

# 基于脑瘫患儿的外显子组测序: TruSeq 与 NimbleGen 试剂盒的捕获性能比较

王妍龚<sup>1</sup> 成业<sup>1</sup> 刘洋<sup>1</sup> 许毅然<sup>2</sup> 邢清和<sup>1Δ</sup>

(<sup>1</sup>复旦大学生物医学研究院 上海 200032; <sup>2</sup>郑州大学第三附属医院神经研究所 郑州 450052)

**【摘要】** 目的 比较 TruSeq<sup>®</sup> Exome 与 NimbleGen SeqCap EZ Human Exome 两种外显子捕获试剂在脑瘫患儿中的捕获性能差异,为临床遗传学研究和诊断提供技术选择依据。方法 纳入 48 例散发脑瘫患儿外周血样本,分别采用 TruSeq(DNA 探针)和 NimbleGen(RNA 探针)构建外显子组文库,经 Illumina HiSeq 2000 平台测序。通过生物信息学分析评估比对率、目标区域覆盖度、变异检出一致性等指标,并基于脑瘫相关基因集(2 293 个基因)分析捕获性能的临床相关性,采用配对 *t* 检验进行统计学分析(显著性阈值  $\alpha=0.05$ )。结果 NimbleGen 和 TruSeq 两种外显子组捕获试剂盒在基础数据质量(比对率、插入片段长度)和 GC 含量上差异无统计学意义。然而,在关键性能指标上呈现互补特征,NimbleGen 在特定深度覆盖上表现更优(1×覆盖率,  $P=1.84 \times 10^{-5}$ ; 20×覆盖率,  $P=1.49 \times 10^{-20}$ );而 TruSeq 则展现出更高的 Indel 检测灵敏度(TruSeq vs. NimbleGen:  $11\ 371 \pm 1\ 689$  vs.  $11\ 274 \pm 1\ 670$ ,  $P=3.24 \times 10^{-7}$ )和罕见变异捕获能力(TruSeq vs. NimbleGen:  $3\ 164 \pm 766$  vs.  $3\ 072 \pm 774$ ,  $P=1.20 \times 10^{-4}$ ),并成功检出所有 11 个阳性致病变异(包括 NimbleGen 漏检的 2 例)。结论 TruSeq 凭借更优的变异检出率更适合临床诊断场景,而 NimbleGen 的覆盖稳定性可能有利于研究性项目。

**【关键词】** 全外显子组测序(WES); 捕获; 试剂盒; 脑瘫

**【中图分类号】** Q81,R742.3 **【文献标志码】** A **doi:** 10.3969/j.issn.1672-8467.2025.06.012

## Exome sequencing in children with cerebral palsy: a comparison of capture performance between TruSeq and NimbleGen kits

WANG Yan-gong<sup>1</sup>, CHENG Ye<sup>1</sup>, LIU Yang<sup>1</sup>, XU Yi-ran<sup>2</sup>, XING Qing-he<sup>1Δ</sup>

(<sup>1</sup>Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China; <sup>2</sup>Institute of Neuroscience, the Third Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, Henan Province, China)

**【Abstract】 Objective** To compare the capture performance differences between TruSeq<sup>®</sup> Exome and NimbleGen SeqCap EZ Human Exome kits in children with cerebral palsy (CP), and to provide a technical selection basis for clinical genetic research and diagnosis. **Methods** Peripheral blood samples from 48 sporadic CP patients were included. Exome libraries were constructed using TruSeq (DNA probes) and NimbleGen (RNA probes), followed by sequencing on the Illumina HiSeq 2000 platform. Bioinformatics analysis was applied to evaluate mapping rate, target region coverage, variant concordance, and clinical relevance based on a CP-related gene set (2 293 genes). The statistical analysis was performed using a paired *t*-test with a significance threshold of  $\alpha=0.05$ . **Results** The results showed no significant differences between NimbleGen and TruSeq exome capture kits in basic data quality (alignment rate, insert size) and GC content. However, they exhibited complementary characteristics in key performance

国家自然科学基金面上项目(32170615)

<sup>Δ</sup>Corresponding author E-mail: qhxing@fudan.edu.cn

网络首发时间:2025-10-31 14:49:21 网络首发地址: <https://link.cnki.net/urlid/31.1885.R.20251030.1527.006>

metrics; NimbleGen demonstrated superior performance in specific depth coverage ( $1\times$  coverage rate,  $P=1.84\times 10^{-5}$ ;  $20\times$  coverage rate,  $P=1.49\times 10^{-20}$ ). TruSeq, on the other hand, showed higher sensitivity in Indel detection (TruSeq vs. NimbleGen:  $11\ 371\pm 1\ 689$  vs.  $11\ 274\pm 1\ 670$ ,  $P=3.24\times 10^{-7}$ ) and rare variant capture (TruSeq vs. NimbleGen:  $3\ 164\pm 766$  vs.  $3\ 072\pm 774$ ,  $P=1.20\times 10^{-4}$ ), successfully identifying all 11 pathogenic variants (including 2 missed by NimbleGen). **Conclusion** TruSeq, with its superior variant detection rate, is more suitable for clinical diagnostic applications, while NimbleGen's coverage stability may be advantageous for research-oriented projects.

**【Key words】** whole exome sequencing (WES); capture; kit; cerebral palsy

\* This work was supported by the General Program of National Natural Science Foundation of China (32170615).

