

二代及三代测序技术在遗传学诊断中的应用进展

高海明, 赵彦艳*

中国医科大学附属盛京医院临床遗传科, 沈阳 110004

摘要: 遗传病的防治是公共卫生领域的重大课题, 而明确病因是遗传病防治的重要环节。高通量测序技术(又称二代测序技术)具有高通量、低成本、高准确度的优点, 为遗传诊断及咨询提供了直接证据, 已成为遗传学检测不可或缺的有力工具; 第三代测序也凭借其长读长的独特优势在临床应用中占据一席之地。二代及三代测序技术各有特点, 互为补充, 临床中针对不同的检测需求有多种类型的测序方案可供选择。基于此, 对二代及三代测序技术的原理、分类及其在遗传学诊断中的应用进展做一综述, 以期为临床测序方案的选择提供思路和指导。

关键词: 高通量测序; 三代测序; 遗传病; 遗传学诊断

DOI:10.19586/j.2095-2341.2020.0100

Application Progress of Next Generation and the Third Generation Sequencing Technology in Genetic Diagnosis

GAO Haiming, ZHAO Yanyan*

Department of Clinical Genetics, Shengjing Hospital, China Medical University, Shenyang 110004, China

Abstract: Prevention and treatment of genetic diseases is a major issue in public health, and it is an important part to clarify the genetic causes. High-throughput sequencing technology (also known as second generation sequencing technology) has the advantages of high throughput, low cost and high accuracy, which provides direct evidence for genetic diagnosis and consultation, and has become an indispensable powerful tool for genetic testing. The third generation sequencing also occupies a place in clinical application by virtue of its unique advantages of long reading time. Next generation and the third generation sequencing technologies have their own characteristics and complement each other. There are many types of sequencing schemes to choose from according to different detection requirements in clinic. Based on this, the principle, classification and application progress of the second and third generation sequencing technologies in genetic diagnosis were reviewed in order to provide ideas and guidance for the selection of clinical sequencing schemes.

Key words: high throughput sequencing; the third generation sequencing; genetic diseases; genetic diagnosis

出生缺陷是导致流产、死胎、新生儿死亡和婴儿死亡的重要原因^[1], 我国每年新增的出生缺陷数约 90~100 万例, 给家庭和社会造成极大的负担, 严重影响人口素质^[2]。因此, 如何有效预防出生缺陷、降低其发生率、增强国民身体素质成为了公共卫生领域的重大课题。21-三体综合征、18-三体综合征及 13-三体综合征是新生儿常见的先天性染色体异常及缺陷疾病, 在孕早期进行有效的胎儿染色体异常筛查对于改善孕妇分娩结局

及保证婴儿健康具有重要意义^[3-4]。遗传学诊断是指采用分子生物学技术对受检者全部或部分基因组信息进行检测和分析, 再结合临床信息来明确患者的遗传学病因, 为临床诊断和遗传咨询提供指导。常用的遗传学诊断技术包括核型分析、一代测序、二代测序及三代测序等。1977 年, 以末端终止法为特征的第一代测序即 Sanger 测序被发明, 其测序读长可达 1 000 bp, 结果准确可靠, 目前仍作为绝大多数遗传学测序诊断的金标

收稿日期: 2020-08-27; 接受日期: 2020-10-20

基金项目: 国家重点研发计划项目(2016YFC1000702)。

联系方式: 高海明 E-mail: gaohm@sj-hospital.org; * 通信作者 赵彦艳 E-mail: yzhaoy@sj-hospital.org

准^[5-6]。但由于 Sanger 测序依赖于 PCR 聚合酶和毛细管电泳,其测序成本高、分析速度慢,难以满足越来越大的测序需求。随着人类基因组计划(human genome project, HGP)的完成,通量更高、更经济的测序技术——高通量测序(high throughput sequencing, HTS)的开发和商业化应运而生。HTS 又被称为二代测序(next generation sequencing, NGS),与 Sanger 测序最大的不同之处在于,NGS 可以同时几十至几百万条 DNA 短片进行独立的大规模平行测序,配合下游的生物信息学分析来实现基因组测序^[7]。此外,为了弥补 NGS 读长较短的问题,第三代测序技术也初露锋芒,凭借其长读长的独特优势在临床应用中占据一席之地。多种测序技术各有特点,互为补充,共同为遗传学诊断提供了有力的检测手段^[8]。基于此,本文对二代及三代测序技术的原理、分类及其在临床遗传学诊断的应用进展做一综述,以期临床测序方案的选择提供思路和指导。

1 二代及三代测序技术的原理和分类

1.1 NGS 原理与分类

高通量测序原理为合成测序(sequencing by synthesis, SBS),是 Sanger 测序的进一步发展,主要步骤包括基因组打断、建库和测序。首先将待测 DNA 分子打断为长约几十到几百 bp 的片段,将其分散至数百万个反应孔中或固定于芯片上,经过 PCR 或等温扩增技术进行 DNA 预扩增反应,并通过检测带不同荧光标记的核苷酸或反应副产物来提供基因组信息^[9]。

目前绝大部分 NGS 平台均采用 SBS,但不同平台的核心技术有所区别,如 Roche(焦磷酸测序)^[10]、Illumina(桥式 PCR)^[11]、Ion PGM(半导体 H⁺测序)^[12]以及华大智造研发的首个国产测序平台 BIGSEQDNA(纳米球测序)^[13]。Roche 454 焦磷酸测序是最早出现的二代测序系统,是通过检测 DNA 合成反应时释放的焦磷酸(PPi)来读取碱基信息。微乳液 PCR(emulsion PCR, emPCR)可以将单个 DNA 片段与磁珠链接,磁珠被小油滴包被后进行独立的扩增,使每个磁珠上均为同一 DNA 片段的数百万个拷贝簇。每一轮碱基合成反应时只加入 1 种 dNTP,如果它刚好能

