

## 基因检测在临床与科研中的应用

刘铁城

解放军总医院 眼科, 北京 100853

**摘要:** 近年来, 基因检测在临床疾病诊断和科学研究中发挥着越来越重要的作用。特别是新一代测序 (next-generation sequencing, NGS) 技术的出现使许多疾病的致病基因被发现, 大量科研成果不断涌现出来, 极大地提高了临床诊断水平和科研水平。本文重点介绍了几种常见的基因测序技术, 比较其优缺点和在疾病诊断及科研中的应用, 旨在对研究生、临床医生在疾病诊断和科研工作中起到促进作用。

**关键词:** 基因; 遗传; 外显子组测序

**中图分类号:** R 394.3 **文献标志码:** A **文章编号:** 2095-5227(2014)06-0648-04 **DOI:** 10.3969/j.issn.2095-5227.2014.06.037

**网络出版时间:** 2014-03-25 17:13

**网络出版地址:** <http://www.cnki.net/kcms/detail/11.3275.R.20140325.1713.004.html>

### Genetic testing in clinical and basic science research

LIU Tie-cheng

Department of Ophthalmology, Chinese PLA General Hospital, Beijing 100853, China

The first author: LIU Tie-cheng. Email: ltc301@sina.com

**Abstract:** Recently genetic testing has played an increasingly substantial role in the diagnosis of clinical disease as well as basic science research. Owing to the application of next-generation sequencing (NGS) technology, more and more pathogenic genes are found, which tremendously improve both clinical practice and science research. This article therefore aims to highlight several genetic sequencing techniques that are commonly used, to compare their advantages and disadvantages in both disease diagnosis and science research, and to facilitate postgraduate students and clinicians in their clinical practice and scientific activities.

**Key words:** genes; inheritance; exome sequencing

随着人类基因组计划的完成, 人类对自身遗传信息的了解和掌握有了前所未有的进步。与此同时, 分子水平的基因检测技术平台不断发展和完善, 使得基因检测技术得到了迅猛发展, 基因检测效率不断提高。从最初第一代以 Sanger 测序为代表的直接检测技术和以连锁分析为代表的间接测序技术, 到 2005 年以 Illumina 公司的 Solexa 技术和 ABI 公司的 SOLiD 技术为标志的新一代测序 (next-generation sequencing, NGS) 的相继出现, 测序效率明显提升, 时间明显缩短, 费用明显降低, 基因检测手段有了革命性的变化。其技术正向着大规模、工业化的方向发展, 极大地提高了基因检测的检出率, 并扩展了疾病在基因水平的研究范围。2009 年 3 月, 约翰霍普金斯大学的研究人员在《Science》杂志上发表了通过 NGS 外显

子测序技术, 发现了一个新的遗传性胰腺癌的致病基因 PALB2 的报道, 标志着 NGS 测序技术成功应用于致病基因的鉴定研究<sup>[1]</sup>。同年,《Nature》发表了采用 NGS 技术发现罕见弗里曼谢尔登综合征 MYH3 致病基因突变,《Nat Genet》发表了遗传疾病米勒综合征致病基因<sup>[2-3]</sup>。此后, 通过 NGS 技术, 与遗传相关的致病基因不断被发现, NGS 技术已成为里程碑式的进步。2010 年,《Science》杂志将这一技术评选为当年“十大科学进展”。

近两年, 基因检测成为临床诊断和科学研究的热点, 得到了突飞猛进和日新月异的发展, 越来越多的临床和科研成果不断涌现出来。同时, 基因检测已经从单一的遗传疾病专业范畴扩展到复杂疾病和个体化应用领域, 其临床检测范围包括高危疾病的新生儿筛查、遗传疾病的诊断和基因携带的检测以及基因药物检测用于指导个体化用药剂量、选择和药物反应等诸多方面的研究<sup>[4-5]</sup>。目前, 基因检测在临床诊断和医学研究的应用正越来越受到医生的普遍重视, 引起研究人员的极大的兴趣。

本文介绍了几种 DNA 水平基因检测的常见方法, 比较其优缺点和在临床诊断和科学研究中的

收稿日期: 2014-02-25

基金项目: 海南省自然科学基金项目 (813221); 解放军总医院临床扶持基金 (2021FC-CXY-1005)

