

DOI:10.3969/j.issn.1002-2694.2015.09.017

高通量测序技术在临床病毒学领域的应用

张拥军

摘要:本文总结了近几年高通量测序平台的发展现状,比较了不同测序平台的性能及特点。同时还对目前高通量测序平台在近两年临床病毒学领域的应用,包括病原学诊断、分子流行病学、宏基因组学和临床治疗转归等几个方面,逐一举例进行了阐述。籍此加深专业人员对高通量测序平台的认识,促进该技术在临床病毒学的转化与应用。

关键词:高通量测序;宏基因组学;病毒学

中图分类号:R373

文献标识码:A

文章编号:1002-2694(2015)09-0864-06

Applications of next generation sequencing technologies on clinical virology

ZHANG Yong-jun

(Fujian Center for Disease Control and Prevention, Fuzhou 350001, China)

Abstract: In this review, current status of several next generation sequencing (NGS) platforms in recent years were summarized, while features and characterization of each NGS platform were compared. Applications of NGS technologies in the field of clinical virology including etiological diagnosis, molecular epidemiology, metagenomics and progression of clinical treatment, etc., was illustrated. This review would facilitate comprehensive understanding of diverse NGS platforms for professionals, and promote the transformation and application of NGS technologies in clinical virology.

Keywords: next generation sequencing; metagenomics; virology

2004年454生命科学公司率先推出商品化的GS-20测序仪,宣告了高通量测序(NGS)时代的到来。经过近10年的不断更新,从技术层面看,各种NGS平台日新月异,仪器试剂逐渐成熟,测序速度、读长及通量得到了大幅度提升,相对测序成本则急剧下降。在应用方面,NGS技术愈来愈多地运用于全基因组测序、宏基因组学(metagenomics)、转录组(transcriptome)测序、目标序列再测序(resequencing)、从头测序(*de novo* sequencing)、染色体免疫共沉淀测序(ChIP-Seq)、外显子组测序(exome sequencing)以及各种RNA测序等方面,已经渗透到生命科学多种领域如临床医学、农业、生态学、法医、微生物学、癌症研究等。本文拟简要介绍目前NGS平台的发展现状,并对近两年NGS平台在临床病毒学方面的部分应用进行概述。

1 NGS平台现状简介

近10年里各种NGS平台不断推陈出新,按时间先后顺序,主要出现了以下主流NGS平台。首先,454生命科学公司于2004年开始销售基于焦磷

酸测序的GS-20测序仪,首次实现了大规模并行测序,能够同时测定约20万个片段,平均读长为110 bp。在被罗氏公司并购之后,454公司又将GS-FLX、GS-Junior等机型推向市场,平均读长达到700~1 000 bp;2005年Solexa公司研制了基因组分析仪(Genome Analyzer),被Illumina公司收购后,又在此基础上陆续发展出HiSeq2000、HiSeq2500、NextSeq500、MiSeq、HiSeq X Ten等系列测序仪,满足了不同层次测序的需求,最大读长已经能够达到600 bp;2007年应用生物系统公司(ABD)推出第一款SOLiD测序系统;而生命技术公司(Life Technologies)收购ABI之后,不仅接连推出了SOLiD3、SOLiD4、SOLiD5500等机型,还分别于2009年和2011年推出了商品化的Helicos测序仪和Ion Torrent个人基因组测序仪,近年又研制了通量更高的Ion Proton、Ion Chef测序仪;作为第三代单分子测序仪的代表,太平洋生物科学公司(Pacific Biosciences)于2011年推出了单分子实时测序仪PacBio-RS,测序读长达到数千个碱基;英国的Oxford Nanopore Technologies公司则研制成功读长更长的GridION、MinION测序仪^[1-3]。

