进行验证测试，以指导实验室决定将哪些基因作为诊断评估或进行测试<sup>[16-17]</sup>。二是采纳美国医学遗传学与基因组学学会（ACMG）和分子病理学协会制订的方针，指导识别致病性基因变异和非致病的变异<sup>[18-19]</sup>。三是开发了评估临床上次要基因变异的方法，将个体身上发现的疑似致病基因变异作为疾病诊断的重要参考指标，但不作为诊疗依据。

在数据共享方面，ClinGen研发人员一是建立了ClinGen等位基因注册中心，开发了符合全球基因组学与健康联盟的基因组知识标准工作流程和通用“基因语言”，实现了基因变异的通用识别<sup>[20]</sup>。二是设立了“满足数据共享最低要求以保证质量的临床实验室列表”，通过设立激励机制，鼓励临床实验室投入更多必要的资源，在数据库中负责地提交并分享其数据，通过与实验室之间的密切合作进一步解决分类差异的问题，改善实验室解读基因变异的能力<sup>[21]</sup>。以上这些措施有效改善了公众可用的基因测序数据质量。FDA于2018年12月4日正式认可ClinGen，基因检测研发人员可以使用ClinGen的基因变异信息，作为证明其检测试剂临床有效性的部分提交材料，支持其在基因检测方面的临床声明，而不必再生成相同的数据。FDA认可ClinGen数据的可靠性，对美国基因测序监管及高通量测序行业发展都具有里程碑式的意义，标志着美国高通量测序标准化数据库基本建成。

## 2 我国高通量测序诊断产品发展概况

我国高通量测序行业还处于发展的初期阶段，2014年以前处于无监管状态。2014年2月，原国家食品药品监督管理总局（以下简称“原国家药监局”）和原卫计委叫停了所有基因测序业务，对行业进行集中整顿。2014年3月，原卫计委发布《关于开展高通量基因检测技术临床应用试点单位申报工作的通知》，通知要求已经开展高通量基因测序技术，且符合申报规定条件的医疗机构可以申请试点，同时明确申请试点的基因测序项目。2014年12月，原卫计委医政医管局发布遗传病诊断、产前筛查与诊断、植入前胚胎遗传学诊断这三个专业的第一批基因测序临床试点名单。随后2015年1月，原卫计委妇幼司正式批准108家医疗服务机构开展无创产前检测高通量测序临床试点，并审核通过13家机构开展植入前胚胎遗传学诊断临床试点。2015年3月，原卫计委医政医管局发布了第一批肿瘤诊断与治疗项目高通量测序临床试点名单，此后原国家药监局相继批准了几款用于高通量测序的仪器和检测试剂。

截至2019年6月13日，一共批准了13个基于高通量测序法的试剂盒，其中用于无创产前21三体筛查的有7个，非小细胞肺癌用药诊断的4个，卵巢癌用药诊断的1个，人乳头状瘤病毒传染性疾病预防的1个。除人乳头状瘤病毒核酸分型检测试剂盒外，其他12个诊断试剂均是条件批准上市，要求其上市后进一步搜集临床数据，在下次延续注册时提供证据。具体产品信息见表1。

EGFR、KRAS、BRAF、PIK3CA、ALK、ROS1等基因变异相关的高通量测序试剂盒（肿瘤用药中1-4项），用于定性检测非小细胞肺癌（NSCLC）肿瘤组织福尔马林固定石蜡包埋切片（FFPE）样本中EGFR、KRAS、BRAF、PIK3CA、ALK、ROS1基因的多种变异。

人类BRCA1基因和BRCA2基因突变检测试剂盒（肿瘤用药中第5项）用于体外定性检测上皮性卵巢癌患者外周血样本DNA中BRCA1基因和BRCA2基因全编码区【包括BRCA1基因（NM\_007294.3）外显子2、3、5~24，BRCA2基因（NM000059.3）外显子2~27】及外显子-内含子连接区、UTR区（非翻译区）和启动子区的点突变和插入缺失突变。