与模板匹配,则会合成到引物 3'端,释放出与光反应偶联的 PPi 从而实现测序^[10-11]。Roche 454 焦磷酸测序读长最长,但成本较高,目前其配套的测序仪已停产,退出了主流的二代测序市场。

目前,Illumina 占据了二代测序市场的绝大部分份额,其核心技术为桥式 PCR 和可逆终止子技术,即将基因组打断、变性为单链,通过与芯片表面的引物碱基互补被固定于芯片上,形成“桥”,经过 PCR 反应使每个分子获得了 1 000 倍扩增,成为单克隆 DNA 簇。测序反应时加入带有 4 种荧光标记的 dNTP“可逆终止子”^[14],通过切割其 3'羟基末端的特殊基团使聚合反应反复起始和终止,并记录荧光信号以得到序列信息。与照相机成像检测相比,带有荧光标记的碱基可以直接成像从而提高检测速度^[11]。Ion Torrent 半导体测序技术的出现拓展了测序市场,其配套测序仪 Ion PGM 是目前市场上最便宜的测序仪,可将核苷酸序列直接转换为半导体芯片上的数字信息。DNA 合成反应中会释放出 H⁺导致溶液 pH 变化,该平台利用离子传感器将溶液 pH 的变化记录为碱基信息,在该过程中不使用荧光或化学发光,不需要相机、光源或扫描仪,因此大大加快了测序过程,使 Ion Torrent 成为二代测序市场上的新流行^[12]。

华大智造作为后起之秀,开启了国产测序仪的高通量时代。其测序平台 BGISEQ 的独特之处在于文库构建过程,在建库时基因组 DNA 被打断、变性、环化为单链环状,利用 DNA 分子滚环扩增技术(rolling circle amplification, RCA)用以制备 DNA 纳米球(DNA nanoballs, DNBs),并使用联合探针锚定聚合技术(combinatorial probe-anchor synthesis, cPAS)使测序引物锚定分子和荧光探针在 DNA 纳米球上进行聚合反应,利用高分辨成像系统对光信号进行采集从而获得序列信息^[13,15]。华大已推出了多种不同通量的测序仪供用户选择,并且配套有相应的自动化样本制备系统,对于临床检测来说具有极大的优势^[16]。上述不同 NGS 平台的比较见表 1。

1.2 三代测序的原理与分类

与二代测序技术相比,三代测序不需要经过 PCR 扩增,并且可以实现超长读长测序。目前市场上可供选择的三代测序技术主要有两大平台:PacBio 公司开发的单分子实时测序(single

表 1 不同 NGS 平台对比
Table 1 Comparison of characteristics of different NGS platforms

厂商	原理	核心技术	测序平台	最大读长 /bp	最大通量	时间 /run	仪器特点和适用范围
Illumina	SBS	桥式 PCR	HiSeq 系列	2×250	1500 Gb	<1~6 d	通量高,灵活性高,成本低,读长短;适合全基因组 测序、靶向重测序、全转录组测序、denovo 测序、表 观遗传调控
			MiSeq 系列	2×300	15 Gb	27 h	适合微生物、病毒全基因组测序,运行速度快,不适 合于 WGS、WES
			NextSeq 500	2×150	120 Gb	11~29 h	适合大型及小型 WGS(人类、植物、动物)、WES、 TGS、转录组测序、甲基化测序等
			MiniSeq	2×150	7.5 Gb	4~24 h	运行时间短,灵活性高,有中通量和高通量 2 种试 剂可选;适合小型全基因组测序(微生物、病毒) TGS,不适合 WGS 和 WES
			NovaSeq	2×150	6 Tb	4h~3 d	适合 WGS(人类、植物、动物、微生物)、WES、TGS、 转录组测序、甲基化测序等
Thermo Fisher	SBS	半导体 H ⁺ 测序	Ion PGM	200	2 Gb	2 h	通量灵活,读长短;可用于小基因组测序组装,不利 于基因组 denovo
			Ion Proton	50	10~15 Gb	1~7 d	读长短,精度最高;适用于大型基因组测序、 WGS、WES
			Ion S5/S5 xl	200 或 400	10~50 Gb	2~24 h	价格低廉,测序速度快;适合于 TGS、WES、转录组 测序和 RNA-seq
华大	SBS	cPAS/ RCA/ DNBs	BGISEQ-500	2×100	260 Gb	24 h	支持多种读长,配套有全自动文库制备系统和样本 加载系统;适用于 NIPT、PGS、WGS\WES\RNA-seq、 转录组测序
			MGISEQ-200	2×150	150 Gb	66 h	支持高、低两种通量,多种读长可选;适合 CNV- seq、小型基因组测序、TGS、WES
			MGISEQ-2000	400 或 2×200	720 Gb	0.5~4 d	适用于大型基因组测序、WGS、WES、TGS
			MGISEQ-T7	2×150	6 Tb	4 h~9 d	适合于大型基因组 WGS、WES、表观基因组测序、 转录组测序和肿瘤 Panel

注:表中数据来自各测序平台公司官方网站(Illumina :<https://www.illumina.com/>;赛默飞:<https://www.thermofisher.com/cn/zh/home.html>;华大:<https://www.mgitech.cn/>)。

molecule real-time, SMRT) 技术^[17] 和纳米孔单分子测序技术 (Oxford nanopore technologies, ONT)^[18]。SMRT 技术可以实时记录被荧光标记的脱氧核苷酸的强度变化,这种荧光标记的脱氧核苷酸不会影响 DNA 聚合酶的活性,并且在荧光被切除之后,合成的 DNA 链和天然的 DNA 链完全一致^[19]。较之二代测序,SMRT 测序速度很快,每秒约 10 个 dNTPs,但是通量低,1 个 SMRT 芯片池上有 150 000 测序单元,但是只有 35 ~ 70 000 个发生有效测序;另外 SMRT 的连续碱基序列(continuous long read,CLR) 错误率高达 11% ~ 15%,但是不同于 NGS 偏向性的错误,SMRT 测序错误是随机的,可以通过足够的测序次数来纠