Supported by the Natural Science Foundation of Hainan Province(813221)  
作者简介: 刘铁城, 男, 硕士, 副主任医师, 硕士生导师。研究方向: 眼底病及眼遗传性疾病基因鉴定及功能研究。Email: ltc301@sina.com

应用,对指导研究生和临床医生课外学习,推进临床科研工作和提升科研教学水平有着指导意义。

## 1 第一代测序

**1.1 Sanger 测序** Sanger 测序采用的是直接测序法。1977年, Sanger 等发明了双脱氧链末端终止法,这一技术随后成为最为常用的基因测序技术<sup>[1]</sup>。2001年, Maxam 和 Gilbert 发明了 Sanger 测序法,并在此后的 10 年内成为基因检测的金标准<sup>[6]</sup>。其基本原理即双脱氧核苷三磷酸 (dideoxynucleoside triphosphate, ddNTP) 缺乏 PCR 延伸所需的 3'-OH, 因此每当 DNA 链加入分子 ddNTP, 延伸便终止。每次 DNA 测序是由 4 个独立的反应组成, 将模板、引物和 4 种含有不同的放射性同位素标记的核苷酸的 ddNTP 分别与 DNA 聚合酶混合形成长短不一的片段, 大量起始点相同、终止点不同的 DNA 片段存在于反应体系中, 具有单个碱基差异的 DNA 序列可以被聚丙烯酰胺变性凝胶电泳分离出来, 得到放射性同位素自显影条带。依据电泳条带读取 DNA 双链的碱基序列。

人类基因组的测序正是基于该技术完成的。Sanger 测序这种直接测序方法具有高度的准确性和简单、快捷等特点<sup>[7]</sup>。目前, 依然对一些临床上小样本遗传疾病基因的鉴定具有很高的实用价值。例如, 临床上采用 Sanger 直接测序 FGFR 2 基因证实单基因 Apert 综合征, 直接测序 TCOF1 基因可以检出多达 90% 的与 Treacher Collins 综合征相关的突变<sup>[8-9]</sup>。

值得注意的是, Sanger 测序是针对已知致病基因的突变位点设计引物, 进行 PCR 直接扩增测序。单个突变点的扩增包括该位点在内的外显子片段即可, 不必将该点所在基因的全部外显子都扩增。因此, 应明确定位要扩增的位点所在的基因外显子和该点的具体位置, 设计包括该点在内的上下游 150 ~ 200 bp 的外显子片段引物。此外, 尽管有 NGS 的出现, 但 Sanger 测序对于致病基因位点明确并且数量有限的单基因遗传疾病的致病基因的检测是非常经济和高效的。到目前为止, Sanger 测序仍然是作为基因检测的金标准, 也是 NGS 基因检测后进行家系内和正常对照组验证的主要手段。

值得注意的是, Sanger 测序目的是寻找与疾病有关的特定的基因突变。对于没有明确候选基因或候选基因数量较多的大样本病例筛查是难以完成的, 此类测序研究还要依靠具有高通量测序能力的 NGS。虽然 Sanger 测序具有高度的分析准确性, 但其准确性还取决于测序仪器以及测序条