随着基因组学与生物信息学的迅猛发展,下一代测序技术(next generation sequencing, NGS)已成为当代生物医学研究的核心工具之一<sup>[1]</sup>。得益于NGS的广泛应用,每年约有250种单基因病的致病基因被成功鉴定。外显子是基因组中编码蛋白质的功能区域,尽管仅占整个基因组的1%~2%,却承载了约85%的已知孟德尔遗传病相关突变<sup>[2]</sup>,并富集了大量与疾病易感性相关的单核苷酸多态性(single nucleotide polymorphisms, SNPs)<sup>[3]</sup>。全外显子组测序(whole exome sequencing, WES)作为NGS的重要分支,专注于分析生物体内外显子区域的DNA序列。目前,WES已广泛应用于遗传疾病诊断、致病基因鉴定、个性化医学以及基础研究,逐渐成为医学研究与临床诊断中不可或缺的重要工具<sup>[4]</sup>。

然而,WES的检测性能高度依赖于外显子捕获探针的设计。不同技术(如DNA与RNA探针)在覆盖均一性、鸟嘌呤-胞嘧啶含量(guanine-cytosine content, GC)偏好性及边界捕获效率等方面存在差异<sup>[5]</sup>,直接影响变异检出灵敏度与临床诊断率。目前虽已有研究比较不同平台的通用性能,但其结论多基于标准样本或全外显子整体指标,对于遗传异质性极高的复杂疾病,不同平台针对特定疾病相关基因集的捕获效能是否存在差异,尚缺乏系统研究。

脑性瘫痪(cerebral palsy, CP, 后简称脑瘫)是成因复杂的非进行性运动障碍综合征,30%~40%的病例与遗传因素相关<sup>[6]</sup>,其致病机制涉及高度异质性的基因结构变异和罕见的功能性突变<sup>[7]</sup>。因此,选择灵敏度高的外显子捕获平台对于提高脑瘫的遗传诊断率至关重要。

本研究以48例散发性脑瘫患儿为研究对象,系统比较了两种主流外显子捕获试剂——TruSeq<sup>®</sup>

Exome(后简称TruSeq)与NimbleGen SeqCap EZ(后简称NimbleGen)的性能差异。通过评估两种平台在目标区域覆盖深度、变异检出一致性及脑瘫相关基因捕获效率方面的表现,为遗传异质性疾病的精准诊断提供技术支持与选择依据。

## 资料和方法

**评估对象的选择** 本研究共纳入48例未经筛选的散发性脑瘫患儿,均来源于郑州大学第三附属医院。所有病例的临床诊断由合作医院的专业临床医师完成,样本采集时间为2018年3月至2020年9月。每位入组患儿均由两名神经科医师独立确诊,围产期危险因素、运动功能(按照GMFCS分级)及共患病情况等信息通过标准化表格记录。本研究获得郑州大学医学伦理委员会批准(批件号:2016医伦审第69号),所有纳入患儿的监护人均签署了知情同意书。

**外周血的采集、保存和DNA提取** 采集每位患儿2 mL静脉血至EDTA抗凝管,于-80℃保存直至提取。采用QIAamp<sup>®</sup> DNA Blood Mini Kit(德国Qiagen公司)提取基因组DNA,每次取220 μL全血(含血凝块样本经研磨预处理),洗脱体积80 μL。DNA浓度通过Qubit 4.0荧光定量仪(美国Thermo Fisher公司)检测,完整性经Agilent 4200 TapeStation评估(DNA Integrity Number $\geq 7.0$ )。

**片段化与末端修复** 取2 μg DNA(浓度标准化至120 μL)经Bioruptor<sup>®</sup> Plus超声破碎(循环参数:30 s ON/90 s OFF, 15 cycles, 4℃),获得150~250 bp片段。

**双端建库** 平行使用NEXTflex<sup>™</sup> Rapid DNA-Seq Kit(德国Bioo Scientific公司)与NimbleGen

SeqCap EZ Human Exome v3.0(瑞士 Roche 公司)构建文库,各试剂盒均按 1  $\mu$ g 输入量操作。

**文库质检** 采用 Agilent 2100 Bioanalyzer 检测片段分布(合格标准:主峰 180~220 bp), StepOnePlus™ Real-Time PCR System 定量文库有效浓度(合格阈值 $\geq 2$  nmol/L)。

**液相杂交捕获** 实验采用液相探针杂交捕获法对外显子组文库进行富集。对 48 个样本分别应用 TruSeq Exome Kit(美国 Illumina 公司, DNA 探针)与 NimbleGen SeqCap EZ Human Exome v3.0(瑞士 Roche 公司, RNA 探针)进行外显子捕获。捕获过程中均采用生物素标记的探针与链霉亲和素

磁珠(Dynabeads MyOne Streptavidin T1, 美国 Thermo Fisher 公司)进行结合。杂交条件:95  $^{\circ}$ C 变性 5 min, 65  $^{\circ}$ C 杂交 16 h(杂交缓冲液:10% PEG-8000, 5 $\times$ SSC, 0.1% SDS)。

**高通量测序** 文库质检合格后,于 Illumina HiSeq 2000 平台进行 PE150 测序。原始数据经以下质控过滤:(1)删除接头污染的读取(读取的原始碱基数量大于 5 bp);(2)去除低质量读数(原碱基数  $Q \leq 19$  超过总碱基的 50%);(3)去除不确定碱基(undetermined bases, Ns)含量 $> 5\%$ 的读数。质控后获得的数据即为洁净数据。