不同于传统的以毛细管电泳为基础的第一代测

序技术,NGS 平台基本上采用边合成边测序(sequencing-by-synthesis, SBS)的策略,也无需预先知测序模板的遗传背景,就能够同时完成数十万到数亿个片段的测序,成为真正意义上的高通量。从测序原理上,454、Ion Torrent 和 SOLiD 体系都采用乳液 PCR 制备样品文库,每一个珠子结合一段 DNA 模板,并均匀分配到测序孔或者玻片,实现一个珠子对应一个测序片段(read)。Illumina/Solexa 技术则是先将测序文库 DNA 变性,一端与固定在玻片上的寡核苷酸接头结合,每个片段的另一端与玻片上的互补接头结合,形成桥式结构,因此将这个在玻片进行的模板扩增称为桥式 PCR。从测序反应来看,454 体系采用焦磷酸测序,测序需要的酶和引物都在反应孔中,每次加入 1 种未标记的核苷酸,有碱基掺入时,释放出的焦磷酸盐产生发光,逐一被光学摄像头记录下来。Ion Torrent 平台与 454 体系类似,只是检测信号为碱基掺入时释放的 H 离子,不需要光学设备记录信号,因此也称半导体测序仪。该平台先后提供了 314、316 和 318 芯片供用户选择,平均读长为 100~400 bp。Illumina/Solexa 体系的试剂包括引物、DNA 聚合酶及 4 种不同标记的可逆终止核苷酸。每掺入一个核苷酸,根据不同荧光颜色记录下来。SOLiD 体系采用的是连接反应来完成测序。通过测序引物与接头杂交,其 5' 端再与相邻序列上杂交的寡核苷酸连接,不同荧光标记的寡核苷酸混合物互相竞争与引物的连接,每次掺入 2 个碱基。PacBio-RS 测序仪也是采取 SBS 策略,不同的是,其测序文库模板不需要扩增,实现了

单分子实时测序。单链基因组 DNA 片段与带发卡结构的接头连接,形成环状分子,然后与固定在该仪器独有的纳米小孔——ZMW 上的 DNA 聚合酶结合,通过记录掺入核苷酸上不同荧光基团达到测序的目的。PacBio-RS 读长能够达到 20 kb 以上,平均为 5 kb。同样,GridION 和 MinION 体系也不涉及模板扩增过程,属于无标记的单分子测序。当单链 DNA 分子通过一种蛋白质纳米孔,链上每一个碱基产生的电信号依次被记录下来。这种测序方法读长可以达到数十万个碱基,是目前读长最长的高通量测序方法^[1-3]。

为了满足一些小实验室的需求,并将 NGS 平台运用于临床诊断市场,这些公司还生产了一些通量较小的台式 NGS 平台。例如,2009 年罗氏公司推出的 GS Junior 测序仪,可以在 10 h 内获得 35 Mb 序列数据,平均读长 400 bp。2011 年 Illumina 公司的 MiSeq 测序仪投放市场,27 h 可以得到 1.5 Gb 数据,平均读长为 150 bp。目前升级后的试剂盒能够产生 2 千 5 百万个片段,每个片段可达 2×300 bp 的读长,也就是一个反应能够得到共计 15 Gb 的序列。而同年生命技术公司推出的台式 NGS 平台 Ion Torrent 能够在 2 h 内获得测序结果,至今仍然是测序时间最短的台式 NGS 仪器。这些平台由于单次运行成本低,数小时内可以获取序列结果,特别适合一些小基因组和 PCR 产物的测序,具有临床辅助诊断的潜力^[1-3]。

以下对目前各种 NGS 平台的一些性能特征进行了简单比较,见表 1。

表 1 不同 NGS 平台主要性能特征比较

Tab. 1 Comparison on performance features of diverse NGS platforms

Manufacturers	Instrument models	Reads (M)	Mean read-length (bp)	Yield (Gb)	Run-time h	Amplification	Advantages & disadvantages
Roche-454	454 GS-FLX	1	700-1000	0.7	29	乳液 PCR	读长较长,试剂相对成本高
	454 GS-Junior	0.1	400	0.04	10		
Life Technologies	Ion Torrent	0.475-4.75	200-400	0.095-1.9	2.3-7.3	乳液 PCR	仪器价廉,快速,读长较短
	Ion Proton	70-500	175	12-87	4-6		
	Ion Chef	—	—	—	—		
Life Technologies	SOLiD-5500	1410	110	155	8 days	乳液 PCR	通量大,测序周期长,仪器昂贵,读长较短
Illumina	MiSeq	1-22	50-600	19-120	5-55	桥式 PCR	操作简便,通量高,成本低
	NextSeq500	130-400	75-300	15-500	11-30		
	HiSeq2500	300-2 000	50-300	—	10 h-11 days		
Pacific Biosciences	PacBio-RS	0.03	3 k	0.09	2	无扩增,单分子测序	读长长,快速,通量中等,错误率高,仪器昂贵
Oxford Nanopore Technologies	Nanopore GridION MinION	0.1-10	9 k-10 k	0.9-100	6-	无扩增,单分子测序	读长最长,快速,仪器昂贵