表 1 我国批准的基于高通量测序的诊断产品

分类	编号	批准日期	名称	生产企业名称
无创 产前 筛查	1	2014-06-30	胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (半导体测序法)	华大基因生物科技 (深圳) 有限公司
	2	2014-06-30	胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (联合探针锚定连接测序法)	华大基因生物科技 (深圳) 有限公司
	3	2014-11-04	胎儿染色体非整倍体 21 三体、18 三体和 13 三体检测试剂盒 (半导体测序法)	中山大学达安基因股份有限公司
	4	2015-02-12	胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (半导体测序法)	博奥生物集团有限公司
	5	2015-03-20	胎儿染色体非整倍体 (T13/T18/T21) 检测试剂盒 (可逆末端终止测序法)	杭州贝瑞和康基因诊断技术有限公司
	6	2017-01-13	胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (联合探针锚定聚合测序法)	华大生物科技 (武汉) 有限公司
	7	2017-03-03	胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (可逆末端终止测序法)	安诺优达基因科技 (北京) 有限公司
肿瘤 用药	8	2018-07-18	人 EGFR/ALK/BRAF/KRAS 基因突变联合检测试剂盒 (可逆末端终止测序法)	广州燃石医学检验所有限公司
	9	2018-08-11	人 EGFR、KRAS、BRAF、PIK3CA、ALK、ROS1 基因突变检测试剂盒 (半导体测序法)	天津诺禾致源生物信息科技有限公司
	10	2018-09-28	EGFR/ALK/ROS1/BRAF/KRAS/HER2 基因突变检测试剂盒 (可逆末端终止测序法)	南京世和医疗器械有限公司
	11	2018-11-16	人类 10 基因突变联合检测试剂盒 (可逆末端终止测序法)	厦门艾德生物医药科技股份有限公司
	12	2019-02-21	人类 BRCA1 基因和 BRCA2 基因突变检测试剂盒 (可逆末端终止测序法)	厦门艾德生物医药科技股份有限公司
传染病	13	2017-06-08	人乳头瘤病毒 (16 种型别) 核酸分型检测试剂盒 (半导体测序法)	华大生物科技 (武汉) 有限公司

人乳头瘤病毒 (16种型别) 核酸分型检测试剂盒, 用于人乳头瘤病毒核酸分型检测时文库的构建, 与已批准的核酸纯化试剂、测序反应通用试剂盒 (半导体测序法)、人乳头瘤病毒核酸分型分析软件联合使用, 用于女性宫颈脱落细胞中16种人乳头瘤病毒 (HPV6、11、16、18、31、33、35、39、45、51、52、56、58、59、66、68) 核酸的定性和分型检测。

从高通量测序产品数量上看, 我国高通量测序已走在世界前列, 但是, 我国高通量测序与美国等发达国家还存在较大差距。例如我国批准的用于指导肿瘤用药的高通量测序产品可检测的基因变异数远少于国外批准的产品, 而且我国缺少对产品测序结果的标准解读规范, 对测序数据的解读能力还

有待提高。

### 3 我国高通量测序数据库建设及数据解读现状

近年来, 随着高通量测序临床应用的增长和测序数据的不断积累, 我国高通量测序标准数据库建设的需求也日益凸显。中国的生物组学数据产量约占全球的40%<sup>[22]</sup>, 是数据产出大国, 但针对中国人群的基因大样本研究仍由分散的企业及临床科研机构承担, 缺乏统一集成的大规模基因信息数据库和被国际认可的数据库系统, 研究人员不得不将我国宝贵的数据资源交给他人代管。另外, 尽管高通量测序已被应用于肿瘤临床诊断、遗传性疾病检测等多个临床诊断领域, 但是这些研究多集中于白种人, 而不同人群中的致病风险基因存在差异, 目前

缺少针对中国人群的高通量测序基因组研究和变异数据库。

2016年12月,中国科学院北京基因组研究所生命与健康大数据中心(BIG Data Center)首次以数据中心为模式整体发布我国生命组学数据资源建设情况,标志着继美国的NCBI、欧洲的EBI及日本的DDBJ之后,我国建成全球第四个综合基因组权威大数据库。目前,该数据库资源系统包括高通量测序的原始组学数据归档库GSA,基于中国人群以及国家重要物种群体的基因组变异数据库GVM等6个分数据库,初步形成了我国生命与健康数据汇交与共享平台<sup>[23]</sup>。2018年7月,东方肿瘤临床研究中心、莲和医疗与希思科(CSCO)临床肿瘤学研究基金会共同发起“肿瘤基因信息学大数据平台”,由35家国内大型三甲医院的权威肿瘤临床专家共同参与建设,旨在建立中国人群特色高发肿瘤的基因信息数据库,为国内肿瘤临床研究和临床实践提供可靠的信息技术支撑。