件的设定。另外, Sanger 测序不能检测出大片段缺失或拷贝数变异等基因突变的类型, 因此对于一些与此相关的遗传性疾病还不能做出基因学诊断。

**1.2 连锁分析** 采用的是间接测序法。在 NGS 出现之前, 国际通用的疾病基因定位克隆策略是建立在大规模全基因扫描和连锁分析基础上的位置候选基因克隆。人类的染色体成对出现, 一条来自父亲, 一条来自母亲, 每一对染色体在同样的位置上拥有相同的基因, 但是其序列并不完全相同, 被称为父系和母系等位基因。遗传标记是指在人群中表现出多态现象的 DNA 序列, 可追踪染色体、染色体某一节段或某个基因座在家系中传递的任何一种遗传特性。它存在于每一个人, 但大小和序列有差别, 具有可遗传性和可识别性。目前采用第二代遗传标记, 即重复序列多态性, 特别是短串联重复序列, 又称微卫星标记。连锁分析是以连锁这种遗传现象为基础, 研究致病基因与遗传性标记之间关系的方法。如果控制某一表型性状的基因附近存在遗传标记, 那么利用某个遗传标记与某个拟定位的基因之间是否存在连锁关系和连锁的紧密程度就能将该基因定位到染色体某一位置上。1986年 Morton 等提出优势对数记分法 (log odds score method, LOD), 主要检测两基因以某一重组率连锁时的似然性。LOD 值为正, 支持连锁; LOD 值为负, 则否定连锁。通过计算家系中的微卫星标记与致病位点之间的 LOD 值, 可以初步估算二者间的遗传距离及连锁程度, 从而确定该基因在染色体上的粗略位置。然后利用该区域的染色体基因图谱, 分析定位区域内所有基因的功能与表达, 选择合适的候选基因进行突变检测, 最终将致病基因定位或克隆。

然而, 采用连锁分析进行基因检测存在很大的局限性。不但所需遗传样本量较大, 一般要求提供三代及以上遗传家系患者血样, 而且数据量大、处理复杂、产出速度较慢、定位不够精确 (一般只能定位在染色体某一区间), 这就使得研究工作繁重和定位基因的时间周期特别长。目前, 连锁分析采用的单核苷酸多肽性和短串联重复序列还在使用, 但经典的间接测序方法, 如单链构象多肽性、变性梯度凝胶电泳和异源双链分析在美国已被淘汰, 而在发展中国家作为研究手段还在有限使用<sup>[7]</sup>。

## 2 新一代测序

主要包括全基因组重测序 (whole-genome sequencing, WGS)、全外显子组测序 (whole-exome

sequencing, WES) 和目标区域测序 (targeted regions sequencing, TRS), 它们属于新一代测序技术。总体而言, NGS 技术具有通量大、时间短、精确度高和信息量丰富等优点, 使得遗传学研究者可以在短时间内对感兴趣的基因进行精确定位。但这些不同的测序技术在测序范围、数据分析量以及测序费用和时间等方面又有很大差异, 如果选择适合的方法, 对于临床诊断和科学研究将起到事半功倍的作用。

**2.1 目标区域测序** 目前常用的是基因芯片技术。其测序原理是基于 DNA 杂交原理, 利用目标基因组区域定制的探针与基因组 DNA 进行芯片杂交或溶液杂交, 将目标基因区域 DNA 富集, 再通过 NGS 技术进行测序。其测序过程是通过把数以万计的 cDNA 或寡聚核苷酸置于芯片上制成阵列, 将芯片上固定好的已知序列的核苷酸探针与溶液中含有荧光标记的相应核酸序列进行互补配对, 根据测序仪所显示强荧光的位置和强度, 获取每组点阵列信息, 再利用生物信息学算法确定目标核苷酸的序列组成。测序所选定的目标区域可以是连续的 DNA 序列, 也可以是分布在同一个染色体不同区域或不同染色体上的片段。目标区域测序技术, 对于以往通过连锁分析将基因突变锁定在染色体某一片段区域内, 但无法找出突变的情况是一个非常好的进一步检测手段。2010 年, Nicholas 等<sup>[10]</sup> 使用基因分型芯片联合连锁分析技术, 成功发现头小畸形的新基因 WDR62, 文章发表在《Nat Genet》杂志。类似的研究在家族性胰腺癌中确定 8 个候选变异位点和在家族性渗出性玻璃体视网膜病变发现易感基因 TSPAN12<sup>[11-12]</sup>。

基因芯片测序技术可以将经过连锁分析锁定了目标范围和经过全基因组筛选的特定基因或区域进行更深一层的研究, 是解决连锁分析无法发现致病基因的有效手段。基因芯片技术对于已知基因突变的筛查具有明显优势, 可以快速、全面地检测出目标基因突变。同时, 由于目标区域受到了限制, 测序范围大幅度减少, 测序时间和费用相应降低。但基因芯片检测所需要大量的 DNA, 由于已提取的 DNA 存在降解的风险, 用于基因芯片研究的血标本最好是冷冻的全血, 这样可以用于检测 DNA 的量有充分保证。