正,15 次测序 CLR 准确率超过 99%^[17]。ONT 是采用电泳技术驱动单个分子逐一通过纳米孔,根据 A、G、C、T 这 4 种碱基的带电性质不同,通过电信号的差异就能检测出通过的碱基类别,从而实现测序^[20-21]。ONT 单碱基错误率高,为随机错误,但可通过提高测序深度进而提高碱基准确率。相比于二代测序,长读长是三代测序的关键优势,而通量低、错误率高、单碱基成本高是三代测序的缺陷^[22]。对于临床检测可以采用二代测序高通量和准确度高的短读长片段对三代测序的长读长片段进行修正^[23],目前三代测序特别适合于全基因组 denovo 测序、甲基化研究、点突变检测、特殊基因区域检测等^[24-27]。

2 二代及三代测序技术在遗传学诊断中的应用

以高通量测序技术、三代测序技术为代表的现代分子生物学技术在遗传学领域的推广应用,正在逐渐改变医学遗传学的现状,无创产前筛查(non-invasive prenatal testing, NIPT)、染色体拷贝数变异检测、单基因病的筛查和诊断都是其应用领域。研究人员或临床医师可以根据不同的检测需求选择适合的测序方案,以提高疾病诊断率,为遗传病家庭带去福音,实现社会资源的最大化利用。

2.1 NGS 在 NIPT 的应用

2.1.1 胎儿染色体非整倍体异常 早在1997年,就有科学家证实妇女在怀孕期间其外周血循环有胎儿游离DNA (cell-free fetal DNA, cfDNA)的存在^[28];2008年Quake团队报道了cfDNA-NIPT技术(cfNIPT),该方法基于高通量测序平台对孕妇外周血中的cfDNA进行低深度全基因组测序,并结合生物信息学分析方法,完成了胎儿13/18/21号染色体非整倍体检测^[29],自此基于cfDNA的NIPT开始在临床中应用。目前,NIPT作为孕期常规筛查手段用于上述常染色体及性染色体非整倍体的筛查^[30]。国内主流的三大NIPT平台为:达安DA8600(基于Ion Proton测序平台)、贝瑞和康NextSeq CN500(与Illumina合作生产)以及华大基因的BGISEQ500。大规模人群研究证实NIPT对于13/18/21号染色体拷贝数异常的检测特异性均大于99.9%;检测敏感性方面T21最高,可达到99.5%,T18为93.1%,T13敏感性略低,为92.7%^[31]。多项研究均证明NIPT对于T13/18/21检出的准确率均高于传统血清学筛查^[32],是一种更适合于高危人群的产前筛查的选择;而对于大众群体来说,NIPT的检测成本与实际效果还需要从政府、社会和个人角度等进一步明确^[33-34]。

2.1.2 胎儿染色体微缺失和微重复 染色体亚结构异常(微缺失和微重复)会导致胎儿的身体或智力发育严重异常,在大约6%的B超结果异常和1.7%的高龄或唐氏综合征筛查阳性的孕妇中存在染色体微缺失和微重复,且其发生率与孕

母年龄无明显相关性^[30]。因此,对所有年龄段孕妇均应在产前行胎儿染色体微缺失和微重复检测。近年来,NIPT也被用于染色体微缺失微重复疾病的筛查^[30]。张春花等^[35]对国内三大NIPT平台染色体微缺失、微重复检测效果进行评估,从数据量、阳性预测值、假阳性率方面比较后发现,BGISEQ500有效数据量和阳性筛查率明显高于DA8600和NextSeq CN500,三大平台的阳性预测值没有显著差异。随着高通量测序的发展和优化,单样本的检测成本在不断下降,对人类基因组变异与疾病间关系的进一步明确和深入,最终以NIPT取代血清学筛查是产前筛查发展的必然趋势,并且NIPT在微缺失/微重复综合征筛查方面的应用将会更加广泛。

2.1.3 胎儿单基因病 单基因病无创产前筛查也是NIPT的重要研究方向,已有亨廷顿舞蹈病、地中海贫血、囊性纤维化、肝豆状核变性、杜氏肌营养不良的NIPT报道^[36-39]。Wang等^[40]针对 α/β 地中海贫血的无创产前检测开发了一套针对HBA1、HBA2和HBB基因的SNP探针,收集地中海贫血家系的gDNA及胎儿游离DNA并对目标区域靶向测序,使用隐马尔科夫模型(hidden Markov model, HMM)和Viterbi算法构建亲本单体型来分析胎儿基因型,从而完成地中海贫血家系的产前诊断。基于二代测序平台的单基因病NIPT需要依赖生物信息学的算法,如相对突变剂量(relative mutation dosage, RMD)和相对单倍剂量(relative haplotype dosage, RHDO),来构建亲本、先证者以及胎儿单倍型来对连锁位点进行检测^[41-42],其准确性还需要更多人群的大规模验证,对于临床应用的实际效用尚需谨慎讨论。另外,采用10×genomics测序进行长读长测序也是单基因病NIPT研究的重要手段。Jang等^[39]采用10×genomics靶向linked-reads测序技术,将单个长片段的DNA分子分配至不同的油滴中,并通过GEM建库技术,将长片段DNA切割成适合测序的大小,并且来源于相同油滴(同一条长片段DNA)的DNA片段会带上相同的DNA序列标记(barcode),经测序拼接后理论上可以得到原先的长片段DNA序列。研究人员收集了5名DMD突变携带者孕妇的血浆DNA并进行靶向linked-reads测序分析,最终成功判断胎儿基因型以及目标区域的染色体重组情况。该方法能够提示

DMD 基因大片段缺失、重复,可以直接对胎儿游离 DNA 进行单倍体分析而无需对先证者 gDNA 测序,但其建库费用十分昂贵,检测周期更长(3 周),而且单基因病 NIPT 的临床适用性研究较少,该方法最终是否可以在临床上推广应用还需进一步探讨。

