

DOI: 10.3724/SP.J.1005.2011.00847

## 全基因组外显子测序及其应用

张鑫, 李敏, 张学军

安徽医科大学皮肤病研究所, 合肥 230032

**摘要:** 近年来, 众多研究小组开展了大量的全基因组关联研究(Genome-wide association studies, GWAS), 发现并鉴定了许多与复杂疾病/性状相关联的遗传变异, 为复杂疾病发病机制的研究提供了重要线索。由于 GWAS 的结果存在假阳性、假阴性、检测到的单核苷酸多态性很少位于功能区以及对稀有变异和结构变异不敏感等问题, 导致了其应用的局限性。而新一代测序技术的进步, 促进了全基因组测序和全基因组外显子测序的快速发展, 为解决上述问题提供了契机。全基因组外显子测序是利用序列捕获技术将全基因组外显子区域 DNA 捕获并富集后进行高通量测序的基因组分析方法。由于其具有对常见和罕见变异高灵敏度, 能发现外显子区绝大部分疾病相关变异以及仅需要对约 1%的基因组进行测序等优点, 促使全基因组外显子测序成为鉴定孟德尔疾病的致病基因最有效的策略, 也被运用于复杂疾病易感基因的研究和临床诊断中。

**关键词:** 外显子测序; 孟德尔疾病; 复杂疾病; 临床诊断

## Exome sequencing and its application

ZHANG Xin, LI Min, ZHANG Xue-Jun

*Institute of Dermatology of Anhui Medical University, Hefei 230032, China*

**Abstract:** In recent years, researchers have identified a large number of complex diseases/traits-associated genetic variants by performing genome-wide association studies (GWAS), which may provide important clues on understanding the mechanisms of related diseases. However, GWAS has its own limitations in terms of being false positive, false negative results, very few SNPs located in the functional areas and insensitive to detect rare and structural variations, which results in the application limitation of this method. With the development of the next-generation sequencing technology, whole genome and exome sequencing developed rapidly and provide an opportunity for us to deal with the problem caused by GWAS. This high-throughput sequencing technology is applied for sequencing the exome (1% of genome) to discover most of the diseases-related variations in exons. Furthermore, it is highly effective to detect common and rare variations. Due to these advantages, exome sequencing has become a powerful and efficient strategy for identifying the genes responsible for mendelian disorders and complex diseases, which will be very helpful for the diseases clinical diagnosis.

**Keywords:** exome sequencing; mendelian disorders; complex diseases; clinical diagnosis

随着人类基因组计划(Human genome project, HGP)和国际人类基因组单体型图计划(The international HapMap project)的完成以及高通量生物芯片技术的成功研发, 人们广泛利用高通量全基因组生

收稿日期: 2000-00-00; 修回日期: 2000-00-00

基金项目: 国家高技术研究发展计划项目(863 计划)(编号: 2007AA02Z161)资助

作者简介: 张鑫, 博士, 研究方向: 复杂疾病遗传学。E-mail: zhangxin68619@163.com

通讯作者: 张学军, 博士, 主任医师, 教授, 博士生导师, 研究方向: 复杂疾病遗传学。E-mail: ayzxj@vip.sina.com

物芯片的技术手段,采用关联分析的方法即全基因组关联研究(Genome-wide association studies, GWAS)来筛选复杂疾病易感基因,取得了前所未有的成就。但由于 GWAS 存在着一定的局限性,同时随着新一代测序技术的迅猛发展,促进了全基因组测序和全基因组外显子测序的发展。本文就全基因组外显子测序的研究方法及其应用作一综述。

## 1 外显子测序的产生背景

GWAS 是通过大规模的人群(病例/对照)DNA 样本进行包括单核苷酸多态性变异 (Single-nucleotide polymorphisms, SNPs)、拷贝数变异(Copy number variation, CNV)在内的全基因组高密度遗传标记并分型,从而寻找与复杂疾病相关的遗传因素的研究方法,掀起了人类基因组研究第 3 次浪潮。自 2005 年以来,利用 GWAS 对多种常见疾病进行了研究,发现和重复验证了近 2 000 个 SNPs 或位点,其中包括以前未检测到的而与疾病密切相关的基因及部分未知基因<sup>[1]</sup>。但此方法有其欠缺,如容易产生假阳性和假阴性,而且发现的与疾病关联的 SNPs 多位于基因间或内含子上,很少位于功能区(如外显子区和 5'UTR 区)。同时,芯片检测位点有一定的有限性,除 Illumina 的新一代全基因组基因分型芯片外,多数是发现常见变异(Minor allele frequency, MAF >5%),而对稀有变异(MAF <5%)和其他结构变异不敏感。目前,GWAS 是基于常见疾病/常见变异的假说,而越来越多的研究表明许多复杂疾病是由稀有变异造成的<sup>[2,3]</sup>,而这种基于芯片的 GWAS 在实验设计时尚未充分考虑这部分信息,故而较难搜寻稀有变异。过去主要通过定位克隆、物理作图和候选基因测序的方法研究孟德尔疾病,但这些方法受到诸多因素的限制,如患者人数少、家系少或小、有意识的减少生育等,许多孟德尔疾病尚未发现其致病基因。

现有的技术和研究方法在研究孟德尔疾病和复杂疾病时均遇见了难以解决的问题,迫切需要新技术的出现。而随着新一代测序技术的迅速发展、测序费用降低和时间缩短,使得全基因组和全基因组外显子测序在大规模人群中的运用成为现实。由于全基因组测序的费用在短时间内很难降到与外显子测序相当的价格,在经费一定的情况下,全基因组外显子测序更适合大批量样本研究以获得高深度的

测序数据。更重要的是,目前一致认为大部分功能变异都潜藏在外显子中<sup>[4]</sup>,这是因为引起孟德尔疾病的突变主要位于基因内造成的<sup>[4,5]</sup>。因此,全基因组外显子测序已成为现阶段基因测序工作的重心。