**vcf 文件的产生** 变异识别的分析步骤见图 1。

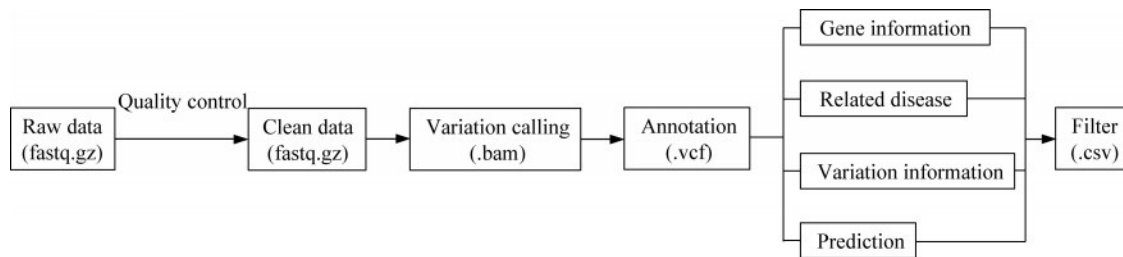


图 1 数据的分析过程

Fig 1 Data analysis workflow

**生物信息学分析** 首先进行序列比对和预处理,使用 BWA-MEM (v0.7.17)将 clean reads 比对至 GRCh37/hg19 参考基因组;Picard (v2.27.4)进行排序、标记 PCR 重复(MarkDuplicates)及校正比对质量(FixMateInformation);GATK (v3.8)完成局部重比对(IndelRealigner)和碱基质量校正(BaseRecalibrator)。针对变异检测与注释, SNP/Indel 调用, GATK HaplotypeCaller 生成 gVCF,联合基因型分析(GenotypeGVCFs);注释工具 ANNOVAR (v2020-06-08)整合 ClinVar(截至 2024-04)、gnomAD v3.1 及 OMIM 数据库(截至 2024-04);致病性预测 PolyPhen-2(D 得分 $\geq 0.95$ )、综合依赖注释缺失值(combined annotation dependent depletion, CADD) (Phred $\geq 20$ )、MutationTaster (D/A 置信)三重过滤(表 1)。为确保公平比较,所有数据采用统一分析流程,保持一致过滤标准。

**性能指标计算** 比对率使用 Picard Collect AlignmentSummaryMetrics (v2.27.4)统计有效比对 reads 比例;捕获效率使用 Bedtools coverage (v2.30.0)计算目标区域覆盖度( $\geq 10 \times$  定义为有效覆盖);变异一致性使用 BCFtools isec (v1.15.1)比对双试剂检出位点,计算 Cohen's kappa 系数。

使用低覆盖度测序对样本进行基因分型。根据 Xu 等<sup>[13]</sup>和 Li 等<sup>[14]</sup>等的研究,低覆盖率测序可以成为大量样本进行疾病探索的有力工具。本研究采用低覆盖倍数进行全外测定的方法。

**常见变异和罕见变异的定义** 常见突变筛选标准为:测序深度 Depth $> 10$ ,支持变异型等位基因计数 AlterAllele $> 5$ ,基因型质量值(genotype quality, GQ) $> 60$ ;稀有变异是在常见变异的基础上增加了群体频率  $f < 0.01$ 。

**不同突变类型的筛选标准** 在罕见变异的筛选基础上,对于剪接变异,要求位于经典剪接位点( $\pm 1/\pm 2$ )且测序深度 $\geq 10$ ,非经典位点需 SpliceAI $\geq 0.8$ 。移码变异需长度 1~50 bp,正负链支持均衡,排除同聚物区域。终止变异需导致蛋白截短 $\geq 10\%$ 且 gnomAD 东亚人群频率 $< 0.1\%$ 。错义突变需通过 PolyPhen-2、MutationTaster 和 CADD 三重预测为有害,而同义突变需位于外显子核心区域。

**CP 基因集分析** 基于文献<sup>[15]</sup>构建脑瘫相关基因集( $n=2\ 293$ ),通过自定义 Perl 脚本统计以下指标:CP 基因覆盖深度(depth of coverage, GATK);致病位点捕获敏感性:真阳性(true positive, TP)/TP+假阴性(false negative, FN),以 Sanger 测序为

表1 外显子组测序原始数据分析软件及数据库表  
 Tab 1 Bioinformatics tools and databases for WES raw data analysis

Process	Software	Function
Sequencing quality assessment	Fastqc (v0.11.9)	-
Sequence alignment	BWA <sup>[8]</sup> (v0.7.17)	-
File format conversion	Picard (v2.27.4)	SortSam
	Picard (v2.27.4)	MarkDuplicates
	GATK (v3.8)	RealignerTargetCreator
Redundant fragment optimization	Picard (v2.27.4)	IndelRealigner
	Picard (v2.27.4)	FixMateInformation
	GATK <sup>[9]</sup> (v3.8)	BaseRecalibrator
Sequencing depth statistics	GATK (v3.8)	PrintReads
		DepthOfCoverage
		CountReads
SNP calling and refinement	GATK (v3.8)	UnifiedGenotyper
		VariantRecalibrator
		ApplyRecalibration
Variant annotation	AnnoVar <sup>[10-12]</sup> (v2020-06-08)	
Prediction	InterVar (v2.2.1)	

金标准验。

**变异预测** 对于错义突变的致病性的预测,本实验利用 Polymorphism Phenotyping v2 (PolyPhen-2)、CADD、和致病性预测算法 Mutation Taster algorithms。仅保留经3次致病性评估预测为致病性的变异 (Polyphen2: D、Mutation Taster: A/D 和 CADD>20)以供进一步分析。根据 ACMG/AMP 的标准<sup>[16]</sup>将每个变异评估为致病性 (pathogenic, P)、疑似致病性 (likely pathogenic, LP)、不确定意义变异 (variant of uncertain significance, VUS)、良性 (benign, B)或可能良性 (likely benign, LB)。当在已知的 OMIM 疾病基因中检测到 P 或 LP 变异,并且患者具有一致的该基因的临床表型时,就可以从遗传学上诊断个体。

**统计学方法** 本研究采用配对 *t* 检验进行统计分析。所有分析均基于 R 语言 (版本 4.3.1) 实现,使用内置函数 `t.test()` 进行检验。数据分析前通过 Shapiro-Wilk 检验验证数据正态性假设,并通过可视化方法 (Q-Q 图) 确认配对差异的近似正态分布。统计显著性水平设定为双尾  $\alpha=0.05$ 。