尽管我国已开始重视高通量测序的发展,也在不断加强测序数据库建设,但是目前仍存在许多不足,主要表现:一是现有的数据库数据来源有限,和临床实验室、临床检测交互功能弱。二是在基因变异解读方面,缺乏统一的解读标准和规范,解读更新不及时。三是基因检测产品在临床应用申请上市前,需要经过国家药监局的检测分析性能评估和临床有效性审查,在检测质量、变异结果解读、产品监管等方面缺少高可靠性且部分不公开的评估用数据库。

目前,急需引入新机制,建立一个全新的、可靠的、针对中国人群的基因变异解读标准数据库,用于系统评估和指导相关检测机构对相关基因的解读。通过大规模收集可靠的测序数据和对解读过程中出现分歧的位点进行详细分析,为测序数据解读提供更细化的意见,以便为临床有效性验证提供参比标准,并为高通量测序产品科学监管提供依据,促进相关医疗和遗传咨询工作的开展。

#### 4 建立基于区块链新技术的高通量测序数据库平台

我国现阶段已有的高通量数据库包含的变异位点数量以及对应的临床信息还比较有限,需要扩大数据库规模。另外,如何进行数据采集的标准化,如何建立数据质量的评估方法,如何考虑各家

检测机构的利益、数据安全和隐私合规的需求;同时,如何做到数据可溯源,流程科学公开透明,不可篡改,符合伦理要求,定期更新维护等,都是大规模高通量测序数据库建立的关切点。为满足上述要求,可以引入区块链技术,通过智能合约、确权存证、分布式存储、点对点加密传输、安全计算等区块链技术特性,形成集成方案,设立中央认证节点,检测机构和临床机构作为数据来源节点,多方协作,共建标准。以此为蓝本,未来可进一步扩展到其余上千种疾病相关基因的解读标准建立。

区块链新技术基因变异标准解读数据库建设可分三个阶段完成:第一阶段完成搭建联盟链,完成全流程的数字化改造,为各方数据确权存证;第二阶段建设治理体系,建立争议位点解读的解决机制,各机构共建解读标准;第三阶段形成对外服务能力,数据库不断迭代更新,可供临床参比对照使用。

通过联合多家机构建立的区块链新技术基因变异标准解读数据库,将形成数据协作共建标准的全新合作范式,可汇聚中国人群的变异解读标准数据库,为产品临床有效性的科学性提供证据,有利于加速基因检测产品的上市审评,让精准医疗真正惠及百姓。

另外,要建成权威、标准、经得起检验、得到业界广泛认可的数据库,还要建立科学完善的评价体系和评价标准;建立广泛的包含临床、科研机构等领域的专家团队,以及相应的数据入库专家组审核规则和程序,并依靠相关程序开展工作;建立相应的机制来激励提交上传数据;依据《中华人民共和国人类遗传资源管理条例》要求,在不损害相关企业的利益前提下做好基因数据的隐私保护和伦理合规性;要遵循中国人基因信息保密的相关法律法规要求;数据库是根据当前科学认知给出的参考,要确立纠错机制和免责条款,建立实时更新制度;对利益诉求要符合相关规定。

#### 5 结语

高通量测序技术已成为临床精准诊疗的重要工具,但是测序数据不能直接提供疾病信息,需要经过数据过滤筛选、基因组对比等多个步骤后,才能得到受检者基因变异信息,为疾病诊疗提供参考,而且测序数据分析解读需要强大的基因数据库支持。现阶段国内外一些区域性基因数据库已在临

床诊疗中发挥了重要作用,但是这些数据库缺乏完善的中国人群的基因组研究和变异数据;另外,我国整个基因信息大数据领域仍然缺乏统一标准,大数据不能充分应用于临床。通过建设基于区块链新技术的高通量测序数据库平台,建立完善成熟的监管体系,有利于规范国内高通量测序技术相关产品,为有效准确的临床诊断提供更强有力的工具。