## 2 外显子测序的基本思路及技术路线

外显子测序主要包括目标区域序列的富集、DNA 测序、生物信息学统计 3 个主要步骤。

### 2.1 目标区域序列的富集

在过去的 20 年里,传统的 Sanger 测序主要是利用 Uniplex PCR 和 Mutlplex PCR 方法富集目标区域序列<sup>[6]</sup>,但其引物合成与 PCR 反应所需费用都很高,且实验周期长、人力资源耗费大。而新一代测序需要对大量外显子进行测序,进而研究疾病相关区域并 SNP 验证,传统的方法不能满足这一要求,在此背景下,NimbleGen 公司和安捷伦相继开发了新的技术以解决这一重大问题,为全基因组外显子测序的广泛运用提供了平台。

#### 2.1.1 NimbleGen 公司的外显子捕获芯片

罗氏 NimbleGen 推出的基于 HD2 平台 2.1M 外显子序列捕获芯片<sup>[6]</sup> (Sequence Capture 2.1M Human Exome microarrays)主要是利用杂交和 DNA 微阵列技术基因分离原理来捕获目标区域。其大致流程是将打断后的基因组 DNA 与定制的序列捕获芯片杂交,洗去未杂交上的片段,随后将富集的目标片段洗脱扩增,最后进行高通量测序。该方法具有高特异性和高覆盖度、捕获区域可按需设计和省时省力等优点。自 2008 年罗氏推出 NimbleGen 序列捕获 2.1M 外显子组芯片以来,全球的研究人员已通过此款序列捕获芯片对成千上万的外显子组进行测序。罗氏随之在 2009 年底推出的 EZ 外显子组序列捕获能提供液相的工作流程并以 DNA 液相探针杂交来捕获目的片段,该技术能提供高质量的目标序列富集以及更为灵活的可扩容性。

#### 2.1.2 安捷伦 SureSelect 靶向序列捕获系统

安捷伦的外显子捕获系统<sup>[7]</sup>是基于寡核苷酸合成技术的液相靶向序列捕获系统。其大致流程是将打断后的基因组 DNA 与 SureSelect 诱饵共孵,通过含有链酶亲和素标记的磁珠,钓出 RNA 诱饵-DNA

杂合体, 洗脱磁珠, 降解 RNA 诱饵, 富集目的区域, 再进行高通量测序。基于液相序列捕获的高通量平行靶向测序技术具有高度特异性、高准确性、卓越的重现性和广泛的应用前景。

## 2.2 测序技术

### 2.2.1 传统的测序技术

最早的 DNA 测序技术是 1977 年 Sanger 的“双脱氧链末端终止法”和 Maxam 及 Gilbert 的“化学降解法”。运用最多的是 Sanger 测序法, 但这种测序仪一般一次最多只能同时进行 96 个或 384 个样品的测序。Sanger 测序技术经过 30 年的不断发展与完善, 现在的测序长度可达 1 000 bp, 每一个碱基的读取准确率高达 99.999%, 费用大约为 0.5 美元/1 000 个碱基, 已成为测序的金标准<sup>[8]</sup>。

### 2.2.2 新一代测序技术(Next-generation sequencing)

传统的 Sanger 测序法无法大幅度的降低测序费用, 而基因变异、RNA 表达、蛋白与 DNA 相互作用以及染色体结构等研究需要使用高通量 DNA 测序技术。而其他学科技术如计算机技术、聚合酶工程技术、数据存储分析技术的发展, 均促使了新一代测序技术的进步。新一代测序技术原理是: 片断化的基因组 DNA 两侧连上接头, 用不同的方法产生几百万个空间固定的 PCR 单克隆阵列, 所有单克隆同时、独立地进行引物杂交和酶延伸反应, 同时拍摄每个延伸所掺入的荧光标记信号来获取测序数据。新一代测序技术广泛运用于 De novo 测序、重测序、细菌基因组和比较基因组研究、小 RNA 测序、古生物学和古 DNA 研究领域以及环境基因组学和感染性疾病等研究领域。

近期市面上出现很多新一代测序仪产品, 例如美国 Illumina 公司和英国 Solexa technology 公司合作开发的 Illumina 测序仪<sup>[9]</sup>、美国 Roche Applied Science 公司的 454 基因组测序仪<sup>[10]</sup>、美国 Applied Biosystems 公司的 SOLiD 测序仪<sup>[11]</sup>、Dover/Harvard 公司的 Polonator 测序仪以及美国 Helicos 公司的 HeliScope 单分子测序仪<sup>[12]</sup>。这些新型测序仪属于合成测序, 大多使用循环芯片测序法(cyclic-array sequencing), 通过对布满 DNA 样品的芯片重复进行基于 DNA 的聚合酶反应(模板变性、引物复性杂交及延伸)以及荧光序列读取反

应。但其中单分子测序仪不需要扩增建立 DNA 库, 而是基于边合成边测序的思想, 将待测序列随机打断成小片段并在 3'末端加上 PolyA, 通过合成互补链技术对数百万个 DNA 片段进行测序。

### 2.3 数据统计分析

不同的新一代测序仪都能产生大量的数据, 但如何在如此庞大的数据中分析出有意义的内容对于新一代测序来说无疑是一个巨大的挑战。数据统计分析主要包括基本的数据分析(如图像的去噪、锐化、定位和偏移校正、依据光强度获得碱基)和生物信息分析(如检验靶区域的测序深度和覆盖度、比对序列、检测和注释 SNPs 和短小的插入或缺失(Short insertion/deletions, indels)), 这对于测序数据的深入发掘具有重要的意义。