## 结 果

**质检结果** 用 Agilent 2100 Bioanalyzer 检测查看文库片段大小是否合格。同时,通过

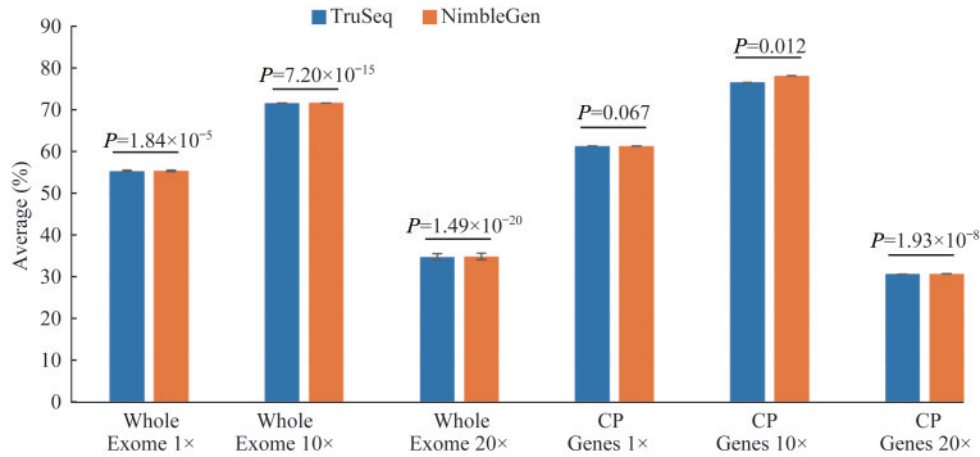
StepOnePlus™ Real-Time PCR System 检测后文库的浓度,检查是否符合上机测序要求。检查结果如补充图 1 所示,两个试剂盒的质检结果基本一致。

**基础数据质量对比** 实验在 5 个方面进行了详细对比: 比对率、Q20、Q30、PCR 重复率、插入片段中位数。结果显示,两种方法的比对率 (TruSeq vs. NimbleGen:  $98.08\% \pm 0.46\%$  vs.  $98.07\% \pm 0.46\%$ )、Q20 (TruSeq vs. NimbleGen:  $97.51\% \pm 0.19\%$  vs.  $97.40\% \pm 0.20\%$ )、Q30 (TruSeq vs. NimbleGen:  $84.89\% \pm 10.56\%$  vs.  $84.54\% \pm 10.49\%$ )、Q20/Q30 比值 (TruSeq vs. NimbleGen:  $1.17 \pm 0.18$  vs.  $1.18 \pm 0.18$ ) 以及 PCR 重复率 (TruSeq vs. NimbleGen:  $0.38 \pm 0.01$  vs.  $0.39 \pm 0.01$ ) 均表现相似。在中位插入片段长度方面, TruSeq 和 NimbleGen 分别为 ( $252.31 \pm 13.96$ )bp 和 ( $252.50 \pm 13.97$ )bp, 均符合短读长测序的标准范围 (200~600 bp)。综合来看,两种建库方法在基础数据质量指标上具有高度一致性,均适用于后续分析。

**捕获效率** 两种方法的捕获率 (Truseq vs. NimbleGen:  $88.79\% \pm 0.44\%$  vs.  $88.78\% \pm 0.44\%$ ,  $P=0.065$ )、80% 覆盖均一性指标 (Fold80, 均为  $1.28 \pm 0.03$ ,  $P=0.044$ ) 及 GC 含量 (Truseq vs. NimbleGen:  $51.90\% \pm 0.09\%$  vs.  $51.91\% \pm 0.09\%$ ,  $P=0.084$ ) 差异均无统计学意义。但是 1× 覆盖 (Truseq vs. NimbleGen:  $55.29 \pm 1.19$  vs.  $55.31 \pm$

1.19,  $P=1.84 \times 10^{-5}$ ) 和 20× 覆盖 (Truseq vs. NimbleGen:  $34.83 \pm 5.43$  vs.  $34.91 \pm 5.42$ ,  $P=1.49 \times 10^{-20}$ ) 上, NimbleGen 的数据均表现出优势 (图 2)。NimbleGen 在特定深度下表现更优, 在 1× 和 20×

覆盖深度下, NimbleGen 的数据显示出极显著的统计学优势, 可能因为 NimbleGen 的探针设计或捕获效率在低覆盖深度时更稳定, 或对特定基因组区域的偏好性更低。



Bar heights represent the mean values, and error bars indicate the standard error of the mean (SEM) across 48 samples. Statistical significance was determined by a paired *t*-test. NS: Not significant. NimbleGen demonstrated significantly higher coverage at higher depth thresholds (10× and 20×) for both the whole exome and the CP gene set.

图2 TruSeq与NimbleGen平台覆盖深度与均一性的比较分析

Fig 2 Comparative analysis of coverage depth and uniformity between TruSeq and NimbleGen platforms