2 在临床病毒学领域的应用

病毒作为一种古老的物种,广泛存在于自然界,几乎能够在任何物种寄生,与人类健康息息相关。虽然大部分病毒感染没有出现明显的临床症状,少数病毒能够引起人类多种疾病,表现出包括发热、腹泻、出血、出疹、呼吸道症状、神经系统症状等,甚至一些病毒感染与肿瘤的发生有关。随着 100 多年来人们对病毒认识的不断深入以及生物技术的发展,人类逐渐积累了多种病毒性疾病的鉴别诊断能力,掌握了部分病毒的致病机制及流行规律,研制出一些病毒的特异性疫苗,并设计生产了各种抗病毒化学药物,极大地降低了病毒性疾病的危害性。在 NGS 技术日渐成熟的过程中,全球科学家也在不断尝试将此技术及时转化到病毒学研究中去,特别是在未知病原学诊断、已知病毒性疾病的分子流行病学、从基因组水平探讨一些病毒的致病机理以及一些疾病的治疗及预后方面,利用生物技术进步促进人类健康。

2.1 发现未知新病原 大多数病毒性疾病都有传染性,能够通过各种途径或媒介传播,因此无论是发达国家还是发展中国家,传染病依然是威胁人类健康的因素之一。近 30 年来涌现的各类新发再发传染病暴发事件中,相当一部分与新病毒特别是 RNA 病毒的出现相关。传统的鉴定未知新病毒的方法,包括病毒分离培养、血清学检测、免疫电镜等,都需要相当长的周期,而一些致病性病毒暂时还没有合适的体外培养体系。NGS 技术不依赖病原体遗传背景和高通量的特点,具有得天独厚的优势,尤其是能够从未经纯培养的复杂样品中,捕捉任何疑似病原体的蛛丝马迹。因此近几年一些新出现传染病疫情的病因调查,时常出现 NGS 平台的身影^[4]。

首先,近几年在以下一些重要新发病毒病的发现过程中,各种 NGS 平台发挥了巨大作用。

1) 新的蜱传播 Thogotovirus^[5] 2014 年春,1 名 50 多岁的美国堪萨斯州男性在户外工作时被蜱叮咬,数日后发病,有恶心、发烧、疲劳、腹泻等症状。患者血小板减少,白细胞减少,起初予以强力霉素治疗,11 天后死于多器官衰竭。美国 CDC 研究人员采用分子和血清学方法检测 10 余种已知蜱相关病原体均为阴性。但用蚀斑减少中和试验测定前几年新发现的 Heartland 病毒抗体时发现一种全新的病毒,经电镜观察疑似为正粘病毒。随后用 ION TORRENT 高通量测序平台,检测到多个节段基因组序列,与 Dhori 病毒相似度达 70%,由此建立特异的实时荧光 RT-PCR 检方法,并在患者血样中证

实了相关片段的存在。因此认为发现了一个正粘病毒科 Thogotovirus 病毒属的新成员。

2) SFTS 调查^[6] 严重发热伴血小板减少综合征(SFTS)是 2010 年首先在中国中部和东北几个省出现的一种新流行性疾病,曾经通过传统的毛细管测序方法证实其病原体为一种新的布尼亚病毒(SFTSV),主要经蜱传播。2012 年秋,日本出现一例 SFTS 病例,患者死于多器官衰竭。研究人员进行回顾性调查时,将患者样品进行病毒分离培养,培养上清采用 MiSeq 平台测序,发现了多个与 SFTSV 同源的 DNA 序列片段。NGS 测序结果为证实日本同样存在 SFTS 病例提供了直接证据。

2) MERS 冠状病毒疫情^[7] 荷兰研究人员 2012 年从 1 例 60 岁沙特急性肺炎转肾衰的死亡病例痰液中,分离到 1 株未知冠状病毒(HCoV-EMC)。通过 454 GS-FLX 测序平台,得到病毒基因组约 90% 的序列。序列分析显示该毒株与蝙蝠冠状病毒 HKU4 和 HKU5 遗传关系接近,为一种全新的 beta 冠状病毒。该病毒临床表现与 2003 年 SARS 相似,后来被命名为中东呼吸综合征(MERS)冠状病毒。2015 年 5 月底 MERS 疫情在韩国暴发,中国出现了首例来自韩国的输入病例。中、韩两国研究人员分别用 ION TORRENT 平台和 Illumina 平台测定病毒基因组序列,用于追溯病毒来源,经初步分析该毒株与今年 2—5 月沙特流行毒株最接近^[8]。至今(2015 年 6 月 12 日)已经累计有 1 289 例实验室确诊 MERS 病例通报给 WHO,至少 455 例死亡^[9]。