**2.2 全外显子组测序** 外显子组是单个个体的基因组 DNA 上所有蛋白质编码序列的总合。人类外显子组序列约占人类全部基因组序列的 1%, 约包含 85% 的致病突变。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。采用的技术平台主要是罗氏公司的 SeqCap EZ 全外显子捕获系统, Illumina 公司的 Solexa 技术和 Agilent 公司的 SureSelect 外显子靶向序列富集系统。其捕获的目标区为 34 ~ 62 M, 不仅包括编码区, 同时也加入了部分非编码区。NGS 的测序过程主要包括 DNA 测序文库的制备、锚定桥接、PCR 扩增、单碱基延伸测序和数据分析。研究者根据测序仪捕获到在测序过程中掺入有不同荧光标记碱基片段, 经计算机将荧光信号转化成不同颜色的测序峰图和碱基序列。基因测序结果与 NCBI 的 SNP 数据库、千人基因组数据库等国际权威数据库比对, 最终确定是否为突变基因。

自 NGS 技术问世以来, 利用 WES 在临床疾病致病基因的鉴定研究中取得前所未有的成果。这些成果不仅集中在单基因遗传疾病, 还在多基因影响的复杂疾病中发现大量相关基因。在单基因遗传性疾病中, 如视网膜色素变性、终端骨发育不良等发现新基因或已知基因新突变<sup>[13-14]</sup>。在一些罕见的疾病中, 如 Kabuki 综合征、家族性混合型低脂血症和脊髓小脑共济失调症等疾病中发现新的致病基因<sup>[2,15-16]</sup>。同时, 在小细胞肺癌、慢性淋巴细胞性白血病等肿瘤研究和诸如肥胖症、脑皮质发育不良等复杂疾病的研究中也取得丰硕成果<sup>[17-20]</sup>。

WES 技术在筛查范围和检出率等方面较其他测序技术具有明显的优势。例如, 对于采用 Sanger 测序和基因芯片测序不能筛查出基因的样本, 可以采用 WES 来进一步基因筛查鉴定。应用 WES 技术能够获得较传统方法对编码区测序更深的覆盖度和更准确的数据。由于信息量的大幅度增加, WES 可以获得更多个体的编码区信息, 因此成为检测致病基因和易感基因位点的有效手段。与连锁分析定位方法比较, WES 对家系的要求并不十分严格, 在单基因遗传病同一家系中有 2 ~ 3 个患者和 1 个正常人即可进行致病基因的鉴定研究, 而不需要连续三代的遗传家系。由于不需要严格的三代以上的遗传家系, WES 使以前不能进行研究的家系成为可能。不仅对于单基因遗传病是一个很好的研究手段, 对于许多常见病, 如肿瘤、糖尿病等疾病也可进行大规模比较研究。

**2.3 全基因组重测序** WGS 是对已知基因组序列的物种进行不同个体的全基因组的测序, 经过数据分析后对序列进行拼接、组装并获得基因组图谱, 或是对不同组织进行测序并分析体细胞突变

的一种研究方法。尽管 WES 可以快速全面地找出个体基因组上的所有突变, 从而找到个体间的差异, 但对于外显子以外的区域则不能有效地进行基因检测。对于此种情况, 目前还要借助 WGS 进行全基因组检测。但由于人类基因组过于庞大, 一次单端全基因组测序很难达到所需要的测序深度。因此, 需要重复测序或双端测序, 由此带来测序成本的大幅度提高, 由于不能达到足够的测序深度导致结果准确性降低。而对于临床疾病诊断和普通科研工作, 其高昂的检测费用也是难以承受的。尽管如此, 对于部分临床研究和 WES 不能解决的科研课题还需要借助 WGS 进行更加全面的基因检测。