2.2 NGS 在染色体病诊断中的应用

我国每年约有 10 万染色体异常患儿出生,每 1 000 个活产婴中就有 5~10 个发生染色体异常^[43]。在基因测序出现之前,临床上针对遗传病主要是在细胞水平对染色体行 G 显带核型分析,该方法可以直观地观察染色体数目及结构变异,是检测染色体异常的金标准^[44-45]。然而,核型分析细胞培养耗时长、效率低,而且仅能分辨 5~10 Mb 以上大小的结构变异,对染色体的微缺失、微重复的分辨不足,且不能精确定位变异染色体位置和大小。低深度全基因组测序(copy number variation sequencing, CNV-seq)的出现弥补了核型分析分辨率不足的缺陷^[46],是高通量测序在染色体拷贝数变异检测中的重要应用。CNV-seq 通过低深度全基因组测序及下游生物信息学分析对染色体拷贝数异常检测,分辨率可达 100 kb。Zhang 等^[47]对 160 名妊娠高风险或遗传病高风险人群进行 CNV-seq 分析,结果显示近一半(44.6%)的高危妊娠孕妇检出染色体异常,不良孕产史和遗传病家族史个体中检出染色体异常的比例更高,分别为 67.6% 和 55%。鉴于 CNV-Seq 对于染色体微缺失、微重复检测的高度特异性和一致性,其已逐渐成为了诊断染色体病的首选方法^[48]。目前临床上已将 CNV-seq 用于染色体病的产前诊断、携带者筛查以及植入前遗传学诊断^[35]。

2.3 NGS 及三代测序在单基因病诊断中的应用

NGS 按照其对基因组的检测范围可分为靶向测序(target gene sequencing, TGS)、全外显子测序(whole exome sequencing, WES)、全基因组测序(whole genome sequencing, WGS)。TGS 是通过选取部分与待检疾病密切相关的基因进行目标区域富集测序,单次检测价格更为低廉,易被患者所接受,在内分泌疾病、神经肌肉病等系统性疾病中应用较多。WES 是利用序列捕获技术将全外显子区域 DNA 捕捉并富集后进行高通量测序的基因分析方法,其对遗传性疾病的临床分子诊断率约

为 25%~50%^[49-51]。由于外显子组序列约占人类全部基因组序列的 1%,但其包含了大约 95% 的致病突变^[52],因此目前对于大部分单基因遗传病来说, WES 是最具性价比的 NGS 手段,而且 WES 具有可回溯性,随着疾病与基因之间的关联认知逐渐深入, WES 数据的重分析对于指导临床诊断具有极大的价值。WGS 是对人类所有基因组信息(包括蛋白质编码区以及大量非编码区)进行测序,具有测序面广、检测基因变异类型全等优点,但过高的价格限制了其在临检中的应用,目前主要用于科研用途。这 3 种方式各有利弊,在临床中应根据每种测序技术的特点和局限性以及临床目的来选择最适合的测序策略。

按照单基因病的致病机制,可将基因组序列变异类型分为单核苷酸变异(single nucleotide variant, SNV)、结构变异(structural variants, SV)、拷贝数变异(copy number variations, CNV)、插入/缺失(insertion-deletion, InDel)以及动态突变、甲基化异常等。NGS 尤其适合于 SNV 及 InDel 的检测,但受限于其技术特点,NGS 对于高 GC 含量或碱基高度重复区域的检测并不准确,因此对于特殊致病机制的单基因病,例如动态突变的检测,可以采用三代测序以弥补 NGS 的缺陷。对于发生于基因组外显子区或明确靶区域的 SNV 或 InDel,可以选择 WES 或 TGS;发生于全基因组的 SNV、InDel 可以选择 WGS^[53];对于复杂的染色体平衡异位或 SV、CNV,NGS 并非最佳检测手段,应首选染色体微阵列分析(chromosomal microarray analysis, CMA)、核型分析或 CNV-Seq 以明确发生变异的染色体,再根据细胞遗传学结果和临床需要采用 WGS 或三代测序进行断点分析。

2.3.1 SNV、Indel 的检测 对于临床诊断较为明确、致病基因有限或突变机制复杂的单基因病,应用 TGS 能够保证足够的测序深度,不仅能够反映 SNV、Indel,还能一定程度上检出 CNV 等复杂致病变异^[54],使临床检测具有更高的针对性;对于临床症状不典型、难以确诊的遗传病, WES 较 TGS 有更大的覆盖范围,能从全外显子组的水平对基因功能区进行评估,有利于快速地诊断。先天性甲状腺功能减低症(congenital hypothyroidism, CH)是一种常见的新生儿内分泌疾病,其中甲状腺发育异常是造成 CH 的最主要原因,孔静等^[55]针对甲状腺发育异常相关致病基

因的热点突变的外显子及外显子-内含子交界区域设计靶向 panel,对 89 例 CH 患儿进行检测,最终有 19 例患者检出潜在致病突变,17 例为 *TSH* 基因突变,其中有 4 个为新发突变。WES 尤其适合单基因病的 SNV、Indel 的临床检测,其可回溯性对于提高临床诊断率具有重要作用。此外,有研究者对比了 WES 和 WGS 在 SNV 和 Indel 的检测中的表现,发现两者都检出了编码区域中超过 96.0% 的高质量 SNV 和 78.0% 的高质量 Indel,但是在 SNV 检测方面 WGS 比 WES 更为强大,其假阳性率分别为 17% (WGS) 和 78% (WES),而且从覆盖深度 (coverage depth, CD)、基因型质量值 (genotype quality, GQ) 和低覆盖区域比例 (minor-read ratio, MRR) 等参数的总体分布来看, WGS 能够提供更为优质均一的测序质量^[56]。