3) H7N7 禽流感^[10] 最近一个意大利团队报道了一起人感染高致病性禽流感 H7N7 病毒的事件。2013 年 8—9 月间,意大利几个农场出现了高致病性 H7N7 禽流感疫情,在处理疫情过程中,先后有 3 名工作人员出现不同程度的结膜炎。调查人员采集患者样品,通过实时荧光 PCR 证实了 H7N7 病毒感染,并用 Ion Torrent PGM 平台对一株病毒进行了全基因组测序。HA 和 NA 基因种系分析显示,毒株与近 3 年欧洲野鸟及家禽中流行的低致病性 H7 毒株类似,而内部基因序列则与该地区鸡群流行的 H7N7 毒株高度相似,也表明此病毒由鸡直接传播到人。

除了以上新发病毒性传染病,同样,NGS 平台对一些常见病毒性感染病例,也能够从背景复杂样品中,更准确地探索相关致病因子。

4) 呼吸道感染调查^[11] 丹麦科学家为了调查儿童急性呼吸道病毒感染的病原体,采集了 500 份

上/下呼吸道感染样品,对其中 92 份样品采用 GS-FLX 平台测序,获得 4 个完整病毒基因组,分别属于人副流感病毒 4 型(HPIV4)的 4a 和 4b 亚型。根据测序结果,他们设计实时荧光 RT-PCR 引物,对所有样品进行筛查,发现 HPIV4 阳性率为 2.6%。而后他们还利用 PacBio RS 平台对代表样品进行了测序,结果证明该平台能够用于复杂临床样品的病毒序列测定。

5) 病毒性脑炎调查^[12] 能够引起病毒性脑膜炎的病毒超过 100 种,包括 HSV-1、流感病毒、肠道病毒以及一些虫媒病毒等。由于涉及的病原种类繁多,只有少数病例被真正明确了实验室诊断。美国犹他大学医学院尝试用 NGS 技术调查一些脑炎病例中病毒类型。他们选择了 7 例脑炎死亡病例的脑组织,提取 RNA 并用随机引物进行逆转录。在 HiSeq 2000 测序仪上,每一例患者和正常对照样品均得到 50~90 M 读长为 50 bp 的片段。在排除人体组织片段以后,7 例患者脑组织里发现有 36 个片段与病毒有关,分别涉及以下病毒科:疱疹病毒科(17)、副粘病毒科(10)、痘病毒科(6)、戊肝病毒科(2)和 黄病毒科(1)。但序列拼接以后 7 例患者样品只有 5 例存在 66~4 019 bp 的片段。经过传统 PCR 扩增及毛细管测序,证实 2 例为麻疹病毒,3 例为 HSV-1,另外 2 例未检出致病病毒序列。上述结果证明深度测序能够用于脑炎患者的病原体调查。

2.2 追踪已知病原的变异及多样性 分子流行病学是采用各种分子生物学手段,在分子或基因水平阐明疾病的病因及其相关的致病过程,并研究疾病的防治和促进健康的策略和措施的科学。NGS 技术应用于病毒性疾病的分子流行病学研究,可以一次性完成一些基因组较大的病毒如疱疹病毒、冠状病毒等的全基因组序列测定,也可以同时进行数十个甚至上百个样本中靶基因的扩增子测序,还可以通过深度测序阐明病毒在疾病发展过程中如何进化变异以及是否出现准种等。要完成这些目标,依靠传统的毛细管电泳测序,几乎是不可能实现的,并且试剂和时间成本会高很多。

具有代表性的典型应用有:

1) 2014 年西非 EBOV 疫情^[13] 埃博拉病毒(EBOV)自 1976 年在非洲发现,主要以非洲东部的散发病例为主。始于 2014 年 2 月的疫情,是首次出现在非洲西海岸几个国家(塞拉利昂、几内亚、利比里亚、尼日利亚),也是迄今为止最大规模的埃博拉疫情。截至 2014 年 8 月 31 日,已经报告 3 685 病例,死亡 1 641 例。哈佛大学、Broad 研究所与塞拉

利昂一家医院合作,从 5 月底至 6 月中旬确诊的 78 例 EBOV 感染病例中,采用高通量测序的手段,共获得了 99 个完整的病毒基因组序列。他们的测序过程主要在 HiSeq2500 和 PacBio-RS 平台完成。结果显示,这次西非疫情毒株与以往 EBOV 基因组存在 341 个碱基变异,而塞拉利昂患者存在 263 个宿主间单个核苷酸变异(iSNV)。根据与以往暴发相关的 20 个毒株基因组进行种系发生比较,表明 2014 年这些西非 EBOV 是近 10 年中从非洲中部传播过来。同时,这些毒株的遗传相似性提示,此次暴发起因是接触单一的 EBOV 天然宿主,随后引起人与人的传播蔓延。序列数据还显示,塞拉利昂差不多同时传入了两种遗传背景不同的 EBOV,该发现与收集到的流行病学调查信息一致。

