

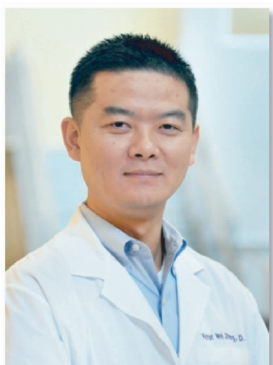
# 单基因病不简单:单基因遗传病的基因诊断技术进展及展望

张巍

(美国贝勒医学院人类和分子遗传系,美国休斯顿 77030)

【中图分类号】 R714.55

【文献标识码】 A



张巍博士(美国临床分子遗传学、临床生化遗传学、线粒体疾病和临床高通量测序专家)。美国医学遗传学院专家委员(FACMG),是美国少数拥有美国医学遗传学会(ABMG)双专科执照认证的诊断医生。1996年毕业于天津医科大学临床医学系,2003年在莱斯(Rice)大学获生化博士学位,2008年埃默里大学(Emory)和贝勒(Baylor)医学院从事临床基因诊断训练。原任美国贝勒医学院医学遗传学实验室副主任、贝勒医学院助理教授,现任贝勒医学院客座副教授和广州嘉检医学主任。

常年从事一线临床基因诊断工作,重点开发临床基因诊断实验室的新一代工作流程及大数据分析算法。他的文章范围包括从蛋白质功能分析、蛋白质结构测定、遗传变异特性、以及大规模测序的临床应用等。他目前的工作重点是通过研究遗传变异在人类基因组中的基因型与临床表型的相关性,为临床医生提供相应的遗传基因检测结果和咨询。以上研究成果共计发表 SCI 论文 50 余篇,并参与编写学术专著 7 部,被多次受邀在国际国内学术大会上做专题报告。

临床、科研和教育培训兴趣包括先天性代谢疾病、染色体异常和出生缺陷遗传病的基因学基础、出生缺陷、产前诊断和新生儿筛查、遗传咨询和临床管理、干预出生缺陷的公共卫生策略。普及临床医学遗传学知识,讲授临床医学遗传学的基本知识、新方法和新技术的发展及应用,内容覆盖细胞遗传、分子遗传、代谢性疾病、线粒体疾病、新生儿筛查、遗传咨询和生物信息学等。

## 1 单基因病的简介和背景

当一种疾病是由致病机制明确的某一个人类基因缺陷而出现临床症状和体征表现,我们一般称它为单基因疾病(single gene disorders)或符合孟德尔遗传方式的疾病(Mendelian inheritance disorders)。现在我们知道的已经有明确致病机制的人类基因疾病是大约 7000 多种,相应来说对于某一个单基因疾病的发病率是非常低的。如何在流行病学上定义一个疾病是不是罕见病,国际上不同国家也有不一样的定义。例如,根据美国 2002 年制订的罕

见病法案,比较清楚地给出按疾病发病率的定义是,在美国 3 亿人口中,如果一种疾病影响到 20 万人以下,也就是对某个病种如果在 1500 人中就有一个这样的病人,这个疾病就可以认为是一种罕见病。日本和欧洲也有类似的疾病发病率的定义,大约是 1/2000~1/2500 左右。如果把这几千种单基因疾病的发病率累加起来,单基因疾病的整体发病率为 2%~3%,占我国的每年新增出生缺陷 90 万例(出生缺陷率根据《中国出生缺陷防治报告 2012》的估计是 5.6%)的很大比重。

一般来讲,一个明确的单基因疾病在全国范围内是少见的,但在某个地区是有可能比较多发的。中国南方常见的蚕豆病是由在 X 染色体上的编码

葡萄糖-6-磷酸脱氢酶 *G6PD* 基因的缺陷造成的,地中海贫血可以由在 16 号染色体的编码 alpha 珠蛋白 *HBA* 基因或 11 号染色体的编码 beta 珠蛋白 *HBB* 基因的缺陷造成。美国常见的单基因疾病是囊性纤维化,这是由在 7 号染色体的编码囊性纤维化跨膜传导调节因子 *CFTR* 基因的缺陷造成的。单基因疾病常见的遗传方式是常染色体隐性遗传、常染色体显性遗传、X 染色体连锁遗传,但是还有非常多的一部分单基因疾病是以比较复杂的遗传方式出现,比如自发性的突变 (*de novo*)、印记缺陷 (*imprinting defect*)、重复扩张突变类 (*repeat expansion*) 等。