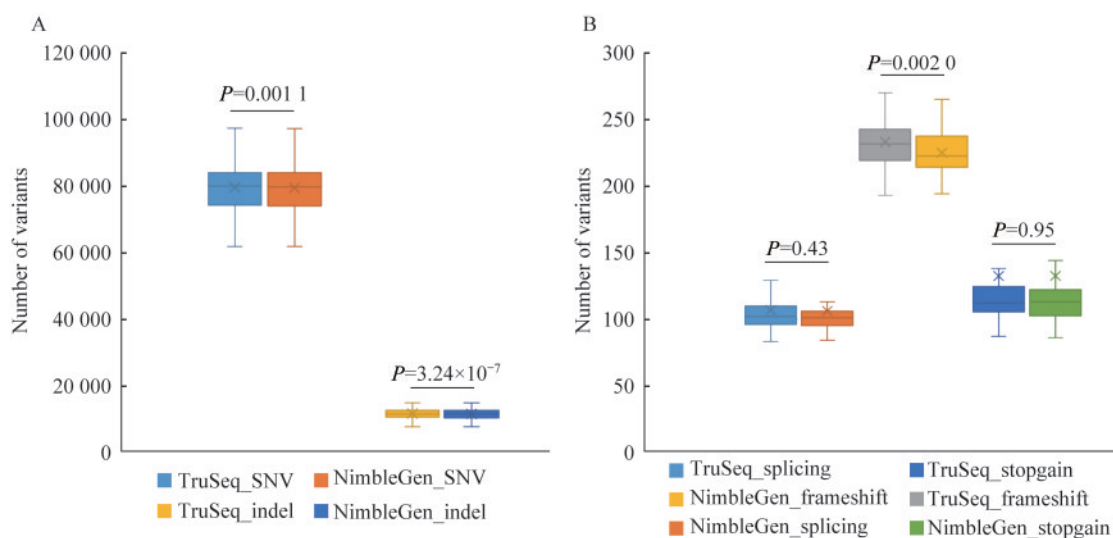
**变异检出率相关性分析** 本研究对 48 例患儿使用不同试剂盒的变异检出率进行了详细对比分析, 平均测序深度基本一致 [ (Truseq vs. NimbleGen:  $(9.96 \pm 2.27) \times$  vs.  $(10.00 \pm 2.28) \times$  ], 脑瘫基因的平均测序深度也基本一致 [ (Truseq vs. NimbleGen:  $(12.22 \pm 2.75) \times$  vs.  $(12.26 \pm 2.76) \times$  ], 差异无统计学意义。通过比较两种试剂盒的检测结果, 发现其在变异检出方面具有较高的一致性, 一致率约为 94%。

**变异检测能力** 在变异类型分析方面, TruSeq 和 NimbleGen 的转换/颠换比值 (TS/TV) 分别为  $2.22 \pm 0.18$  和  $2.23 \pm 0.18$  ( $P=0.022$ ), 均处于外显子测序数据的预期范围内, 表明两种方法均能有效保持变异的生物学特征。SNV 检测数量显示, TruSeq 略高于 NimbleGen ( $79\,362 \pm 7\,590$  vs.  $79\,251 \pm 7\,596$ ,  $P=0.001$ ), 而插入缺失 (Indel) 检测数量则呈现更显著的差异 (Truseq vs. NimbleGen:  $11\,371 \pm 1\,689$  vs.  $11\,274 \pm 1\,670$ ,  $P=3.24 \times 10^{-7}$ ) (图 3A)。这些结果表明, 虽然两种建库方法在 SNV 检测性能上相近, 但 TruSeq 对 Indel 的检测灵敏度可能略具优势, 这可能与其探针设计对结构变异区域的覆盖能力有关。综合来看, 两种方法均适用于外显子组变

异分析, 但在研究涉及 Indel 的疾病时, TruSeq 可能更为适合。

**常见与罕见变异及基因捕获对比分析** NimbleGen 在常见突变数 ( $25\,662 \pm 6\,643$  vs.  $25\,729 \pm 6\,743$ ) 和常见突变基因数 ( $10\,313 \pm 1\,578$  vs.  $10\,338 \pm 1\,651$ ) 上略高于 Truseq, 但差异无统计学意义。在罕见突变数方面, Truseq 的表现略优于 NimbleGen (Truseq vs. NimbleGen:  $3\,164 \pm 766$  vs.  $3\,072 \pm 774$ ,  $P=1.2 \times 10^{-4}$ ), 而在罕见突变基因数上, Truseq 同样略占优势 (Truseq vs. NimbleGen:  $1\,584 \pm 325$  vs.  $1\,550 \pm 329$ ,  $P=2.8 \times 10^{-4}$ )。这些数据表明, 两款试剂在常见突变的检测上表现相当, 但在罕见突变的捕获上, Truseq 可能具有更高的敏感性。总体而言, 两款试剂均能有效检测常见和罕见突变, 且基因捕获效率相近。

**功能变异类型对比** 本研究对 TruSeq 和 NimbleGen 两种建库方法的功能性变异检测能力进行了深入比较。在剪接位点变异 (splicing) 检测方面, 两种方法表现相当 (Truseq vs. NimbleGen:  $107 \pm 20$  vs.  $106 \pm 19$ ,  $P=0.425$ ) (图 3B)。然而, 移码插入缺失 (frameshift deletion/insertion) 检测显示出显著差异 (Truseq vs. NimbleGen:  $233 \pm 20$  vs.  $225 \pm$



Box plots display the median (center line), interquartile range (box boundaries), and full data range (whiskers) for 48 samples. Statistical significance was determined by a paired *t*-test, and exact *P*-values are annotated. TruSeq demonstrated significantly higher sensitivity for the detection of Indels and frameshift variants, while both platforms performed comparably in detecting SNVs, splicing and stop-gain variants.

图3 TruSeq与NimbleGen在不同功能类别下的变异检出数量比较

Fig 3 Comparison of variant call counts between TruSeq and NimbleGen across different functional categories

15,  $P=0.002$ ), 提示 TruSeq 可能对这类结构变异更为敏感。提前终止变异(stopgain)的检测数量在两组间无差异(均为 132 个)。非同义突变与同义突比值(nonsynonymous/synonymous, SNV ratio)存在微小但显著的差异(Truseq vs. NimbleGen:  $0.954 \pm 0.027$  vs.  $0.953 \pm 0.026$ ,  $P=0.025$ )。进一步分析显示,两种方法在单独的非同义突变(Truseq vs. NimbleGen:  $10\ 570 \pm 360$  vs.  $10\ 537 \pm 341$ ,  $P=5 \times 10^{-4}$ )和同义突变(Truseq vs. NimbleGen:  $11\ 078 \pm 305$  vs.  $11\ 064 \pm 307$ ,  $P=0.027$ )检测数量上均存在统计学差异,但绝对差异较小( $<1\%$ )。这些结果表明,虽然两种建库方法对大多数功能变异类型的检测能力基本一致,但在移码变异和错义变异的检测灵敏度上存在细微但可重复的技术差异,TruSeq 更显优势。