2) 荷兰 H7N7 人禽流感疫情^[14] 2003 年 2—5 月,荷兰发生高致病性禽流感 H7N7 病毒暴发,9 周内波及 255 个养鸡场,导致近 3 千万只鸡被扑杀。另外有 89 人被感染,1 名兽医死亡。调查人员起初用第一代测序技术观察到死亡病例病毒株存在 PB2-E627K 和 HA-K416R 的突变。时隔几年之后 NGS 技术成熟之时,他们又利用 Illumina 公司的 GAIIX 和 HiSeq 2000 平台,对原始样品进行了深度测序。从原来养鸡场几份鸡样品中并未发现上述突变,但意外在禽样品中发现流感病毒缺陷 RNA 片段,表明病毒在感染禽类过程中合成了缺损的干扰病毒。在几份人样品中,还发现 PB2-E627K 突变随感染过程检出频率增高,强烈支持人类适应标志 PB2-E627K 是个体感染过程中逐渐形成的,减少人类暴露于禽流感病毒的机会,对于降低病毒适应人类是非常重要的。

3) 乙型流感病毒基因组^[15] 自 1980 年代以来,乙型流感病毒(IBV)同时存在两种抗原性和遗传背景不同谱系的毒株:B/Victoria/2/87-like 和 B/Yamagata/16/88-like。为了对大量乙型流感病毒进行基因组测序并建立序列数据库,美国 J. Craig Venter 研究所(JCVD)建立了一套通用的基因组扩增方法,以便测序、构建疫苗和反向遗传学研究。他们将扩增产物在 Ion Torrent PGM、HiSeq 2000、MiSeq 以及 454 (Roche)平台进行测序。经过 1 000 多个 IBV 临床样品的测试,证明该方法灵敏、有效、不依赖序列,适用于 NGS 测序基因组诊断以及快速克隆反向遗传学质粒。

4) 鼻咽癌中 EBV 基因组^[16] 虽然 40 年前就已经知道未分化鼻咽癌(NPC)与 EB 病毒感染有关,但迄今仅测定了 10 株 EB 病毒毒株全基因组序

列。EBV 基因组全长近 200 kb,若用第一代毛细管电泳测序,成本高、耗时长。为了探索 NPC 样品中 EB 病毒基因组变异及多样性,香港大学团队首先在 Miseq 测序仪上测试了 3 株参比 EBV 序列,而后在此基础上利用 GA IIx 平台完成了来自患者病理切片的 8 株 EBV 基因组序列。与参比毒株比较,共发现 1 736 处变异(包括 1 601 个碱基替换、64 处碱基插入、71 处碱基删除),编码多种蛋白的基因出现歧义突变。以上结果说明,从全基因组水平能够发现 NPC 患者 EBV 基因组的序列多样性,通过比较正常样品和 NPC 样品中 EBV 的基因组变异,有助于评价某些位点变异与 NPC 致病机制的联系。

2.3 宏基因组学(病毒组) 宏基因组学是研究不同生态环境中所有微生物组成,遗传结构及群落功能的科学。近几年有关人类宏基因组研究先后涉及皮肤、口腔、血液、粪便等,病毒宏基因组(viral metagenome)或者病毒组(virome)指的是人类、动物、植物或者特定环境样品中所有病毒的集合。人类病毒组包括引起人类急性/持续性或潜伏感染的病毒,以及诸如内源性逆转录病毒之类的整合在人类基因组上面的病毒,开展这类研究有助于揭示新病毒及其与已知未知疾病的联系^[17-19]。NGS 平台对于这类相对较小的基因组进行深度测序,也不需要为每一个种类一一进行分离培养,具有与其它方法不过比拟的优越性。