#### 参考文献:

- [1] Goldfeder R L, Wall D P, Khoury M J, et al. Human Genome Sequencing at the Population Scale: A Primer on High-Throughput DNA Sequencing and Analysis[J]. *Am J Epidemiol*, 2017, 186 ( 8 ) : 1000-1009.
- [2] Koboldt D C, Steinberg K M, Larson D E, et al. The Next-generation Sequencing Revolution and Its Impact on Genomics[J]. *Cell*, 2013, 155 ( 1 ) : 27-38.
- [3] 诺禾致源科技服务部. 高通量测序与大数据分析医学篇: 第一部分人类基因组研究 (2016版) [M]. 2016, 2-182.
- [4] König I R, Fuchs O, Hansen G, et al. What Is Precision Medicine?[J]. *Eur Respir J*, 2017, 50 ( 4 ) : 1-12.
- [5] 燕赛英, 王晓巍, 田彩娟, 等. 精准医疗在恶性肿瘤中的应用进展[J]. *广东医学*, 2017, 38 ( 21 ) : 3374-3381.
- [6] 张乐吟, 孙磊涛, 沈敏鹤. 基因变异检测技术在恶性肿瘤精准医疗中的应用[J]. *中国肿瘤生物治疗杂志*, 2019, 26 ( 1 ) : 22-28.
- [7] 张文力. 高通量测序数据分析现状与挑战[J]. *集成技术*, 2012, 1 ( 3 ) : 20-24.
- [8] 黎籽秀, 刘博, 徐凌丽, 等. 高通量测序数据分析和临床诊断流程的解读[J]. *中国循证儿科杂志*, 2015, 10 ( 1 ) : 19-24.
- [9] Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing[J]. *Arch Pathol Lab Med*, 2017, 141 ( 11 ) : 1544-1557.
- [10] Hardwick S A, Deveson I W, Mercer T R. Reference Standards for Next-generation Sequencing[J]. *Nat Rev Genet*, 2017, 18 ( 8 ) : 473-484.
- [11] 国家食品药品监督管理总局. 胎儿染色体非整倍体 (T21、T18、T13) 检测试剂盒 (高通量测序法) 注册技术审查指点原则[S]. 2017.
- [12] 国家卫生计生委个性化医学检测技术专家委员会. 肿瘤个性化治疗检测技术指南 (试行) [S]. 2015.
- [13] 熊筱晶. NCBI高通量测序数据库SRA介绍[J]. *生命的化学*, 2010, 30 ( 6 ) : 959-963.
- [14] Melissa J. Landrum, et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants[J]. *Nucleic Acids Research*, 2016, 44 ( D1 ) : D862 - D868.
- [15] Caudle K E, Keeling N J, Klein T E, et al. Standardization Can Accelerate the Adoption of Pharmacogenomics: Current Status and the Path Forward[J]. *Pharmacogenomics*, 2018, 19 ( 10 ) : 847-860.
- [16] Strande N T, Riggs E R, Buchanan A H, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-based Framework Developed by the Clinical Genome Resource[J]. *Am J Hum Genet*, 2017, 100 ( 6 ) : 895-906.
- [17] DiStefano M T, Hemphill S E, Oza A M, et al. ClinGen Expert Clinical Validity Curation of 164 Hearing Loss Gene-disease Pairs[J]. *Genet Med*, 2019.
- [18] Rivera-Munoz E A, Milko L V, Harrison S M, et al. ClinGen Variant Curation Expert Panel Experiences and Standardized Processes for Disease and Gene-level Specification of the ACMG/AMP Guidelines for Sequence Variant Interpretation[J]. *Hum Mutat*, 2018, 39 ( 11 ) : 1614-1622.
- [19] Nykamp K, Anderson M, Powers M, et al. Sherlock: A Comprehensive Refinement of the ACMG-AMP Variant Classification Criteria[J]. *Genet Med*, 2017, 19 ( 10 ) : 1105-1117.
- [20] Pawliczek P, Patel R Y, Ashmore L R, et al. ClinGen Allele Registry Links Information about Genetic Variants[J]. *Hum Mutat*, 2018, 39 ( 11 ) : 1690-1701.
- [21] ClinGen. Clinical Laboratories Meeting Minimum Requirements for Data Sharing to Support Quality Assurance[EB/OL].[2019-06-20]. <https://clinicalgenome.org/tools/clinical-lab-data-sharing-list/>.
- [22] 中国科学院. 生物组学数据库获国际认可[EB/OL]. (2016-12-20) [2019-06-20]. <http://www.chinanews.com/gn/2016/12-20/8099687.shtml>.
- [23] The BIG Data Center: From Deposition to Integration to Translation[J]. *Nucleic Acids Res*, 2017, 45 ( D1 ) : D18-D24.

(收稿日期 2019年7月9日 编辑 郑丽娥)