## 3 全基因组外显子测序的相关研究

### 3.1 在正常人中的相关研究

Ng 等<sup>[13]</sup>对一个体的全基因组外显子进行测序, 结果发现 10 389 个非同义 SNPs (nsSNPs), 其中 5 604 个是杂合子, 4 785 个是纯合子, 大部分变异属于常见变异, 且为中性。同时检查了与疾病相关的已知基因, 未发现该个体存在患某种重大疾病的风险。这是首次对单个个体的外显子进行研究, 不仅丰富了 NCBI 数据库的参考序列, 而且为大规模的外显子测序提供了较具体的研究方法。

Li 等<sup>[14]</sup>利用 NimbleGen 2.1M 芯片捕获外显子, Genome Analyzer II 测序系统对 200 个丹麦人测序, 在群体中共发现了 121 870 个 SNPs, 包括 53 081 个编码区的 SNP 位点(cSNP)。通过统计学方法对群体数据进行 SNP 的鉴定和等位基因频率的估算, 获得了 MAF 在 2%以上的 cSNP 的频率图谱; 发现 MAF 为 2%~5%的低频率区域, 稀有的有害非同义 cSNP 是同义 cSNP 的 1.8 倍。从而间接说明了全基因组外显子测序能发现低频突变。

### 3.2 在孟德尔疾病和罕见综合症中的运用

利用连锁分析方法研究孟德尔疾病需要足够多的患者、多代遗传的大家系, 因此不适合难以收集到大家系和只有散发病例的孟德尔疾病, 而全基因组外显子测序没有这些限制。为验证外显子测序在孟德尔

疾病研究中的可行性, Ng 等<sup>[15]</sup>利用安捷伦 244K 芯片捕获外显子, Genome Analyzer II 系统对 12 例研究对象进行测序, 其中 8 例为 HapMap 计划中的个体, 4 例为无亲缘关系的弗里曼谢尔登综合征 (Freeman-sheldon syndrome, FSS) 患者, FSS 是一种由 *MYH3* 基因突变引起的常染色体显性遗传病。同 HapMap 数据、全基因组测序数据(NA18507)和 Sanger 测序结果比较表明了外显子测序数据的可靠性和研究遗传变异的敏感性与特异性。通过逐步滤过法, 同 HapMap 数据库和 dbSNP 数据库比对滤掉普遍变异和个人独具的变异后, 验证了 FSS 的致病基因为 *MYH3*。以上研究表明, 全基因组外显子测序可用于研究少数散发病例和有小家系的孟德尔疾病, 为后续的研究提供了基于全基因组

外显子测序策略的研究思路。

目前全基因组外显子测序除已广泛运用于孟德尔疾病的研究中, 也能运用于一些代谢性疾病的研究。依据全基因组外显子测序研究对象情况以及是否有定位区域, 将目前的研究现状进行了总结(表 1)。我国在该方面的研究也取得一定的成绩: 2010 年中南大学湘雅医院研究发现小脑共济失调新的致病基因 *TGM6*<sup>[16]</sup>, 这是我国科学家应用外显子组测序技术进行孟德尔疾病研究的一项突破, 对促进国内孟德尔疾病研究的发展具有重要意义。随后通过该技术发现并验证了 *NCSTN* 基因的突变可导致逆向性痤疮的发生, 对 *NCSTN* 基因突变的检测和逆向性痤疮及其相关疾病的诊断、治疗具有十分重要的意义<sup>[17]</sup>。

表 1 全基因组外显子测序研究孟德尔疾病或综合征文献

疾病	遗传模式	致病基因	全基因组外显子测序	
			捕获芯片	测序平台
1. 散发病例				
弗里曼谢尔登综合征 <sup>[15]</sup>	AD	<i>MYH3</i>	SureSelect Human All Exon kit	Illumina GAI
Kabuki 综合征 <sup>[18]</sup>	AD	<i>MLL2</i>	SureSelect Human All Exon kit	Illumina GAI
Schinzel-Giedion 综合征 <sup>[19]</sup>	AR	<i>SETBP1</i>	SureSelect Human All Exon kit	SOLiD
Sensenbrenner 综合征 <sup>[20]</sup>	AR	<i>WDR35</i>	SureSelect Human All Exon kit	SOLiD
Fowler 综合征 <sup>[21]</sup>	AR	<i>FLVCR2</i>	SureSelect Human All Exon kit	Illumina GAIIX
Perrault 综合征 <sup>[22]</sup>	AR	<i>HSD17B4</i>	SureSelect Human All Exon kit	Illumina GAIIX
Hajdu-Cheney 综合征 <sup>[23]</sup>	AD	<i>NOTCH2</i>	SureSelect Human All Exon kit	Illumina GAIIX
成骨不全 <sup>[24]</sup>	AR	<i>SERPINF1</i>	SureSelect Human All Exon kit	SOLiD
复合物 I 缺乏症 <sup>[25]</sup>	代谢性疾病	<i>ACAD9</i>	未提及	SOLiD
2. 散发和家系内病例				
米勒综合征 <sup>[26]</sup>	AR	<i>DHODH</i>	SureSelect Human All Exon kit	Illumina GAI
Brown-Vialletto-van Laere 综合征 <sup>[27]</sup>	AR	<i>C20orf54</i>	NimbleGen 2.1M array	Illumina GAIIX
3. 家系内病例				
血磷酸脂酶过多智力迟钝综合征 <sup>[28]</sup>	AR	<i>PIGV</i>	SureSelect Human All Exon kit	SOLiD
家族性 β-脂蛋白过少血症 <sup>[29]</sup>	AD	<i>ANGPTL3</i>	SureSelect Human All Exon kit	Illumina GAIIX
色素性视网膜炎 <sup>[30]</sup>	AR	<i>DHDDS</i>	NimbleGen 2.1M array	Illumina GAIIX
4. 家系内病例结合定位区域				
非综合征性耳聋 <sup>[31]</sup>	AR	<i>GPSM2</i>	SureSelect Human All Exon kit	Illumina GAIIX
肾脏相关性 ciliopathy(NPHP-RC) <sup>[32]</sup>	AR	<i>SDCCAG8</i>	NimbleGen 385K array *	Illumina GAI
Carnevale, Malpuech, Michels 和 OSA 综合征 <sup>[33]</sup>	AR	<i>MASP1</i>	SureSelect Human All Exon kit	Illumina GAIIX
原发性淋巴管性水肿 <sup>[34]</sup>	AD	<i>GJC2</i>	SureSelect Human All Exon kit	Illumina GAIIX
肌萎缩性侧索硬化(ALS) <sup>[35]</sup>	AD	<i>VCP</i>	SureSelect Human All Exon kit	Illumina GAIIX
非综合征的智力迟钝 <sup>[36]</sup>	AR	<i>TECR</i>	Agilent 244K microarray	Illumina GAI
Van Den Ende-Gupta 综合征 <sup>[37]</sup>	AR	<i>SCARF2</i>	SureSelect Human All Exon kit	Illumina GAIIX
自身免疫性淋巴组织增生症(ALPS) <sup>[38]</sup>	AR	<i>FADD</i>	SureSelect Human All Exon kit	Illumina GAIIX
小脑共济失调 <sup>[16]**</sup>	AD	<i>TGM6</i>	NimbleGen 2.1M array	Illumina GAI
逆向性痤疮 <sup>[17]**</sup>	AD	<i>NCSTN</i>	SureSelect Human All Exon kit	Illumina HiSeq2000