根据不同组织部位的宏基因组研究结果,健康人体组织存在的病毒很多属于噬菌体,主要与人体胃肠道的正常菌群有关。除了噬菌体,不同组织中占优势的病毒种类也不一致。人体皮肤主要检出过包括 Merkel 细胞多瘤病毒(一种与皮肤癌有关的 DNA 病毒)在内的一些多瘤病毒(HPyV6、7、9),以及多种 β -和 γ -人乳头状瘤病毒(HPV)和圆环病毒科(Circoviridae)的一些成员;呼吸道常见 EB 病毒和逆转录病毒;口腔和鼻咽部可以检测到多种呼吸道病毒,如人呼吸道合胞病毒(RSV)、流感病毒和鼻病毒,健康儿童呼吸道样品还发现腺病毒、小 RNA 病毒和冠状病毒;健康人体粪便中检出过单链 RNA 病毒、单链 DNA 病毒、双链 DNA 病毒以及逆转录病毒等,包括多种病毒如 anellovirus、picobirnavirus 和 parechovirus,而博卡病毒、C 组和 F 组腺病毒、Aichi 病毒、星状病毒、轮状病毒等则较少检出。还有部分为植物病毒,可能与膳食来源有关。另外,70%的健康人血浆中可能有非致病性的 anelloviruses,一些人疱疹病毒如 EB 病毒、CMV、HHV6、HHV7 也经常出现在血液中。疾病状态下,血液中

可能出现其它种类的病毒:如淋巴瘤或肉瘤(EBV、HHV8),脑炎(HSV、CMV、EBV 和 HHV6),消化道症状(CMV、HHV6)^[17-20]。

不同于上述横断面调查,美国宾州大学的研究团队^[21]为了研究人类肠道中病毒组的进化,在 2.5 年内对一名健康人连续采样 16 次,获取了 24 份粪便样品。通过 Illumina 公司 HiSeq2000 平台的深度宏基因组测序,得到 560 亿病毒序列数据,拼接后共有 478 个 contigs,其中 60 个 contigs 能够组装成环状,意味着得到了这些病毒全部环状基因组。87%属于未知病毒,13%属于以下噬菌体科(Microviridae、Podoviridae、Myoviridae、Siphoviridae)。进一步分析还发现,其中 80%基因组在 2.5 年中长期稳定存在。因此,他们认为,人体肠道中病毒的多样性来自两个方面,少部分是广泛存在的病毒组,其余是快速进化的长期病毒组成员。

2.4 指导临床治疗及转归 一些病毒感染机体之后,病毒和宿主之间的相互作用,会导致病毒基因组部分位点发生变异,以逃逸免疫系统对病毒的清除。而在机体接受抗病毒药物治疗的过程中,一些病毒可能在药物作用靶位产生突变,从而对这类药物具有耐药性。甚至有的流行毒株,本身有可能就对部分药物耐受。这些位点如果突变频率不高,采用传统的毛细管电泳测序很难捕捉到。只有覆盖靶基因位点达数十倍甚至上百倍的深度测序,才能清晰展现突变位点,为临床治疗提供及时有效的参考依据。目前这方面的应用,主要集中在 HIV、HCV 感染患者的抗病毒治疗方面。

1) 长期接受抗逆转录病毒药物治疗患者体内的 HIV 整合 对于 HIV 感染患者,主要采用联合抗逆转录病毒治疗(cART)来抑制病毒的复制,而患者体内持续存在的被感染细胞是阻碍 HIV 感染治疗的关键。为了探索 HIV 前病毒 DNA 在人类基因组的整合位点,美国国家癌症研究所(NCI)的科学家^[22]对 5 名长期(7.2~14.5 年)接受 cART 治疗的 HIV 感染者进行了观察。通过提取患者外周血单核细胞(PBMCs)或 CD4⁺ T 细胞的 DNA,在 MiSeq 平台测定患者基因组 DNA 中病毒/宿主序列的连接点和突破点。他们发现患者治疗前及刚参加 cART 的时刻,体内存在多个病毒群体;而经过长时间的 cART 治疗(平均 11.7 年)之后,无论在 DNA 水平还是在 RNA 水平,都出现相同的病毒序列。5 名患者中,累计发现的整合位点涉及 985 个不同基因。以上结果说明 HIV 的整合位点对于患者体内感染细胞的克隆化扩散以及持续存在发挥了

重要作用。

2)复杂的 HCV 基因型^[23] HCV 分为 7 个基因型及 67 个亚型,不同基因型之间存在 30% 核苷酸序列差异,不同亚型之间差异达 20%。目前 HCV 感染的治疗方案取决于基因型,因此准确区分基因型至关重要。研究人员选择了世界各地 2 363 例患者中,有 12 例 INNO-LiPA 分型属于基因型 2,而 NS5B 序列分型却属于基因型 1 的患者,首先对其中 8 例用传统测序方法测病毒核心区序列证实为基因 2 型。然而,他们随后利用 MiSeq 平台测定病毒基因组序列发现,这些病毒实际属于重组基因型 2/1。其中 11 株病毒重组区域发生在 NS2/NS3 交界处。这些患者对索非布韦/利巴韦林治疗方案的应答则类似于基因 1 型患者。