## 2 单基因病的临床异质性

单基因病的临床异质性是指当一个基因的缺陷可以导致不同种类临床表型,影响病人的基因突变的基因型和临床症状相关性不明确,会对医生和患者造成极大的诊断不确定性。在美国及欧洲等发达国家,由于生化和串联质谱的临床应用以及新生儿筛查工作的大规模开展,对于一些遗传代谢性疾病会有灵敏度高的遗传代谢产物的生成,会对这一类代谢性单基因疾病有直接临床指导意义<sup>[1]</sup>。但是大部分的单基因病缺乏这种异常代谢产物和基因型的关系,也为临床医生在诊断和治疗方面提出了极大的挑战。例如,以 DNA 聚合酶  $\gamma$  亚单位 (*POLG*) 基因突变相关性的疾病就可以归纳出多种临床症状,如进行性眼外肌麻痹、Alpers 综合征、帕金森症状、少年脊髓小脑性共济失调、癫痫综合征等<sup>[2,3]</sup>。电压控钠离子通道 (*SCN4A*) 基因突变相关性的疾病也可以多种临床形式表现出来,如低血钾性周期性麻痹、高血钾性周期性麻痹、钾离子相关性周期性瘫痪、先天性副肌强直、微管聚集性肌病、先天性肌无力综合征<sup>[4,5]</sup>。然而,对于这些临床表现已经很复杂的单基因病,新的临床症状和表型还是被不断地发现,而同一个突变在一个家庭的不同成员或不同家庭间也会存在一些临床症状的差异性。这也更进一步说明单基因病的复杂性和环境因素对症状的影响作用。所以,由于病人的症状的复杂性,能否准确确定疾病的基因缺陷也是一项具有挑战性的工

作。我们通常学习的医学教科书往往只提供一种疾病的典型临床症状,而在临床实践中每个单基因疾病都有可能是具有挑战性的,这是由于疾病的变化和进展,使得多数病人的临床症状是非典型的表现。随着高通量测序在临床的广泛应用,一方面可以发现人类新的致病基因,而更多方面是对已知的单基因疾病的非典型临床症状或新的临床症状的做进一步理解。对于以前已经有清楚认知的单基因疾病,由于出现许多新的表型,表型扩展的现象 (*phenotype expansion*) 也正变得越来越普遍,同时这些新的发现也会逐渐加强我们对个体疾病的准确认知<sup>[6]</sup>。

## 3 同一类疾病的基因异质性

像影响遗传代谢性病的糖原合成及分解的糖原贮积病 (*glycogen storage diseases*), 患者的初期临床表现通常是间歇期酮症性低血糖,一部分可以有肝功能异常,包括肝酶异常、肝肿大等,另外一部分患者有异常肌肉功能的表现,导致运动不耐受、肌肉无力和痉挛等。依靠明确的分子诊断是病人个性化管理和治疗的关键,但是之前需要依赖创伤性的肝脏或肌肉活检进行生化方法和免疫组化检测和诊断,在临床上会有一定的局限性和不确定性,现已知有 16 个以上的基因是与糖原贮积病的症状有关。如果用传统的一代测序 (*Sanger* 测序) 的方法进行分子诊断,将很难全面地对患者进行分析,有可能造成漏诊或误诊。最近的临床研究表明,如果选用合理的高通量测序方案,辅以严格的临床检测流程,可以达到 65% 的临床确诊率<sup>[7]</sup>。

遗传性神经肌肉疾病也是遗传异质性相当高的一系列疾病,虽然临床可以分成不同的亚型,包括先天性肌病、肢带型肌营养不良症、*CMT* 病、代谢肌病、肌强直综合征和离子通道疾病等。虽然以神经源性和肌源性系统受影响的程度和部位不同,初期临床表现多以肌无力、强直、萎缩或疼痛等症状出现,但遗传性神经肌肉疾病可以有几百种相关基因。最近的临床研究对 236 个神经肌肉疾病相关的基因进行临床级别的检测,可以达到 83% 的确诊率<sup>[8]</sup>。

具有表型明确的单基因性孟德尔疾病,由于诊

断技术的进步,之前认为是疑难病症的疾病,现在也可以相对容易地根据已知临床症状,缩小分子诊断的范围,为患者和医生提供非常宝贵的最初的诊断线索<sup>[9]</sup>。