注: AD 是常染色体隐性遗传模式, AR 是常染色体显性遗传模式; GA 是 Genome Analyzer 的缩写; \*表示候选基因定制芯片;

\*\*表示该研究是中国的研究成果。

### 3.3 在复杂疾病中的应用

全基因组外显子测序能准确地找到孟德尔遗传疾病的致病基因, 那么外显子测序同样也可用于研究常见疾病。目前已有利用全基因组外显子测序的方法结合或不结合连锁分析结果研究复杂疾病的案例, 例如: 胰岛素抵抗动脉粥样硬化家庭研究主要是研究遗传和环境因素对西班牙裔和非裔美国人的葡萄糖稳态和肥胖症的影响, Bowden 等<sup>[39]</sup>利用安捷伦 SureSelect 外显子捕获系统和 Genome Analyzer IIx 系统对浆乙二醇水平无显著性差别的 2 个家系中 3 个患者进行测序, 发现 *ADIPOQ* 基因的低频突变 (1.1%)G45R, 能解释 17% 的西班牙裔美国人的血浆乙二醇水平, 63% 的家族存在该突变; 脑皮质发育异常是一组大脑功能障碍的疾病, 影响到神经细胞增殖、迁移、组织形成等, Bilguvar 等<sup>[40]</sup>运用 NimbleGen 2.1M 芯片捕获外显子和 Genome Analyzer II 测序系统对 1 例患者测序, 发现 *WDR62* 基因与脑皮质发育异常疾病相关, 从而能解释一系列的严重皮质畸形的症状, 包括小头畸形、巨脑回畸形等, 该研究提示外显子测序对具有遗传异质性和诊断亚型困难的疾病具有一定的帮助。早期主要运用传统的 Sanger 测序方法研究肿瘤患者的外显子以期发现新的体细胞突变, 但耗时耗经费, 随着新一代测序技术的发展, 加快了对肿瘤的研究步伐。

#### 3.3.1 肿瘤基因组计划

国际肿瘤基因组在 *Nature* 上发表的专题报道“International network of cancer genome projects”中详细描述了国际肿瘤基因组协作研究计划启动的历程、研究目标、研究策略、技术规范、数据分析与交汇及伦理学等核心问题<sup>[41]</sup>。国际肿瘤基因组协作组 (ICGC) 对人类常见的 50 种肿瘤 (或亚型) 进行全基因组变异分析, 明确肿瘤发生发展过程中基因型与表型变异的关系, 为全面解析肿瘤的生物学特性与机制提供基础数据。我国肿瘤学和基因组学的科学家携手组建了“中国肿瘤基因组研究协作组 (CCGC)”, 启动了中国肿瘤基因组研究项目并选择我国常见高发肿瘤胃癌作为参与 ICGC 的研究任务, 拟采用外显子测序和典型病例的全基因组测序的方案。

在过去的 3 年中, 运用外显子测序的研究方法, 研究人员在卡波西肉瘤、小细胞肺癌、非小细胞肺

癌、Hodgkin 淋巴瘤、急性髓性白血病、小叶基底乳癌、慢性淋巴细胞性白血病、骨癌、葡萄膜黑色素瘤等肿瘤的研究中取得了显著的成就<sup>[42-48]</sup>。目前我国在白血病、黑色素瘤、多发性内分泌腺瘤综合征/家族性甲状腺髓样癌的研究中也取得了一定的成绩<sup>[43,49,50]</sup>。

#### 3.3.2 运用新一代测序技术研究肿瘤的案例

新一代测序技术可以用于研究肿瘤 DNA, 以发现核苷酸序列的改变和染色体结构改变, 例如 Harbour 等<sup>[48]</sup>利用 NimbleGen 2.1M 芯片捕获外显子, Genome Analyzer II 测序系统对 31 例转移性葡萄膜黑色素瘤患者测序, 在 84% 的转移性肿瘤中发现了 *BAP1* 基因的失活性体细胞突变。肿瘤 DNA 序列存在染色体易位、倒位、重复和缺失, 随着双向测序的出现, 可以通过组装, 确定 DNA 片段的物理距离的原理研究肿瘤 DNA 的结构变异, Stephen 等<sup>[47]</sup>研究发现许多肿瘤中 2%~3% 的患者存在大量的基因组重排现象。

同时可以通过对原发灶、转移灶和异种移植植物 DNA 测序, 研究肿瘤的进展情况, 例如对基底样腺癌的原发肿瘤和转移灶或异种移植植物测序发现, 在转移灶或异种移植物的突变来源于原发灶, 但是会选择性扩增<sup>[51]</sup>。而小叶乳腺癌的转移灶的多数突变 (约 60%) 不存在原发灶中<sup>[44]</sup>。