3)HCV 耐药性评估模型^[24] 为了计算药物对病毒群体的作用以及群体对个体变异的抗性,研究人员选择了 HCV 和干扰素抗性作为证据。他们选择 16 名未接受过任何治疗的基因 1 型慢性 HCV 患者,注射 IFN- α 后 48 h 内采血 9 次。然后扩增病毒基因组 E1/E2 交界处包含 HCV 高变区 1 (HVR1)区域的 309 nt 片段,将 PCR 产物进行 GS FLX 平台测序,共测定 1.7 M 片段,平均每个时间点有 12 332 个片段。根据出现变异的频率变化及病毒群体滴度改变计算出干扰素抗性系数,认为病毒群体抗性在 IFN- α 2a/利巴韦林治疗第 12 周和干扰素变异的出现高度相关。因此认为,该方法能够在短时间内准确评估因单纯病毒滴度变化或者相对病毒变异频率引起的药物抗性。

3 小结与展望

综上所述,经过这 10 年的持续发展,NGS 技术和实验体系日臻完善,在临床病毒学的多个领域得到了应用。尽管与 10 年前相比,测序成本下降了许多,但目前的 NGS 技术应用还主要局限于研究阶段,离大规模临床应用还有相当远的距离。只有在进一步降低实验成本的前提下,实验室技术人员建立起可行的标准化操作方案,生物信息学人员及时提供合理可靠的分析结果,让临床医生能够正确理解 NGS 数据,并充分认识到 NGS 技术为患者带来的便利,NGS 平台才能真正进入临床,早日服务于需要的患者群体。

参考文献:

[1]van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology[J]. Trends Genet, 2014, 30

(9): 418-426. DOI: 10.1016/j.tig.2014.07.001

- [2]Quinones-Mateu ME, Avila S, Reyes-Teran G, et al. Deep sequencing: Becoming a critical tool in clinical virology[J]. J Clin Virol, 2014, 61(1): 9-19. DOI: 10.1016/j.jcv.2014.06.013
- [3]Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications[J]. Biochim Biophys Acta, 2014, pii: S0925-4439(14)00180-X. DOI: 10.1016/j.bbdis.2014.06.015
- [4]Chiu CY. Viral pathogen discovery[J]. Curr Opin Microbiol, 2013, 16(4): 468-478. DOI: 10.1016/j.mib.2013.05.001
- [5]Kosoy OI, Lambert AJ, Hawkinson DJ, et al. Novel *Thogotovirus* species associated with febrile illness and death, United States, 2014 [J]. Emerg Infect Dis, 2015. DOI: 10.3201/eid2105.150150
- [6]Takahashi T, Maeda K, Suzuki T, et al. The first identification and retrospective study of Severe Fever with Thrombocytopenia Syndrome in Japan[J]. J Infect Dis, 2014, 209(6): 816-827. DOI: 10.1093/infdis/jit603
- [7]Zaki AM, van Boheemen S, Bestebroer TM, et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia [J]. N Engl J Med, 2012, 367: 1814-1820. DOI: 10.1056/NEJMoa1211721
- [8]virological.org. MERS coronavirus [EB/OL]. (2015-06-08) [2015-07-02]. <http://virological.org/t/preliminary-analysis-of-middle-east-respiratory-syndrome-coronavirus-mers-cov-sequences-from-korea-and-china/143>
- [9]World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV)-Republic of Korea [EB/OL]. (2015-06-12) [2015-06-18]. <http://www.who.int/csr/don/12-june-2015-mers-korea/en/>
- [10]Puzelli S, Rossini G, Facchini M, et al. Human infection with highly pathogenic A (H7N7) avian influenza virus, Italy, 2013 [J]. Emerg Infect Dis, 2014. DOI: 10.3201/eid2010.140512
- [11]Alquezar-Planas DE, Mourie T, Bruhn CAW, et al. Discovery of a divergent HPIV4 from respiratory secretions using second and third generation metagenomic sequencing [J]. Sci Rep, 2013, 2468. DOI: 10.1038/srep02468
- [12]Chan BK, Wilson T, Fischer KF, et al. Deep sequencing to identify the causes of viral encephalitis[J]. PLoS One, 2014, 9(4): e93993. DOI: 10.1371/journal.pone.0093993
- [13]Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak[J]. Science, 2014. DOI: 10.1126/science.1259657
- [14]Jonges M, Welkers MR, Jeeninga RE, et al. Emergence of the virulence-associated PB2 E627K substitution in a fatal human case of highly pathogenic avian influenza virus A(H7N7) infection as determined by Illumina ultra-deep sequencing[J]. J Virol, 2014, 88(3): 1694-1702. DOI: 10.1128/JVI.02044-13
- [15]Zhou B, Lin X, Wang W, et al. Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics[J]. J Clin Microbiol, 2014, 52(5): 1330-1337. DOI: 10.1128/JCM.03265-13