#### 4 基因突变谱的异质性

一种单基因疾病的临床表型不但随着病情的发生阶段而不同,而且也与引起疾病基因的突变类型有关。通常来讲由于拷贝数变异引起基因组致病结构性变化的是与常染色体显性遗传或 X 连锁性遗传病有关,而与单基因相关的疾病常常是常染色体隐性遗传方式表现。致病性的基因突变类型大致可以包括,有一个或几个核苷酸变化的点突变(SNV)或多个核苷酸的变化引起较大区域的拷贝数突变(CNV)。以常见的侏儒畸形之一软骨发育不全为例,虽然大部分的突变是新生的,但是 *FGFR3* 基因的同样 c. 1138G>A (p. Gly380Arg) 点突变占了 98% 以上的病例。这是一类以点突变为主的单基因疾病。而以假肥大型肌营养不良症(DMD),也称 Duchenne 型肌营养不良症,约 65% 的基因突变是引起较大的区域(至少一个外显子)缺失和重复突变,其余的约 35% 是点突变<sup>[10]</sup>。由于在同一个基因因为突变类型造成临床表征不同的比较典型的疾病是成骨不全又称为脆骨症(osteogenesis imperfecta,OI),这个疾病是由 *COL1A1* 基因的编码的  $\alpha 1$  型胶原缺陷导致的全身性结缔组织和相应骨骼的病变。当患者有 *COL1A1* 基因无义突变(nonsense)时的骨骼表现比病人有错义突变(mis-sense)临床症状轻微<sup>[11]</sup>。对于一个单基因的疾病,临床医生不但需要熟知这个疾病的不同临床症状表现对病人进行治疗和处理,也需要了解引起这个疾病基因的结构和相应的突变谱。为了获得一个单基因病的基因的完整突变信息,可以采取分步检测和排除的方式,只有这样才能提高检测的阳性率,同时为病人减轻经济负担。在目前来讲,可能有一个相对全面的分子诊断项目可以同时进行点突变和拷贝数变异的检测<sup>[12]</sup>。

#### 5 单基因病的基因诊断技术进展及展望

已经在我国进行新生儿筛查的氨基酸代谢的病

苯丙酮尿症,通常认为是由于苯丙氨酸羟化酶(PAH)的缺陷引起的单基因病,但是有关四氢生物蝶呤(BH<sub>4</sub>)的缺陷也会引起苯丙酮尿症。这样总共是有 6 个基因可以引起这一个疾病。当一代测序刚普及的时候,对这 6 个基因的进行检测是非常昂贵的,一个 PAH 基因检测阴性而氨基酸代谢产物阳性的患者是没有能力进行全部余下基因检测的。随着全基因组芯片和高通量测序在临床分子诊断领域的广泛应用,我们可以对几十到 20 000 个基因同时进行检测,而且这些技术对人类新基因发现也是非常有价值的<sup>[13,14]</sup>。由于临床的需要和这些技术的独特性,这些新的诊断技术在美国临床基因诊断实验室得以广泛应用。一方面是满足个性化医学的临床诊断的要求,另一方面也为精准医学提供准确的基因组学的信息<sup>[15]</sup>。

但是这些新技术无疑也会有相应的不足之处,以杂交捕获方法为例,因为某些基因组区域的结构和化学组成不同,不少关键基因的测序覆盖程度不均匀或是没有覆盖,这种情况会造成检测结果的假阴性。在假基因和同源性很高的基因组区域,会有异常的高覆盖情况,也会容易引起结果的假阳性<sup>[16]</sup>。另外测序化学相关的错误和分析算法的偏差也需要在临床验证实验时进行考虑<sup>[17]</sup>。建立一个满足“临床标准”工作流程和相应的算法是一个非常具有挑战性的任务,最近的文献发现有开放源码的生物信息学分析软件包对于变异的低符合率也要引起我们足够的警觉<sup>[18]</sup>,所以了解每个环节的局限性无疑是有助于帮助病人和医生做出正确的选择。

在临床基因诊断工作中,我们目睹了分子诊断的快速发展,基因组技术的进步已经可以把每个人的全基因组测序成为可能。新的高通量技术已经从实验室的科研工具变成临床诊断的主要工具,可以提高诊断率,减少所需的时间和成本,以疾病表型为主的基因检测项目和全基因组测序的会更加普及应用,也为精准医疗发展奠定了基础<sup>[6]</sup>。

#### 参 考 文 献

- [1] Piero R, Silvia T, Dietrich M. Recent developments and new applications of tandem mass spectrometry in newborn screen-