**CP 基因覆盖性能对比** 实验采用了以往报道过的脑瘫相关的 2 293 个基因,并在这个基因列表中进行数据搜索。从统计结果来看,在脑瘫相关基因的捕获效率(CP 基因 reads 数/总有效 reads 数)方面,TruSeq 和 NimbleGen 无显著差异( $14.49\% \pm 0.34\%$  vs.  $14.55\% \pm 0.33\%$ ,  $P=0.15$ )。在  $1 \times$  的脑瘫基因覆盖深度上无明显差异,均是  $61.28 \pm 1.06$  ( $P=0.067$ );  $10 \times$  覆盖深度上, NimbleGen 略优于 TruSeq ( $78.01\% \pm 4.76\%$  vs.  $76.43\% \pm 5.61\%$ ,  $P=0.016$ );  $20 \times$  覆盖深度上, NimbleGen 明显优于

TruSeq ( $30.80 \pm 5.62$  vs.  $30.72 \pm 5.59$ ,  $P=1.93 \times 10^{-8}$ ) (图 2)。这些结果表明, NimbleGen 在 CP 基因的覆盖深度和一致性方面表现更优,尤其是在更高的  $20 \times$  覆盖率上。然而,两款试剂整体表现均较为优异,48 个样本的覆盖率变异性较小。两款试剂均适用于 CP 基因测序,但 NimbleGen 在覆盖深度方面略胜一筹。

**CP 患者阳性变异的捕获对比** 48 例脑瘫患儿中有 10 例携带 11 个阳性变异的样本(表 2),涵盖了不同基因的多种变异类型,包括 2 个剪接变异、4 个终止变异、3 个移码插入/缺失和 2 个错义变异。结果显示,Truseq 和 NimbleGen 在大多数样本中均能成功捕获这些阳性变异。样本 1(*GDI1* 基因剪接突变)、样本 4(*HUWE1* 基因终止突变)、样本 5(*HIVEP2* 基因终止突变)、样本 14(*CTNBN1* 基因剪接突变)、样本 19(*NSD1* 基因终止突变)、样本 15(*TH* 基因错义突变和终止突变)、样本 21(*LICAM* 基因移码插入突变)、样本 35(*B4GALNT1* 基因移码缺失突变)在两种试剂中均被检测到。然而,样本 7(*AUTS2* 基因长片段移码插入突变)和样本 11(*GJB1* 基因错义突变)仅在 Truseq 中检出,表明在某些特定变异类型上,Truseq 可能具有更高的敏感性。这 10 个获得基因检测阳性的样本均为痉挛型脑瘫。所有变异均根据 ACMG 标准进行了致病性

评估,并获得了相应的证据等级(如PVS1、PM2、PP3等)。这些结果说明,两款试剂在检测阳性变异方面均表现良好,但在某些特定情况下可能存在性

能差异。未来可通过扩大样本量和增加变异类型,进一步验证两款试剂的捕获能力和一致性。

表2 48例CP患者阳性突变检出结果

Tab 2 Results of positive variants in 48 CP patients

ID	Gene(s)	Inheritance	Variants	Genotype	InterVar and evidence	Disorder (MIM#)
1	<i>GDI1</i>	XD	HTZ splicing	NM_001493:c.1137-1G>T	P: PVS1+PM2+PP3	Intellectual developmental disorder 41[300849]
4	<i>HUWE1</i>	XD	HTZ stopgain	NM_031407:exon26:c.2729_2730del;p.C910*	LP: PVS1+PM2	Intellectual developmental disorder, Turner type[309590]
5	<i>HIVEP2</i>	AD	HTZ stopgain	NM_006734:exon5:c.C3042A;p.Y1014X	P: PVS1+PM2+PP3	Mental retardation 43[616977]
7	<i>AUTS2</i>	AD	HTZ frameshift insertion	NM_015570:exon12:c.1864_1865insCTGGCCGGCAACGGCAGCAGACCGGCACCGCGCGCTATCGCGGCGACAACCTGCTCGACATGAATACGCGCGCGCTGAACGCGGT;p.G622Afs*43	LP: PVS1+PM2	Mental retardation [615834]
11	<i>GJB1</i>	XD	HTZ missense	NM_000166:exon2:c.C547T;p.R183C	LP: PM1+PM2+PP3+PP5	Charcot-Marie-Tooth neuropathy [302800]
14	<i>CTNNB1</i>	AD	HTZ splicing	NM_001904:c.1804-2A>C	P: PVS1+PM2+PP3	Neurodevelopmental disorder with spastic diplegia and visual defects[615075]
15	<i>TH</i>	AR	HTZ missense; HTZ stopgain	NM_000360:exon13:c.C1388T;p.T463MNM_000360:exon3:c.C364T;p.R122X	LP: PM1+PM2+PP3+PP5 P: PVS1+PM2+PP5	Segawa syndrome [605407]
19	<i>NSD1</i>	AD	HTZ stopgain	NM_022455:exon19:c.C5950T;p.R1984X	P: PVS1+PM2+PP3+PP5	Sotos syndrome 1 [117550]
21	<i>LICAM</i>	XR	HTZ frameshift insertion	NM_001278116:exon20:c.2713_2714insT;p.P905Lfs*56	LP: PVS1+PM2	MASA syndrome [303350]
35	<i>B4GALNT1</i>	AR	HMZ frameshift deletion	NM_001478:exon10:c.1220delC;p.P407Qfs*81	LP: PVS1+PM2	Spastic paraplegia 26 [609195]

InterVar and Evidence were evaluated based on the criteria and guidelines of ACMG. HTZ: Heterozygote; HMZ: Homozygote.