- 1005-6661. 2005. 04. 003 (in Chinese)
- 曹建平, 韩海勃, 刘述先, 等. 日本血吸虫硫氧还蛋白 DNA 疫苗的构建和鉴定[J]. 中国血吸虫病防治杂志, 2005, 17(4): 250-254. DOI: 10.3969/j.issn.1005-6661.2005.04.003
- [44] Cao JP, Xu YX, Han HB, et al. Study on the protective effect of *Schistosoma japonicum thioredoxin* DNA vaccine against challenge infection in mice[J]. Chin J Schisto Ctrl, 2005, 17(4): 255-259. DOI: 10.3969/j.issn.1005-6661.2005.04.004 (in Chinese)
- 曹建平, 徐徐信, 韩海勃, 等. 日本血吸虫硫氧还蛋白 DNA 疫苗小鼠保护性免疫研究[J]. 中国血吸虫病防治杂志, 2005, 17(4): 255-259. DOI: 10.3969/j.issn.1005-6661.2005.05.004
- [45] Han HB, Cao JP, Liu SX, et al. Protective immunity induced by recombinant *Schistosoma japonicum* thioredoxin in mice[J]. Chin J Parasitol Parasit Dis, 2006, 24(3): 175-178. DOI: 10.3969/j.issn.1000-7423.2006.03.004 (in Chinese)
- 韩海勃, 曹建平, 刘述先, 等. 日本血吸虫重组硫氧还蛋白小鼠保护性免疫研究[J]. 中国寄生虫学与寄生虫病杂志, 2006, 24(3): DOI: 10.3969/j.issn.1000-7423.2006.03.004
- [46] Wen YL. Studies on cloning, expression and immunological reaction of the gene encoding thioredoxin of chinese stain of *Schistosoma japonicum* in picllia pastor[D]. Nanjing: Nanjing Medical University, 2007. (in Chinese)
- 闻礼永. 日本血吸虫大陆株硫氧还蛋白在毕氏酵母菌的克隆表达及其免疫反应性研究[D]. 南京医科大学, 2007.
- 收稿日期: 2015-01-29; 修回日期: 2015-06-04
-
- (上接第 869 页)
- [16] Kwok H, Wu CW, Palser AL, et al. Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples[J]. J Virol, 2014, 88(18): 10662-10672. DOI: 10.1128/JVI.01665-14
- [17] Hofer U. Viral evolution: Variation in the gut virome[J]. Nat Rev Microbiol, 2013, 11(9): 596. DOI: 10.1038/nrmicro3092
- [18] Delwart E. A roadmap to the human virome[J]. PLoS Pathog, 2013, 9(2): e1003146. DOI: 10.1371/journal.ppat.1003146
- [19] Lecuit M, Eloit M. The human virome: new tools and concepts [J]. Trends Microbiol, 2013, 21(10): 510-515. DOI: 10.1016/j.tim.2013.07.001
- [20] Duerkop BA, Hooper LV. Resident viruses and their interactions with the immune system[J]. Nat Immunol, 2013, 14(7): 654-659. DOI: 10.1038/ni.2614
- [21] Minot S, Bryson A, Chehoud C, et al. Rapid evolution of the human gut virome[J]. Proc Natl Acad Sci U S A, 2013, 110(30): 12450-12455. DOI: 10.1073/pnas.1300833110
- [22] Maldarelli F, Wu X, Su L, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells[J]. Science, 2014, 345(6193): 179-183. DOI: 10.1126/science.1254194
- [23] Hedskog C, Doehle B, Chodavarapu K, et al. Characterization of hepatitis C virus inter-genotypic recombinant strains and associated virologic response to sofosbuvir/ribavirin[J]. Hepatology, 2014. DOI: 10.1002/hep.27361
- [24] Campo DS, Skums P, Dimitrova Z, et al. Drug resistance of a viral population and its individual intrahost variants during the first 48 hours of therapy[J]. Clin Pharmacol Ther, 2014, 95(6): 627-635. DOI: 10.1038/clpt.2014.20
- 收稿日期: 2015-02-13; 修回日期: 2015-07-20