## 4 在基因诊断中的运用

由于全基因组外显子测序技术在孟德尔疾病和复杂疾病研究中的优势, 目前已有成功运用该技术进行基因诊断和分子诊断的报道。Choi 等<sup>[52]</sup>通过 NimbleGen 2.1 M 芯片和 Illumina 测序系统, 对疑似 Bartter 综合征的 5 例高加索人进行测序, 结果发现其中 1 例患者的 *SLC26A3* 基因 (为先天性氯腹泻的位点) 的一个错义突变 D652N, 说明该患者是先天性氯腹泻, 而不是 Bartter 综合征, 其临床也支持这个诊断。后期对 39 例疑似 Bartter 综合征和 95 例对照进行 Sanger 测序验证, 发现其中 5 例患者是 *SLC26A3* 纯合子突变, 且临床表现也符合先天性氯腹泻诊断。上述研究表明全基因组外显子测序可以用于发现疾病相关的基因信息, 利用测序信息结合临床特征可对相关遗传病做出正确的诊断, 从而进一步证

明全基因组外显子测序能运用于疾病的分子诊断。除此之外,还有利用全基因组外显子测序的策略对其他临床上难诊断的疾病进行诊断的相关报道,如新生儿糖尿病、难治性炎症性肠病和 Charcot-Marie-Tooth atrophy 综合征的分子水平的诊断<sup>[53-55]</sup>。

## 5 展望

目前,全基因组外显子测序已在孟德尔疾病或罕见综合征的研究中取得了重大突破,证实了全基因组外显子测序对鉴定孟德尔疾病或罕见综合征的致病基因是行之有效的;可以运用在复杂疾病的研究中,目前在肿瘤方面的研究较多,但由于该技术存在一些尚待完善的问题,导致全基因组外显子测序在常见疾病的研究进展缓慢,但随着技术的发展和后续分析手段的丰富,预期在不久的将来,外显子测序会广泛地运用到其他复杂疾病的研究中;也可以用于疾病的分子诊断,为临床一些难以诊断的疾病提供了一种新诊断思路。因此, Biesecker 等<sup>[56]</sup>在 *Nature Genetics* 发表评论说:外显子测序使得医学基因组学成为现实。全基因组外显子测序也存在不足之处:(1) 该技术对结构变异与非编码区变异的研究具有局限性,而结构变异和非编码区变异也可能与疾病相关,可通过全基因组测序研究非编码区变异,同时进行基因组组装,同基因序列对比后发现一些结构变异,如 CNV、indels,此外还可以结合芯片研究检测 CNV;(2) 在目标区域的捕获时,存在捕获不均、捕获偏差等现象,可以通过增加测序深度,获得更多的序列信息进行统计分析,以尽可能的弥补这些偏差;(3) 研究常见疾病的罕见突变时,需要庞大的样本量,也导致了测序的费用的升高;目前用于验证的基因分型的平台如 Sequenom、TaqMan 主要适用于研究常见变异,需要定制芯片或通过 Sanger 测序,但这些方法耗时和价格较昂贵;数据分析的方法仍然不够完善,尚待解决,难以从海量的数据中迅速发现具有重大价值的信息。尽管如此,由于外显子是与疾病及表型相关的最具特征性的区域,并且迄今为止较难评价非编码区域对疾病的影响,所以在全基因组测序费用居高不下的今天,全基因组外显子测序仍然不失为一个很好的选择。外显子测序为阐明疾病的发病机理提供了新的线索,在疾病的基因诊断和致病基因的研究方面有

广阔的前景,为之后的功能研究奠定了理论基础,从而为临床应用铺平了道路。

## 参考文献(References):

- [1] Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? *J Hum Genet*, 2010, 55(4): 195–206.
- [2] McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry*, 2007, 190: 194–199.
- [3] McClellan J, King MC. Genetic heterogeneity in human disease. *Cell*, 2010, 141(2): 210–217.
- [4] Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, 2003, 33(3): 228–237.
- [5] Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeyasinghe S, Krawczak M, Cooper DN. Human gene mutation database (HGMD®): 2003 update. *Hum Mutat*, 2003, 21(6): 577–581.
- [6] Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? *Nat Biotechnol*, 2008, 26(10): 1125–1133.
- [7] Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*, 2009, 27(2): 182–189.
- [8] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*, 2008, 26(10): 1135–1145.
- [9] Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, Magrini VJ, Richt RJ, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Schedl T, Wilson RK, Mardis ER. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods*, 2008, 5(2): 183–188.
- [10] Droege M, Hill B. The Genome Sequencer FLX System—longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol*, 2008, 136(1–2): 3–10.
- [11] Hashimoto S, Qu W, Ahsan B, Ogoshi K, Sasaki A, Nakatani Y, Lee Y, Ogawa M, Ametani A, Suzuki Y, Sugano S, Lee CC, Nutter RC, Morishita S, Matsushima K. High-resolution analysis of the 5'-end transcriptome using a next generation DNA sequencer. *PLoS One*, 2009, 4(1): e4108.