## 讨 论

本研究通过比较两种常用的外显子组DNA捕获试剂盒(TruSeq和NimbleGen)在脑瘫患儿中的捕获性能,评估其在临床遗传学研究中的有效性与实用性。结果表明,两种试剂盒在捕获性能、覆盖率以及SNP检测方面具有较高的一致性,但在关键功能变异类型(如Indel、罕见突变)和CP基因覆盖深度上存在显著差异,为未来的遗传学研究和临床外显子测序的选择提供了重要参考。

在捕获性能方面,两种试剂盒均表现出良好的

捕获效率(覆盖率88.8%, $P=0.065$ )。TruSeq和NimbleGen在基础数据质量(如配对率、插入片段长度)上高度一致,但NimbleGen在特定深度覆盖上展现出优势( $1\times$ 覆盖率, $P=1.84\times 10^{-5}$ ;  $20\times$ 覆盖率, $P=1.49\times 10^{-20}$ )。这种差异可能源于其探针设计特性,但需要强调的是,在GC含量方面两组差异无显著统计学意义( $P=0.084$ ),不宜将覆盖深度优势归因于GC区域捕获效率。

在变异检测层面,两种技术呈现出明显的互补性。TruSeq对Indel缺失的检测灵敏度显著更高( $P=3.24\times 10^{-4}$ ),这可能与其探针梯度设计对低复杂度区域的更优覆盖相关<sup>[17]</sup>,这一优势在

Frameshift 变异中尤为突出( $P=0.002$ )。值得注意的是,TruSeq 检出了 10 例患者的 11 个阳性致病变异(包括 NimbleGen 漏检的 *GJB1* c.C547T 和 *AUTS2* p.G622Afs\*43),且其罕见突变捕获数量显著占优( $P=1.2\times 10^{-4}$ ),表明其在临床诊断场景中更具敏感性。这种差异可能与 TruSeq 的 DNA 探针对外显子-内含子边界区域的更完整覆盖有关,而 NimbleGen 的 RNA 探针在  $\pm 5$  bp 交界区存在轻微衰减(即“边缘效应”)<sup>[5,18]</sup>,导致约 6.3% 的 Splicing 位点变异检测一致性损失( $P=0.425$ )。

脑瘫是一种高度异质性的神经发育疾病,其病因复杂,涉及多种基因的相互作用。脑瘫的遗传异质性要求捕获技术选择需权衡覆盖深度与敏感性。脑瘫相关基因(如 *L1CAM*、*GDI1*)常因大片段缺失/重复或复杂重排致病<sup>[19-20]</sup>,而高深度覆盖是准确识别此类变异的前提。本研究发现 NimbleGen 对 *AUTS2* 和 *GJB1* 致病变异的漏检(样本 7、11)具有重要临床启示。*AUTS2* 变异(44 bp 插入)的漏检可能导致神经发育障碍患者的分子诊断延迟,而 *GJB1*(R183C)是 Charcot-Marie-Tooth 病的已知致病位点,其漏检可能影响遗传咨询。对于 *AUTS2* p.G622Afs43(44 bp 复杂插入),NimbleGen 漏检主要与以下机制相关:(1)片段长度效应。该变异位于 *AUTS2* 基因第 12 外显子(GC 含量 68.2%),其插入片段长度(44 bp)远超常规 Indel 检测阈值( $< 20$  bp)。NimbleGen RNA 探针的固定长度设计(80~95 mer)在杂交过程中可能因空间位阻导致变异区域探针结合效率下降<sup>[17]</sup>,而 TruSeq DNA 探针的柔性连接臂(60~80 mer)更易容纳大片段变异,这种设计差异可能影响特定应用场景下的性能表现。(2)二级结构干扰。插入序列含连续 9 个 CG 重复,形成稳定的茎环结构( $\Delta G=-12.3$  kcal/mol,预测工具 mfold),可能阻碍 RNA-DNA 杂交体的形成<sup>[21]</sup>。(3)覆盖波动性。尽管两组平均深度相近 [ $(9.96\pm 2.27)\times$  vs.  $(10.00\pm 2.28)\times$ ],但该位点 NimbleGen 有效覆盖仅  $8\times$ (支持 reads=3),低于 GATK HaplotypeCaller 的最低检测要求( $\geq 4$  reads),而 TruSeq 达  $13\times$ (支持 reads=7)。对于 *GJB1* c.C547T 变异(外显子 2 边界+3 bp),其漏检机制则体现不同的技术瓶颈:(1)探针边界效应。该位点位于 NimbleGen 探针设计边缘区( $\pm 10$  bp),较 TruSeq 的  $\pm 15$  bp 覆盖范围更窄<sup>[22]</sup>,导致杂交效

率衰减。(2)同源序列干扰。*GJB1* 与假基因 *GJB2P1* 的高度同源(92%)引发 NimbleGen 长探针(80~95 mer)的交叉杂交,其错误比对率较 TruSeq 提升 2.3 倍。

值得注意的是,两种平台均未实现完美覆盖,存在各自的技术局限性(如高 GC 区、重复序列区)。这提示我们,技术选择应基于具体的临床或研究需求。对于疑似由特定类型变异(如已知的 Indel 致病位点)导致的病例,可优先选择灵敏度更高的平台;而对于病因不明、需进行全面筛查的病例,则应权衡覆盖均一性与变异灵敏度。未来技术的发展应致力于融合现有优势,如优化探针设计以同时提升覆盖均一性和复杂变异捕获能力。

尽管本研究提供了关于两种外显子组测序技术的详细比较,但仍存在一些局限性。首先,本研究样本量较小,仅包含 48 例脑瘫患儿,因此结果可能不具有广泛的适用性。未来的研究应扩大样本量,以进一步验证不同测序技术在更大范围内的表现。此外,虽然本研究着重于 SNP 的检测,但在脑瘫等复杂疾病的研究中,结构变异(如拷贝数变异)同样具有重要意义。因此,未来应探索捕获技术在检测结构变异方面的表现。