2.3.2 CNV 的检测 WES 和 WGS 可以从基因组全局水平提示基因组大片段 CNV,该优势可以提高致病位点检出率。Pfundt 等^[57]对 2 603 例遗传性患者的 WES 数据进行分析,最终确定了 123 个致病性 CNV,大小从 727 bp 到 15.3 Mb 不等,总体诊断率提高了约 2%。随着对基因与疾病认识的不断深入,相关数据库也在更新和完善,未明确诊断的患者的 WES 数据重分析对于提高临床诊断率也不容忽视。Ewans 等^[58]对来自 37 个家系共 54 名遗传病患者的 WES 进行了重新分析以鉴定其遗传病因,包括对患者表型的准确分析、内部变异数据库检索以及新疾病基因的鉴定,最终明确了 4 个家系的遗传学病因,整体诊断率由最初的 30% 提高到 41%。另外从成本效益角度分析, WES 与传统诊断手段相比可以大大节省遗传病的诊断成本。研究人员建议对于遗传病患者应尽早进行基因组测序,以便为后续疾病诊断、治疗乃至生育指导提供依据。

2.3.3 SV 的检测 结构变异 (structural variants, SV) 的鉴定对于疾病的遗传学诊断也尤为重要,约有 5%~20% 的遗传病由 SV 致病^[59-62],核型分析和染色体微阵列虽然价格低廉,但不能检出准确的断点信息以及小片段 SV。理论上 WGS 能提供全部基因组信息, Currall 等^[63]开发了一种针对大片段 SV (>5 000~10 000 bp) 的特殊 WGS 建库技术——large inserts 文库 (large inserts libraries, liWGS) 来最大程度地提高覆盖率。染色体平衡异常 (balanced chromosomal abnormalities, BCA) 常

用核型分析和 CMA 来检测,有研究表明有多达 40% 的 BCA 可以被 CMA 作为非平衡异常而检出,因此对于复杂结构变异或平衡异常,可以通过传统细胞遗传学手段检测后再采用 liWGS 进一步分析断点信息,提高 SV 检出率。Redin 等^[64]对 273 名预先经过 CMA 确认的 BCA 患者采用 liWGS 测序,最终有 90.8% 患者明确了其平衡异常的断点,但有 9% 的断点无法检出。

随着测序技术和分析能力的提高,三代测序通过长读长检测可以直接反映基因组结构变异,成为了 SV 鉴别的重要手段。Stancu 等^[65]使用 MinION 纳米孔测序仪和 NanoSV 算法分析了 2 名具有复杂表型的先天性异常患者基因组,其中对于 2 号患者检出了 29 个新发 SV,提示了 1、5 和 9 号染色体的复杂重排,与 46, XY, t(1;9;5) (complex) dn 相吻合,该研究证明,可以利用纳米孔测序来确定染色质断点的亲本起源,其在新发染色质重排检测方面优于 NGS。

2.3.4 动态突变的检测 随着基因检测技术的发展,三代测序技术的应用逐步提升,目前有研究将其用于全基因组短串联重复变异检测。脆性 X 综合征 (fragile X syndrome, FXS) 是遗传性智力障碍和自闭症的最常见原因之一,是由 *FMR1* 基因 (CGG)_n 区域重复次数过多 (大于 200 次) 而致。该区域为大量 GC 碱基重复,短读长的 NGS 对位于此类区域的读取表现不佳,为了解决该问题, Loomis 等^[59]采用 SMRT 技术对 FXS 患者 *FMR1* 基因 (CGG)_n 区域进行检测,并且最终完成了长达 2 kb 的 (CGG)_n 重复片段鉴定,其检出的重复数高达 750 次。不同突变类型的检测方式及测序平台选择见表 2。

3 展望

自 NGS 发明以来,相应的测序平台不断推陈出新,技术平台的优化和成本的急速降低促成了其在临床检测应用的不断扩大。NGS 技术极大地推动了精准医学的发展,全面地改变了遗传学领域的发展面貌,也使遗传学分子诊断迎来了一个新的技术时代。然而,不断增加的测序需求量也对 NGS 平台的硬件及软件提出了更高的要求,对人类基因组测序的能力已大大超过遗传变异的解读能力,冗余的测序数据对数据管理、分析及变

表 2 不同变异类型的检测方式及测序平台选择

Table 2 Detection methods and sequencing platform selections of different variations

变异类型	测序策略	可选择的测序平台
SNV	TGS/WES/WGS	HiSeq、NextSeq 500、Ion Proton、BGISEQ-500 等
Indel	TGS/WES/WGS	HiSeq、NextSeq 500、Ion Proton、BGISEQ-500 等
CNV	WGS/WES	HiSeq、NextSeq 500、Ion Proton、BGISEQ-500 等
SV	WGS/三代测序	HiSeq、NextSeq 500、Ion Proton、BGISEQ-500 ONT 测序
动态突变	三代测序	SMRT 测序

异解读、数据库的建立及更新提出了严峻的挑战^[66]。NGS 产生的大量原始序列数据依赖于生物信息学的分析,因此,生物信息技术对于 NGS 起着至关重要的作用。主要生物信息分析步骤均包括原始数据的质量控制、数据对比、对比后处理、变异识别、注释、过滤及解读,不同的 NGS 平台所产生的测序数据类型有一定的差异,如 Illumina 和华大 BGISEQ 平台产生的测序数据主要包含序列信息和相应的碱基质量值,而 Ion Torrent 平台产生的测序数据除了以上信息外,还包含测序碱基信号值(flow signal)^[12]。目前,应用于临床基因检测的生物信息学分析流程都是实验室内部自建或第三方服务商提供,不同的测序平台和测序范围的数据分析流程及所使用的软件有所差别,为了保证检测的准确性和可靠性,各个实验室在建立生物信息分析流程时,需要设置必要的质控标准包括 reads 质量、GC 含量、比对率、reads 比对质量以及测序深度,还要明确检测范围、准确度、分析敏感性、特异性等来确认生物信息分析的性能,以保证达到预期的检测目的^[67-68]。