- [12] Rothberg JM, Leamon JH. The development and impact of 454 sequencing. *Nat Biotechnol*, 2008, 26(10): 1117–1124.
- [13] Ng PC, Levy S, Huang JQ, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC. Genetic variation in an individual human exome. *PLoS Genet*, 2008, 4(8): e1000160.
- [14] Li YR, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao HZ, Korneliusson T, Grarup N, Guo YR, Hellman I, Jin X, Li QB, Liu JT, Liu X, Sparso T, Tang MF, Wu HL, Wu RH, Yu C, Zheng HC, Astrup A, Bolund L, Holmkvist J, Jørgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang XQ, Li RQ, Yang HM, Wang J, Hansen T, Pedersen O, Nielsen R, Wang J. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet*, 2010, 42(11): 969–972.
- [15] Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 2009, 461(7261): 272–276.
- [16] Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS. *TGM6* identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain*, 2010, 133(Pt 12): 3510–3518.
- [17] Liu Y, Gao M, Lv YM, Yang X, Ren YQ, Jiang T, Zhang X, Guo BR, Li M, Zhang Q, Zhang P, Zhou FS, Chen G, Yin XY, Zuo XB, Sun LD, Zheng XD, Zhang SM, Liu JJ, Zhou YW, Li YR, Wang J, Yang HM, Yang S, Li RQ, Zhang XJ. Confirmation by Exome Sequencing of the Pathogenic Role of NCSTN Mutations in Acne Inversa (Hidradenitis Suppurativa). *J Invest Dermatol*, 2011, 131(7): 1570–1572.
- [18] Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMILLIN MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura KI, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat Genet*, 2010, 42(9): 790–793.
- [19] Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BBA, Veltman JA. *De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat Genet*, 2010, 42(6): 483–485.
- [20] Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NVAM, Veltman JA, Brunner HG. Exome sequencing identifies *WDR35* variants involved in Sensenbrenner syndrome. *Am J Hum Genet*, 2010, 87(3): 418–423.
- [21] Lalonde E, Albrecht S, Ha KCH, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. *Hum Mutat*, 2010, 31(8): 918–923.
- [22] Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet*, 2010, 87(2): 282–288.
- [23] Simpson MA, Irving MD, Asilmaz E, Gray MJ, Dafou D, Elmslie FV, Mansour S, Holder SE, Brain CE, Burton BK, Kim KH, Pauli RM, Aftimos S, Stewart H, Kim CA, Holder-Espinasse M, Robertson SP, Drake WM, Trembath RC. Mutations in *NOTCH2* cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat Genet*, 2011, 43(4): 303–305.
- [24] Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, Bergmann C, Rohrbach M, Koerber F, Zimmermann K, de Vries P, Wirth B, Schoenau E, Wollnik B, Veltman JA, Hoischen A, Netzer C. Exome sequencing identifies truncating mutations in Human *SERPINF1* in Autosomal-Recessive osteogenesis imperfecta. *Am J Hum Genet*, 2011, 88(3): 362–371.
- [25] Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J, Rolinski B, Iuso A, Biskup S, Schmidt T, Mewes HW, Wittig I, Meitinger T, Zeviani M, Prokisch H. Exome sequencing identifies *ACAD9* mutations as a cause of complex I deficiency. *Nat Genet*, 2010, 42(12): 1131–1134.
- [26] Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 2009, 42(1): 30–35.
- [27] Johnson JO, Gibbs JR, van Maldergem L, Houlden H, Singleton AB. Exome sequencing in brown-vialletto-van laere syndrome. *Am J Hum Genet*, 2010, 87(4): 567–569.
- [28] Krawitz PM, Schweiger MR, Rödelberger C, Marcellis C, Kölsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Köhler S, Jäger M, Grünhagen J, de Condor BJ, Doelken S,

- Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet*, 2010, 42(10): 827–829.
- [29] Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernysky A, Gonzalez E, Rudzicz N, Engert JC, Depristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S. Exome sequencing, *ANGPTL3* mutations, and familial combined Hypolipidemia. *N Engl J Med*, 2010, 363(23): 2220–2227.
- [30] Züchner S, Dallman J, Wen R, Beecham G, Naj A, Farooq A, Kohli MA, Whitehead PL, Hulme W, Konidari I, Edwards YJK, Cai GQ, Peter I, Seo D, Buxbaum JD, Haines JL, Blanton S, Young J, Alfonso E, Vance JM, Lam BL, Pericak-Vance MA. Whole-Exome sequencing links a variant in *DHDDS* to retinitis pigmentosa. *Am J Hum Genet*, 2011, 88(2): 201–206.
- [31] Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein *GPSM2* as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet*, 2010, 87(1): 90–94.
- [32] Otto EA, Hurd TW, Airik R, Chaki M, Zhou WB, Stoetzel C, Patil SB, Levy S, Ghosh AK, Murga-Zamalloa CA, van Reeuwijk J, Letteboer SJF, Sang LY, Giles RH, Liu Q, Coene KLM, Estrada-Cuzcano A, Collin RWJ, McLaughlin HM, Held S, Kasanuki JM, Ramaswami G, Conte J, Lopez I, Washburn J, MacDonald J, Hu JH, Yamashita Y, Maher ER, Guay-Woodford LM, Neumann HPH, Obermüller N, Koenekoop RK, Bergmann C, Bei XS, Lewis RA, Katsanis N, Lopes V, Williams DS, Lyons RH, Dang CV, Brito DA, Dias MB, Zhang XM, Cavalcoli JD, Nürnberg G, Nürnberg P, Pierce EA, Jackson PK, Antignac C, Saunier S, Roepman R, Dollfus H, Khanna H, Hildebrandt F. Candidate exome capture identifies mutation of *SDCCAG8* as the cause of a retinal-renal ciliopathy. *Nat Genet*, 2010, 42(10): 840–850.
- [33] Sirmaci A, Walsh T, Akay H, Spiliopoulos M, Şakalar YB, Hasanefendioğlu-Bayrak A, Duman D, Farooq A, King MC, Tekin M. *MASPI* mutations in patients with facial, umbilical, coccygeal, and auditory findings of carnevale, malpuech, OSA, and michels syndromes. *Am J Hum Genet*, 2010, 87(5): 679–686.
- [34] Ostergaard P, Simpson MA, Brice G, Mansour S, Connell FC, Onoufriadis A, Child AH, Hwang J, Kalidas K, Mortimer PS, Trembath R, Jeffery S. Rapid identification of mutations in *GJC2* in primary lymphoedema using whole exome sequencing combined with linkage analysis with delineation of the phenotype. *J Med Genet*, 2011, 48(4): 251–255.
- [35] Johnson JO, Mandrioli J, Benatar M, Abramzon Y, van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wu J, Ding JH, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurro MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G, Galassi G, Scholz SW, Taylor JP, Restagno G, Chio A, Traynor BJ. Exome sequencing reveals *VCP* mutations as a cause of familial ALS. *Neuron*, 2010, 68(5): 857–864.
- [36] Caliskan M, Chong JX, Uricchio L, Anderson R, Chen P, Sougnez C, Garimella K, Gabriel SB, dePristo MA, Shakir K, Matern D, Das S, Waggoner D, Nicolae DL, Ober C. Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the *TECR* gene on chromosome 19p13. *Hum Mol Genet*, 2010, 20(7): 1285–1289.
- [37] Anastasio N, Ben-Omran T, Teebi A, Ha KCH, Lalonde E, Ali R, Almureikhi M, Der Kaloustian VM, Liu JH, Rosenblatt DS, Majewski J, Jerome-Majewska LA. Mutations in *SCARF2* are responsible for Van Den Ende-Gupta syndrome. *Am J Hum Genet*, 2010, 87(4): 553–559.
- [38] Bolze A, Byun M, McDonald D, Morgan NV, Abhyankar A, Premkumar L, Puel A, Bacon CM, Rieux-Laucat F, Pang K, Britland A, Abel L, Cant A, Maher ER, Riedl SJ, Hambleton S, Casanova JL. Whole-Exome-Sequencing-Based discovery of Human *FADD* deficiency. *Am J Hum Genet*, 2010, 87(6): 873–881.
- [39] Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, Hawkins GA, Guo X, Rotter JI, Chen YD, Wagenknecht LE, Langefeld CD. 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. *Hum Mol Genet*, 2010 19(20): 4112–4120.
- [40] Bilgüvar K, Öztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoğlu D, Tüysüz B, Çağlayan AO, Gökben S, Kaymakçalan H, Barak T, Bakircioğlu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yılmaz S, Dinçer A, Johnson MH, Bronen RA, Koçer N, Per H, Mane S, Pamir MN, Yalçinkaya C, Kumandaş S, Topçu M, Özmen M, Šestan N, Lifton RP, State MW, Günel M. Whole-exome sequencing identifies recessive *WDR62* mutations in severe brain malformations. *Nature*, 2010, 467(7312): 207–210.