### 3 展望

NGS 的出现为新兴的基因组技术增添了无限的活力和想象空间。特别是基因芯片的问世和其在临床上应用于大样本的疾病筛查和基因诊断中所展现出的活力, 以及其商业化发展的模式都令人鼓舞。眼科是单基因病最常见的学科, 利用芯片技术进行 Leber 病的筛查已使很多病因不明的视神经萎缩得到明确诊断。而原发性开角型青光眼是眼科最具隐蔽性和危险性的致盲性眼病, 其致病基因或突变的鉴定研究对疾病筛查将有着非常重要的临床价值, 并有巨大的商业价值。在新生儿糖尿病的筛查中采用基因芯片技术可以更快、全面、经济, 避免第一代测序的过于繁琐和漏检。基因芯片技术在产前诊断中更加具有发展前景。只要对孕妇进行 DNA 血液检查即可进行遗传疾病的筛查, 避免以往通过羊膜穿刺抽取羊水进行有创检查的局限性和危险性。目前, 随着基因检测技术水平的提升和检测费用的不断降低, 使发展大规模个体化基因检测在不久的将来成为可能。同时, 药物易感性基因和疾病发生易感基因的检测的深入开展, 个体化医疗将在基因检测的基础上得以实现。有理由相信, 随着人们生活水平的不断提高和健康意识的不断增强, 基因检测在未来医学发展中的应用前景将十分光明。

### 参考文献

- Jones S, Hruban RH, Kamiyama M, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene [J]. *Science*, 2009, 324 (5924): 217.
- Ng SB, Turner EH, Robertson PD, et al. Targeted capture and

massively parallel sequencing of 12 human exomes [J]. *Nature*, 2009, 461 (7261): 272-276.

- Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder [J]. *Nat Genet*, 2010, 42 (1): 30-35.
- Sequeiros J, Paneque M, Guimarães B, et al. The wide variation of definitions of genetic testing in international recommendations, guidelines and reports [J]. *J Community Genet*, 2012, 3 (2): 113-124.
- Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits [J]. *Nat Genet*, 2012, 44 (6): 623-630.
- Bakker E. Is the DNA sequence the Gold standard in genetic testing? Quality of molecular genetic tests assessed [J]. *Clin Chem*, 2006, 52 (4): 557-558.
- Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics [J]. *Nat Rev Genet*, 2013, 14 (6): 415-426.
- Robin NH, Falk M J and Haldeman-Englert CR. FGFR-related craniosynostosis syndromes [J/OL]. <http://www.ncbi.nlm.nih.gov/books/NBK1455> (updated 7 Jun 2011).
- Katsanis SH and Jabs EW. Treacher Collins syndrome [J]. <http://www.ncbi.nlm.nih.gov/books/NBK1532> (updated 30 Aug 2012).
- Nicholas AK, Khurshid M, D é sir J, et al. WDR62 is associated with the spindle Pole and is mutated in human microcephaly [J]. *Nat Genet*, 2010, 42 (11): 1010-1014.
- Nikopoulos K, Gilissen C, Hoischen A, et al. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy [J]. *Am J Hum Genet*, 2010, 86 (2): 240-247.
- Rosa-Rosa JM, Gracia-Azn á rez FJ, Hodges E, et al. Deep sequencing of target linkage assay-identified regions in familial breast cancer: methods, analysis pipeline and troubleshooting [J]. *PLoS One*, 2010, 5 (4): e9976.
- Liu T, Jin X, Zhang X, et al. A novel missense SNRNP200 mutation associated with autosomal dominant retinitis pigmentosa in a Chinese family [J]. *PLoS One*, 2012, 7 (9): e45464.
- Sun Y, Almomani R, Aten E, et al. Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene [J]. *Am J Hum Genet*, 2010, 87 (1): 146-153.
- Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia [J]. *N Engl J Med*, 2010, 363 (23): 2220-2227.
- Wang JL, Yang X, Xia K, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing [J]. *Brain*, 2010, 133 (Pt 12): 3510-3518.
- Byun M, Abhyankar A, Lelarge V, et al. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma [J]. *J Exp Med*, 2010, 207 (11): 2307-2312.
- Harbour JW, Onken MD, Roberson ED, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas [J]. *Science*, 2010, 330 (69): 1410-1413.
- Bowden DW, An SS, Palmer ND, et al. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study [J]. *Hum Mol Genet*, 2010, 19 (20): 4112-4120.
- Bilg ü var K, Ozt ü rk AK, Louvi A, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations [J]. *Nature*, 2010, 467 (7312): 207-210.