NGS 在不同领域的应用标准、伦理问题及检测后的遗传咨询规范等也需要进一步探讨和完善^[69],2017 年唐北沙等^[53]制定了符合我国现实情况的人类孟德尔遗传性疾病基因组序列变异解析规范,从临床资料采集、二代测序选择、质控管理、生物信息分析、变异解读原则及伦理学等方面对于遗传病的变异解读进行了详细的阐述。NGS 的高速发展伴随着机遇和挑战,随着其技术的不断升级优化和临床经验的不断积累,通过不同领域专家共同协作以建立更加精准的“基因-疾病”解读指南是未来遗传学检测的发展方向,NGS 终将会成为遗传学领域主要且常规的检测手段。

三代测序是人类遗传学领域的重要里程碑,其特别适合于复杂的基因组组装、串联重复以及复杂的结构重排、单倍型构建等,但高昂的价格限制了其在常规临检的应用,目前仅用于科研用途。三代测序技术在读取的准确性、相应的分析算法以及检测成本方面需要进一步优化和提升,在未来可以与 NGS 互为补充,为临床遗传学服务^[70]。

参 考 文 献

[1] HE C, LIU L, CHU Y, *et al.*. National and subnational all-cause and cause-specific child mortality in China, 1996–2015: a systematic analysis with implications for the Sustainable Development Goals[J]. *Lancet Glob. Health*, 2017, 5(2):e186–e197.

[2] 吴怡, 程蔚蔚. 出生缺陷概况及产前筛查[J]. *中国计划生育和妇产科*, 2016 (1):29–33, 52.

[3] YI H, HALLOWELL N, GRIFFITHS S, *et al.*. Motivations for undertaking DNA sequencing-based non-invasive prenatal testing for fetal aneuploidy: a qualitative study with early adopter patients in Hong Kong[J/OL]. *PLoS ONE*, 2013, 8 (11): e81794 [2020–11–01]. <https://doi.org/10.1371/journal.pone.0081794>.

[4] JUVET L K, ORMSTAD S S, STOINSKA-SCHNEIDER A, *et al.*. Non-invasive prenatal test (NIPT) for identification of trisomy 21, 18 and 13: 2016–18[R/OL]. (2016–12–16) [2020–11–24]. <https://pubmed.ncbi.nlm.nih.gov/29553653/>.

[5] KIRCHER M, KELSO J. High-throughput DNA sequencing—concepts and limitations[J]. *Bioessays*, 2010, 32 (6): 524–536.

[6] 刘亚兰, 蒋小红, 孙捷, 等. 实时荧光 PCR 熔解曲线法在非综合征型耳聋基因检测中的应用研究[J]. *中华耳鼻咽喉头颈外科杂志*, 2019(4): 286–291.

[7] MCGINN S, GUT I G. DNA sequencing—spanning the generations[J]. *New Biotechnol.*, 2013, 30(4):366–372.

[8] GRADA A, WEINBRECHT K. Next-generation sequencing: methodology and application[J/OL]. *J. Invest. Dermatol.*, 2013, 133(8):e11[2020–11–01]. <https://doi.org/10.1038/jid.2013.248>.

[9] SEO T S, BAI X, KIM D H, *et al.*. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides[J]. *Proc. Natl. Acad. Sci. USA*, 2005, 102(17): 5926–5931.