- [41] Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, Watanabe K, Yang H, Yuen MM, Knoppers BM, Bobrow M, Cambon-Thomsen A, Dressler LG, Dyke SO, Joly Y, Kato K, Kennedy KL, Nicolás P, Parker MJ, Rial-Sebbag E, Romeo-Casabona CM, Shaw KM, Wallace S, Wiesner GL, Zeps N, Lichter P, Biankin AV, Chabannon C, Chin L, Clément B, de Alava E, Degos F, Ferguson ML, Geary P, Hayes DN, Hudson TJ, Johns AL, Kasprzyk A, Nakagawa H, Penny R, Piris MA, Sarin R, Scarpa A, van de Vijver M, Futreal PA, Aburatani H, Bayés M, Botwell DD, Campbell PJ, Estivill X, Gerhard DS, Grimmond SM, Gut I, Hirst M, López-Otín C, Majumder P, Marra M, McPherson JD, Nakagawa H, Ning Z, Puente XS, Ruan Y, Shibata T, Stratton MR, Stunnenberg HG, Swerdlow H, Velculescu VE, Wilson RK, Xue HH, Yang L, Spellman PT, Bader GD, Boutros PC, Campbell PJ, Flicek P, Getz G, Guigó R, Guo G, Haussler D, Heath S, Hubbard TJ, Jiang T, Jones SM, Li Q, López-Bigas N, Luo R, Muthuswamy L, Ouellette BF, Pearson JV, Puente XS, Quesada V, Raphael BJ, Sander C, Shibata T, Speed TP, Stein LD, Stuart JM, Teague JW, Totoki Y, Tsunoda T, Valencia A, Wheeler DA, Wu H, Zhao S, Zhou G, Stein LD, Guigó R, Hubbard TJ, Joly Y, Jones SM, Kasprzyk A, Lathrop M, López-Bigas N, Ouellette BF, Spellman PT, Teague JW, Thomas G, Valencia A, Yoshida T, Kennedy KL, Axton M, Dyke SO, Futreal PA, Gerhard DS, Gunter C, Guyer M, Hudson TJ, McPherson JD, Miller LJ, Ozenberger B, Shaw KM, Kasprzyk A, Stein LD, Zhang J, Haider SA, Wang J, Yung CK, Cross A, Liang Y, Gnaneshan S, Guberman J, Hsu J, Bobrow M, Chalmers DR, Hasel KW, Joly Y, Kaan TS, Kennedy KL, Knoppers BM, Lowrance WW, Masui T, Nicolás P, Rial-Sebbag E, Rodriguez LL, Vergely C, Yoshida T, Grimmond SM, Biankin AV, Bowtell DD, Cloonan N, deFazio A, Eshleman JR, Etemadmoghadam D, Gardiner BA, Kench JG, Scarpa A, Sutherland RL, Tempero MA, Waddell NJ, Wilson PJ, McPherson JD, Gallinger S, Tsao MS, Shaw PA, Petersen GM, Mukhopadhyay D, DePinho RA, Thayer S, Shazand K, Beck T, Sam M, Timms L, Ballin V, Lu Y, Ji J, Zhang X, Chen F, Hu X, Zhou G, Yang Q, Tian G, Zhang L, Xing X, Li X, Zhu Z, Yu Y, Yu J, Yang H, Lathrop M, Tost J, Brennan P, Holcatova I, Zaridze D, Brazma A, Egevard L, Prokhorchouk E, Banks RE, Uhlén M, Cambon-Thomsen A, Viksna J, Ponten F, Skryabin K, Stratton MR, Futreal PA, Birney E, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Martin S, Reis-Filho JS, Richardson AL, Sotiriou C, Stunnenberg HG, Thoms G, van de Vijver M, van't Veer L, Calvo F, Birnbaum D, Blanche H, Boucher P, Boyault S, Chabannon C, Gut I, Masson-Jacquemier JD, Lathrop M, Pauporté I, Pivot X, Vincent-Salomon A, Tabone E, Theillet C, Thomas G, Tost J, Treilleux I, Bioulac-Sage P, Clément B, Decaens T, Degos F, Franco D, Gut I, Gut M, Heath S, Lathrop M, Samuel D, Thomas G, Zucman-Rossi J, Lichter P, Eils R, Brors B, Korbel JO, Korshunov A, Landgraf P, Lehrach H, Pfister S, Radlwimmer B, Reifemberger G, Taylor MD, von Kalle C, Majumder PP, Sarin R, Rao TS, Bhan MK, Scarpa A, Pedersoli P, Lawlor RA, Delledonne M, Bardelli A, Biankin AV, Grimmond SM, Gress T, Klimstra D, Zamboni G, Shibata T, Nakamura Y, Nakagawa H, Kusada J, Tsunoda T, Miyano S, Aburatani H, Kato K, Fujimoto A, Yoshida T, Campo E, López-Otín C, Estivill X, Guigó R, de Sanjosé S, Piris MA, Monserrat E, González-Díaz M, Puente XS, Jares P, Valencia A, Himmelbaue H, Quesada V, Bea S, Stratton MR, Futreal PA, Campbell PJ, Vincent-Salomon A, Richardson AL, Reis-Filho JS, van de Vijver M, Thomas G, Masson-Jacquemier JD, Aparicio S, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Stunnenberg HG, van't Veer L, Easton DF, Spellman PT, Martin S, Barker AD, Chin L, Collins FS, Compton CC, Ferguson ML, Gerhard DS, Getz G, Gunter C, Guttmacher A, Guyer M, Hayes DN, Lander ES, Ozenberger B, Penny R, Peterson J, Sander C, Shaw KM, Speed TP, Spellman PT, Vockley JG, Wheeler DA, Wilson RK, Hudson TJ, Chin L, Knoppers BM, Lander ES, Lichter P, Stein LD, Stratton MR, Anderson W, Barker AD, Bell C, Bobrow M, Burke W, Collins FS, Compton CC, DePinho RA, Easton DF, Futreal PA, Gerhard DS, Green AR, Guyer M, Hamilton SR, Hubbard TJ, Kallioniemi OP, Kennedy KL, Ley TJ, Liu ET, Lu Y, Majumder P, Marra M, Ozenberger B, Peterson J, Schafer AJ, Spellman PT, Stunnenberg HG, Wainwright BJ, Wilson RK, Yang H. International network of cancer genome projects. *Nature*, 2010, 464(7291): 993-998.
- [42] Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao CN, Jouanguy E, Feske S, Abel L, Casanova JL. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J Exp Med*, 2010, 207(11): 2307-2312.
- [43] Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute

- monocytic leukemia. *Nat Genet*, 2011, 43(4): 309–315.
- [44] Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao YJ, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, 2009, 461(7265): 809–813.
- [45] Lee W, Jiang ZS, Liu JF, Haverty PM, Guan YH, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennermer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang ZM. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*, 2010, 465(7297): 473–477.
- [46] Saarinen S, Aavikko M, Aittomäki K, Launonen V, Lehtonen R, Franssila K, Lehtonen HJ, Kaasinen E, Broderick P, Tarkkanen J, Bain BJ, Bauduer F, Unal A, Swerdlow AJ, Cooke R, Makinen MJ, Houlston R, Vahteristo P, Aaltonen LA. Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. *Blood*, 2011.
- [47] Stephens PJ, Greenman CD, Fu BY, Yang FT, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia MM, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 2011, 144(1): 27–40.
- [48] Harbour JW, Onken MD, Roberson EDO, Duan SH, Cao L, Worley LA, Council ML, Matatall KA, Helms C, Bowcock AM. Frequent mutation of *BAP1* in metastasizing uveal melanomas. *Science*, 2010, 330(6009): 1410–1413.
- [49] Wei XM, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y. Exome sequencing identifies *GRIN2A* as frequently mutated in melanoma. *Nat Genet*, 2011, 43(5): 442–446.
- [50] Qi XP, Ma JM, Du ZF, Ying RB, Fei J, Jin HY, Han JS, Wang JQ, Chen XL, Chen CY, Liu WT, Lu JJ, Zhang JG, Zhang XN. *RET* germline mutations identified by exome sequencing in a Chinese multiple endocrine neoplasia Type 2A/Familial medullary thyroid carcinoma family. *PLoS One*, 2011, 6(5): e20353.
- [51] Ding L, Ellis MJ, Li SQ, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang QY, Chen L, Lin L, Wendl MC, McMichael JF, Margrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du FY, Hawkins AE, O’Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*, 2010, 464(7291): 999–1005.
- [52] Choi M, Scholl UI, Ji WZ, Liu TW, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Özen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA*, 2009, 106(45): 19096–19101.
- [53] Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med*, 2010, 13(3): 255–262.
- [54] Bonnefond A, Durand E, Sand O, De Graeve F, Gallina S, Busiah K, Lobbens S, Simon A, Bellanné-Chantelot C, Létourneau L, Scharfmann R, Delplanque J, Sladek R, Polak M, Vaxillaire M, Froguel P. Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. *PLoS One*, 2010, 5(10): e13630.
- [55] Montenegro G, Powell E, Huang J, Spezziani F, Edwards YJK, Beecham G, Hulme W, Siskind C, Vance J, Shy M, Züchner S. Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family. *Ann Neurol*, 2011, 69(3): 464–470.
- [56] Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet*, 2011, 43(1): 13–14.