- ing[J]. *Curr Opin Pediatr*,2004,16(4): 427-433.
- [ 2 ] Milone M, Benarroch EE, Wong LJ. POLG-related disorders: defects of the nuclear and mitochondrial genome interaction[J]. *Neurology*,2011, 77(20): 1847-1852.
- [ 3 ] Tang S, Wang J, Lee NC, et al. Mitochondrial DNA polymerase gamma mutations; an ever expanding molecular and clinical spectrum[J]. *J Med Genet*,2011,48(10): 669-681.
- [ 4 ] Hudson AJ, Ebers GC, Bulman DE. The skeletal muscle sodium and chloride channel diseases[J]. *Brain*,1995, 118 (Pt 2): 547-563.
- [ 5 ] Brunklaus A, Ellis R, Reavey E, et al. Genotype phenotype associations across the voltage-gated sodium channel family [J]. *J Med Genet*,2014,51(10): 650-658.
- [ 6 ] Yu H, Zhang VW. Precision Medicine for Continuing Phenotype Expansion of Human Genetic Diseases[J]. *Biomed Res Int*, 2015,2015: 745043.
- [ 7 ] Wang J, Cui H, Lee NC, et al. Clinical application of massively parallel sequencing in the molecular diagnosis of glycogen storage diseases of genetically heterogeneous origin[J]. *Genet Med*,2013,15(2): 106-114.
- [ 8 ] Tian X, Liang WC, Feng Y, et al. Expanding genotype/phenotype of neuromuscular diseases by comprehensive target capture/NGS[J]. *Neurology Genetics*,2015,1(2): e14.
- [ 9 ] Liewluck T, Tian X, Wong LJ, et al. Dystrophinopathy mimicking metabolic myopathies[J]. *Neuromuscul Disord*, 2015, 25(8): 653-657.
- [10] Tayeh MK, Chin EL, Miller VR, et al. Targeted comparative genomic hybridization array for the detection of single and multiexon gene deletions and duplications [J]. *Genet Med*,2009,11(4): 232-240.
- [11] Byers PH, Steiner RD. Osteogenesis imperfecta[J]. *Annu Rev Med*19923: 269-282.
- [12] Feng Y, Chen D, Wang GL, et al. Improved molecular diagnosis by the detection of exonic deletions with target gene capture and deep sequencing[J]. *Genet Med*,2015,17(2): 99-107.
- [13] Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders [J]. *N Engl J Med*,2013,369(16): 1502-1511.
- [14] Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing[J]. *JAMA*, 2014, 312(18): 1870-1879.
- [15] Zhang W, Cui H, Wong LJ. Application of next generation sequencing to molecular diagnosis of inherited diseases[J]. *Top Curr Chem*.2014,336: 19-45.
- [16] Yu H, van Karnebeek C, Sinclair G, et al. Detection of a novel intragenic rearrangement in the creatine transporter gene by next generation sequencing[J]. *Mol Genet Metab*,2013, 110(4): 465-471.
- [17] Gargis AS, Kalman L, Bick DP, et al. Good laboratory practice for clinical next-generation sequencing informatics pipelines[J]. *Nat Biotechnol*, 2015, 33(7): 689-693.
- [18] O'Rawe J, Jiang T, Sun G, et al. Low concordance of multiple variant-calling pipelines; practical implications for exome and genome sequencing[J]. *Genome Med*,2013,5(3): 28.

(收稿日期:2015-09-21)

编辑:刘邓浩

---

## 读者 · 作者 · 编者

### 本刊对文稿撰写的要求

文稿应具有科学性、实用性,论点明确,资料可靠,数据准确,层次清楚,文字精练,用字规范,文稿附图量不限,提倡多附图片和视频(音频)内容。论著性文章 4000 字左右,综述、讲座 5000 字左右,论著摘要、经验交流、病例报告等一般不超过 2000 字,欢迎以图像为主的来稿,并贯穿文字说明和评析,专家视频讲座为 30~40 分钟(分成 3~4 段)。当报告是以人为研究对象的试验时,作者应该说明其遵循的程序是否符合负责人体试验的委员会(单位性的、地区性的或国家性的)所制定的伦理学标准并得到该委员会的批准,是否取得受试对象的知情同意。文题力求简明,且能反映出文章的主题。中文文题一般不超过 20 个汉字。

中国产前诊断杂志(电子版)编辑部