- [10] RONAGHI M, KARAOHAMED S, PETERSSON B, *et al.* Real-time DNA sequencing using detection of pyrophosphate release[J]. *Anal. Biochem.*, 1996, 242(1):84-89.
- [11] BENTLEY D R, BALASUBRAMANIAN S, SWERDLOW H P, *et al.* Accurate whole human genome sequencing using reversible terminator chemistry[J]. *Nature*, 2008, 456(7218):53-59.
- [12] ROTHBERG J M, HINZ W, REARICK T M, *et al.* An integrated semiconductor device enabling non-optical genome sequencing[J]. *Nature*, 2011, 475(7356):348-352.
- [13] ZHU F Y, CHEN M X, YE N H, *et al.* Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants [J/OL]. *Plant Methods*, 2018, 14:69 [2020-11-01]. <https://doi.org/10.1186/s13007-018-0337-0>.
- [14] RUPAREL H, BI L, LI Z, *et al.* Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis [J]. *Proc. Natl. Acad. Sci. USA*, 2005, 102(17):5932-5937.
- [15] MAK S, GOPALAKRISHNAN S, CARØE C, *et al.* Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing [J]. *Gigascience*, 2017, 6(8):1-13.
- [16] HUANG J, LIANG X, XUAN Y, *et al.* A reference human genome dataset of the BGISEQ-500 sequencer [J]. *Gigascience*, 2017, 6(5):1-9.
- [17] RHOADS A, AU K F. PacBio sequencing and its applications [J]. *Genom. Proteom. Bioinf.*, 2015, 13(5):278-289.
- [18] QUICK J, LOMAN N J, DURAFFOUR S, *et al.* Real-time, portable genome sequencing for Ebola surveillance[J]. *Nature*, 2016, 530(7589):228-232.
- [19] CARNEIRO M O, RUSS C, ROSS M G, *et al.* Pacific biosciences sequencing technology for genotyping and variation discovery in human data[J/OL]. *BMC Genomics*, 2012, 13:375 [2020-11-01]. <https://doi.org/10.1186/1471-2164-13-375>.
- [20] LAVER T, HARRISON J, O'NEILL P A, *et al.* Assessing the performance of the Oxford nanopore technologies MinION[J]. *Biomol. Detect. Quantif.*, 2015, 3:1-8.
- [21] SALVÀ-SERRA F, JAÉN-LUCHORO D, JAKOBSSON H E, *et al.* Complete genome sequences of *Streptococcus pyogenes* type strain reveal 100%-match between PacBio-solo and Illumina-Oxford nanopore hybrid assemblies [J/OL]. *Sci. Rep.*, 2020, 10:11656 [2020-11-01]. <https://doi.org/10.1038/s41598-020-68249-y>.
- [22] MIZUGUCHI T, TOYOTA T, ADACHI H, *et al.* Detecting a long insertion variant in SAMD12 by SMRT sequencing: implications of long-read whole-genome sequencing for repeat expansion diseases[J]. *J. Hum. Genet.*, 2019, 64(3):191-197.
- [23] CHAISSON M J, HUDDLESTON J, DENNIS M Y, *et al.* Resolving the complexity of the human genome using single-molecule sequencing[J]. *Nature*, 2015, 517(7536):608-611.
- [24] ISTACE B, FRIEDRICH A, D'AGATA L, *et al.* De novo assembly and population genomic survey of natural yeast isolates with the Oxford nanopore MinION sequencer[J]. *Gigascience*, 2017, 6(2):1-13.
- [25] CHAISSON M, SANDERS A D, ZHAO X, *et al.* Multi-platform discovery of haplotype-resolved structural variation in human genomes [J/OL]. *Nat. Commun.*, 2019, 10:1784 [2020-11-01]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6467913/>. DOI: 10.1038/s41467-018-08148-z.
- [26] SIMPSON J T, WORKMAN R E, ZUZARTE P C, *et al.* Detecting DNA cytosine methylation using nanopore sequencing [J]. *Nat. Methods*, 2017, 14(4):407-410.
- [27] KULESHOV V, XIE D, CHEN R, *et al.* Whole-genome haplotyping using long reads and statistical methods[J]. *Nat. Biotechnol.*, 2014, 32(3):261-266.
- [28] LO Y M, CORBETTA N, CHAMBERLAIN P F, *et al.* Presence of fetal DNA in maternal plasma and serum[J]. *Lancet*, 1997, 350(9076):485-487.
- [29] FAN H C, BLUMENFELD Y J, CHITKARA U, *et al.* Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood[J]. *Proc. Natl. Acad. Sci. USA*, 2008, 105(42):16266-16271.
- [30] WAPNER R J, BABIARZ J E, LEVY B, *et al.* Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes [J/OL]. *Am. J. Obstet. Gynecol.*, 2015, 212(3):332 [2020-11-02]. <https://pubmed.ncbi.nlm.nih.gov/25479548/>. DOI: 10.1016/j.ajog.2014.11.041.
- [31] Health Quality Ontario. Noninvasive prenatal testing for trisomies 21, 18, and 13, sex chromosome aneuploidies, and microdeletions: a health technology assessment [J/OL]. *Ont. Health Technol. Assess. Ser.*, 2019, 19(4):1-166 [2020-11-24]. <https://pubmed.ncbi.nlm.nih.gov/30847010/>. PMID: 30847010; PMCID: PMC6395059.
- [32] BIANCHI D W, PLATT L D, GOLDBERG J D, *et al.* Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing[J]. *Obstet. Gynecol.*, 2012, 119(5):890-901.
- [33] WALKER B S, NELSON R E, JACKSON B R, *et al.* A cost-effectiveness analysis of first trimester non-invasive prenatal screening for fetal trisomies in the United States[J/OL]. *PLoS ONE*, 2015, 10(7):e0131402 [2020-11-02]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489811/>. DOI: 10.1371/journal.pone.0131402.
- [34] LÉGARÉ F, ST-JACQUES S, GAGNON S, *et al.* Prenatal screening for Down syndrome: a survey of willingness in women and family physicians to engage in shared decision-making[J]. *Prenat. Diagn.*, 2011, 31(4):319-326.
- [35] 张春花, 贺权泽, 乔龙威, 等. 3类无创产前筛查平台检测胎儿染色体微缺失微重复效果分析[J]. *中国计划生育学杂志*, 2020, 28(2):203-207.
- [36] LYU W, WEI X, GUO R, *et al.* Noninvasive prenatal testing for Wilson disease by use of circulating single-molecule amplification and resequencing technology (cSMART) [J]. *Clin. Chem.*, 2015, 61(1):172-181.
- [37] GONZÁLEZ-GONZÁLEZ M C, GARCÍA-HOYOS M, TRUJILLO M J, *et al.* Prenatal detection of a cystic fibrosis mutation in fetal DNA from maternal plasma[J]. *Prenat. Diagn.*, 2002, 22(10):946-948.
- [38] GONZÁLEZ-GONZÁLEZ M C, TRUJILLO M J, RODRÍGUEZ DE ALBA M, *et al.* Huntington disease-unaffected fetus diagnosed from maternal plasma using QF-PCR [J]. *Prenat. Diagn.*, 2003, 23(3):232-234.
- [39] JANG S S, LIM B C, YOO S K, *et al.* Targeted linked-read sequencing for direct haplotype phasing of maternal DMD alleles: a practical and reliable method for noninvasive prenatal diagnosis[J/OL]. *Sci. Rep.*, 2018, 8(1):8678 [2020-11-02]. <https://pubmed.ncbi.nlm.nih.gov/29875376/>. DOI: 10.1038/s41598-018-26941-0.