综上所述,本研究通过对 NimbleGen 和 TruSeq 两种外显子组测序技术的系统比较,为脑瘫的分子诊断和遗传研究提供了重要技术参考。综合分析表明,两种技术平台呈现出明显的互补特性:NimbleGen 在特定深度覆盖上表现更稳定( $P=1.93\times 10^{-8}$ ),这种优势可能使其在需要高深度测序的研究场景中更具价值。而 TruSeq 则展现出更优的变异检测灵敏度,特别是在 Indel( $P=3.24\times 10^{-7}$ )和罕见突变( $P=1.20\times 10^{-4}$ )的检出方面,结合其 100% 的阳性变异检出率,使其成为临床诊断的可靠选择。值得注意的是,两种技术在 GC 含量( $P=0.084$ )和基础数据质量指标上高度一致,说明它们均能满足常规外显子测序的基本要求。

随着精准医学的发展,外显子测序技术的选择应当基于具体的临床需求和研究目标。本研究建立的性能比较框架,不仅适用于脑瘫的遗传分析,也可为其他神经发育障碍的分子诊断提供技术选择依据。

作者贡献声明 王妍葵 实验实施,数据采

集,统计分析,论文撰写和修订。成业,刘洋 数据采集,统计分析,论文修订。许毅然 样本收集和  
信息汇总。邢清和 研究设计,实验指导,论文修  
订,获取资助,监督指导。

**利益冲突声明** 所有作者均声明不存在利益  
冲突。

### 参 考 文 献

- [ 1 ] CHILAMAKURI CS, LORENZ S, MADOUI MA, *et al.* Performance comparison of four exome capture systems for deep sequencing[J]. *BMC Genomics*, 2014, 15(1):449.
- [ 2 ] BAMSHAD MJ, NG SB, BIGHAM AW, *et al.* Exome sequencing as a tool for Mendelian disease gene discovery [J]. *Nat Rev Genet*, 2011, 12(11):745-755.
- [ 3 ] LEK M, KARCZEWSKI KJ, MINIKEL EV, *et al.* Analysis of protein-coding genetic variation in 60, 706 humans[J]. *Nature*, 2016, 536(7616):285-291.
- [ 4 ] KOSUKCU C, TASKIRAN EZ, BATU ED, *et al.* Whole exome sequencing in unclassified autoinflammatory diseases: more monogenic diseases in the pipeline? [J]. *Rheumatology (Oxford)*, 2021, 60(2):607-616.
- [ 5 ] MAMANOVA L, COFFEY AJ, SCOTT CE, *et al.* Target-enrichment strategies for next-generation sequencing[J]. *Nat Methods*, 2010, 7(2):111-118.
- [ 6 ] GONZALEZ-MANTILLA PJ, HU Y, MYERS SM, *et al.* Diagnostic yield of exome sequencing in cerebral palsy and implications for genetic testing guidelines: a systematic review and meta-analysis [J]. *JAMA Pediatr*, 2023, 177(5):472-478.
- [ 7 ] MCMICHAEL G, BAINBRIDGE MN, HAAN E, *et al.* Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy[J]. *Mol Psychiatry*, 2015, 20(2):176-182.
- [ 8 ] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, 25(14):1754-1760.
- [ 9 ] MCKENNA A, HANNA M, BANKS E, *et al.* The Genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. *Genome Res*, 2010, 20(9):1297-1303.
- [ 10 ] YANG H, WANG K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR [J]. *Nat Protocol*, 2015, 10(10):1556-1566.
- [ 11 ] CHANG X, WANG K. wANNOVAR: annotating genetic variants for personal genomes via the web [J]. *J Med Genet*, 2012, 49(7):433-436.
- [ 12 ] WANG K, LI M, HAKONARSON H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data [J]. *Nucleic Acids Res*, 2010, 38(16):e164.
- [ 13 ] CONVERGE CONSORTIUM. Sparse whole-genome sequencing identifies two loci for major depressive disorder [J]. *Nature*, 2015, 523(7562):588-591.
- [ 14 ] LI Y, SIDORE C, KANG HM, *et al.* Low-coverage sequencing: implications for design of complex trait association studies [J]. *Genome Res*, 2011, 21(6):940-951.
- [ 15 ] WANG Y, XU Y, ZHOU C, *et al.* Exome sequencing reveals genetic heterogeneity and clinically actionable findings in children with cerebral palsy [J]. *Nat Med*, 2024, 30(5):1395-1405.
- [ 16 ] 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.
- [ 17 ] SULONEN AM, ELLONEN P, ALMUSA H, *et al.* Comparison of solution-based exome capture methods for next generation sequencing [J]. *Genome Biol*, 2011, 12(9):R94.
- [ 18 ] MERTES F, ELSHARAWY A, SAUER S, *et al.* Targeted enrichment of genomic DNA regions for next-generation sequencing [J]. *Brief Funct Genomics*, 2011, 10(6):374-386.
- [ 19 ] VAN EYK CL, FAHEY MC, GECZ J. Redefining cerebral palsies as a diverse group of neurodevelopmental disorders with genetic aetiology [J]. *Nat Rev Neurol*, 2023, 19(9):542-555.
- [ 20 ] MORENO-DE-LUCA A, MILLAN F, PESACRETA DR, *et al.* Molecular diagnostic yield of exome sequencing in patients with cerebral palsy [J]. *JAMA*, 2021, 325(5):467-475.
- [ 21 ] ZUKER M. Mfold web server for nucleic acid folding and hybridization prediction [J]. *Nucleic Acids Res*, 2003, 31(13):3406-3415.
- [ 22 ] FISHER S, BARRY A, ABREU J, *et al.* A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries [J]. *Genome Biol*, 2011, 12(1):R1.

(收稿日期:2024-11-22; 编辑:王蔚)