- [40] WANG W, YUAN Y, ZHENG H, *et al.*. A pilot study of non-invasive prenatal diagnosis of alpha- and beta-thalassemia with target capture sequencing of cell-free fetal DNA in maternal blood[J]. *Genet. Test. Mol. Biomarkers*, 2017, 21(7):433-439.
- [41] VERMEULEN C, GEEVEN G, DE WIT E, *et al.*. Sensitive monogenic noninvasive prenatal diagnosis by targeted haplotyping[J]. *Am. J. Hum. Genet.*, 2017, 101(3):326-339.
- [42] LO Y M, CHAN K C, SUN H, *et al.*. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus[J/OL]. *Sci. Transl. Med.*, 2010, 2(61):61ra91[2020-11-02]. <https://stm.sciencemag.org/content/2/61/61ra91.long>. DOI: 10.1126/scitranslmed.3001720.
- [43] 马慧. 优生检查中增加染色体核型分析的重要性[J]. *临床医药文献电子杂志*, 2016, 3(23):4553-4554.
- [44] 李燕青, 刘夏莹, 谢俊杰, 等. 超声异常胎儿的染色体核型分析及单核苷酸多态性芯片检测应用分析[J]. *国际妇产科学杂志*, 2020, 47(3):286-290.
- [45] 余宏盛, 郭红, 沈双双, 等. 染色体核型分析联合 SNP-array 技术在不良妊娠史孕妇产前诊断中的应用[J]. *中华妇产科杂志*, 2018, 53(3):155-159.
- [46] XIE C, TAMMI M T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing[J/OL]. *BMC Bioinformatics*, 2009, 10:80[2020-11-02]. <https://doi.org/10.1186/1471-2105-10-80>.
- [47] ZHANG R, CHEN X, WANG D, *et al.*. Prevalence of chromosomal abnormalities identified by copy number variation sequencing in high-risk pregnancies, spontaneous abortions, and suspected genetic disorders[J]. *J. Int. Med. Res.*, 2019, 47(3):1169-1178.
- [48] LIANG D, PENG Y, Lyu W, *et al.*. Copy number variation sequencing for comprehensive diagnosis of chromosome disease syndromes[J]. *J. Mol. Diagn.*, 2014, 16(5):519-526.
- [49] YANG Y, MUZYNY D M, REID J G, *et al.*. Clinical whole-exome sequencing for the diagnosis of mendelian disorders[J]. *N. Engl. J. Med.*, 2013, 369(16):1502-1511.
- [50] YANG Y, MUZYNY D M, XIA F, *et al.*. Molecular findings among patients referred for clinical whole-exome sequencing[J]. *JAMA*, 2014, 312(18):1870-1879.
- [51] BRAUN D A, SCHUELER M, HALBRITTER J, *et al.*. Whole exome sequencing identifies causative mutations in the majority of consanguineous or familial cases with childhood-onset increased renal echogenicity[J]. *Kidney Int.*, 2016, 89(2):468-475.
- [52] SHAMSELDIN H E, MADDIREVULA S, FAQEIH E, *et al.*. Increasing the sensitivity of clinical exome sequencing through improved filtration strategy[J]. *Genet. Med.*, 2017, 19(5):593-598.
- [53] 唐北沙, 曾胜, 李凯. 人类孟德尔遗传性疾病基因组序列变异解析与临床规范[J]. *中国现代神经疾病杂志*, 2017, 17(7):471-476.
- [54] 陈召. 二代测序在几个神经系统罕见孟德尔病的基因诊断中的应用[D]. 长沙:中南大学, 硕士学位论文, 2013.
- [55] 孔静, 戴静宜, 龙伟, 等. 先天性甲状腺发育异常相关致病基因的突变分析[J]. *重庆医学*, 2019, 48(12):2087-2089, 2094.
- [56] BELKADI A, BOLZE A, ITAN Y, *et al.*. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants[J]. *Proc. Natl. Acad. Sci. USA*, 2015, 112(17):5473-5478.
- [57] PFUNDT R, DEL ROSARIO M, VISSERS L, *et al.*. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders[J]. *Genet. Med.*, 2017, 19(6):667-675.
- [58] EWANS L J, SCHOFIELD D, SHRESTHA R, *et al.*. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders[J]. *Genet. Med.*, 2018, 20(12):1564-1574.
- [59] LOOMIS E W, EID J S, PELUSO P, *et al.*. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene[J]. *Genome Res.*, 2013, 23(1):121-128.
- [60] CARSS K J, ARNO G, ERWOOD M, *et al.*. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease[J]. *Am. J. Hum. Genet.*, 2017, 100(1):75-90.
- [61] COOPER G M, COE B P, GIRIRAJAN S, *et al.*. A copy number variation morbidity map of developmental delay[J]. *Nat. Genet.*, 2011, 43(9):838-846.
- [62] MILLER D T, ADAM M P, ARADHYA S, *et al.*. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies[J]. *Am. J. Hum. Genet.*, 2010, 86(5):749-764.
- [63] CURRALL B B, ANTOLIK C W, COLLINS R L, *et al.*. Next generation sequencing of prenatal structural chromosomal rearrangements using large-insert libraries[J]. *Methods Mol. Biol.*, 2019, 1885:251-265.
- [64] REDIN C, BRAND H, COLLINS R L, *et al.*. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies[J]. *Nat. Genet.*, 2017, 49(1):36-45.
- [65] CRETU STANCU M, VAN ROOSMALEN M J, RENKENS I, *et al.*. Mapping and phasing of structural variation in patient genomes using nanopore sequencing[J/OL]. *Nat. Commun.*, 2017, 8(1):1326[2020-11-02]. <https://pubmed.ncbi.nlm.nih.gov/29109544/>. DOI: 10.1038/s41467-017-01343-4.
- [66] DAHLÖ M, SCOFIELD D G, SCHAAL W, *et al.*. Tracking the NGS revolution: managing life science research on shared high-performance computing clusters[J]. *Gigascience*, 2018, 7(5):giy028. <https://doi.org/10.1093/gigascience/giy028>.
- [67] 孙隽, 黄颐, 王小冬, 等. 遗传病二代测序临床检测全流程规范化共识探讨(3)——数据分析流程[J]. *中华医学遗传学杂志*, 2020(3):345-351.
- [68] 黄辉, 沈亦平, 顾卫红, 等. 临床基因检测报告规范与基因检测行业共识探讨[J]. *中华医学遗传学杂志*, 2018, 35(1):1-8.
- [69] ZHANG X, LIANG Z, WANG S, *et al.*. Application of next-generation sequencing technology to precision medicine in cancer: joint consensus of the tumor biomarker committee of the chinese society of clinical oncology[J]. *Cancer Biol. Med.*, 2019, 16(1):189-204.
- [70] MIDHA M K, WU M, CHIU K P. Long-read sequencing in deciphering human genetics to a greater depth[J]. *Hum. Genet.*, 2019, 138:1201-1215.