

高通量测序技术的发展及其在临床检测中的应用现状

王玉静, 陆梓涔, 陈俊煜, 陈毅歆*

(厦门大学公共卫生学院, 国家传染病诊断试剂与疫苗工程技术研究中心, 福建 厦门 361102)

摘要: 随着分子生物学的迅速发展, 检测技术也随之迅猛发展, 如今的高通量测序技术已经历不同代次的更迭, 性能和技术日趋成熟与稳定, 在临床样本的病原体鉴定、肿瘤检测、疾病诊断、遗传病检测等领域为精准医学提供了一个高效的分析检测工具。通过对 DNA 或 RNA 序列的测定与拼接, 能够发掘更多与疾病、微生物相关的基因组学信息, 迅速把握病情的原因以及与患者临床特征相关的感染情况, 对于疾病的快速精准诊断、治疗效果与预后情况的预测都有十分重要的帮助。本文将对测序技术的发展、不同高通量测序技术平台的介绍及其在临床中的应用进行综述。

关键词: 高通量测序; 二代测序; 三代测序; 临床检测

中图分类号: R 331 **文献标志码:** A

测序技术的出现, 直观而深刻地揭露了核酸分子的深层信息, 为人类进一步探索基因结构与功能提供了决定性的技术手段。高通量测序在过去的约二十年中得到了迅猛的发展, 这些测序技术也在过去的二十年中也实现了成功的商业化, 与之相关的基础应用、科研探究以及临床应用也随之大幅增加^[1]。随着“精准医疗”概念的提出, 临床应用上对于高通量测序的需求越来越大, 病原学诊断、检测与遗传病、肿瘤等疾病的精准诊断等应用领域对于高通量测序技术的要求也越来越高。而在高通量测序技术出现之后, 发生的几次世界性范围的传染性疫情中, 高通量测序技术也逐渐扮演起越来越重要的角色^[2]。高通量测序技术作为精准医疗的重要基石, 对精准医疗做出了极大的贡献, 在临床相关的病原微生物检测、临床肿瘤学、16S RNA 以及 ITS 测序、遗传疾病检测、传染病监测以及新型病毒的发掘等方面发挥了高通量测序技术的优势与特点^[3]。对于高通量测序技术, 其发展历程、不同平台介绍、测序原理的比较以及不同领域的应用都是较受关注的焦点, 后文将对以上关注焦点进行详细介绍与讨论; 同时, 本综述也将对高通量测序技术在临床检测中的应用进行详细阐述。

收稿日期: 2020-06-20 **录用日期:** 2020-08-10

基金项目: 福建省科技重大专项 (2020YZ014001)

通信作者: yxchen2008@xmu.edu.cn

1 测序技术的发展历程

Frederick Sanger 于 1975 年发明了“双脱氧链终止法”基因测序技术，这是科学史上出现的第一个基因测序技术^[4]；另外一种基因测序技术是 1977 年 Walter Gilbert 发明的“化学测序法”，这两种测序技术均作为一代测序的标志性技术而广泛应用，其中“双脱氧链终止法”因为操作更为简便、更为稳定，而更被广泛应用^[5]。在过去的 20 年内，基因测序技术有了较大的进步与发展，一代测序仍然稳定占据部分市场，二代测序、三代测序也高速飞快占据了较大市场份额，基因测序技术在规模、通量以及应用上都有了极大的发展。一代测序虽然在过去五十年内占有着极大的测序市场，然而其存在通量低、数据产出较低以及成本较高等问题，虽然一代测序在目前测序市场上有着不可取代的地位，还是仍然无法满足当前分子生物学、医学研究以及临床诊断对于高通量、高效率、高产出的测序需求。二代测序相对于一代测序而言测序的准确率略微降低，但是通量增加、产出增加，可以实现同时对多个样本进行测序，在单位时间内数据产出量比一代测序实现了数量级的增长。自 2005 年第一台二代测序仪器罗氏 454 焦磷酸测序平台诞生以来，随后陆续推出的二代测序技术平台包括有：2006 年 Illumina 公司推出的 Solexa 测序平台、2007 年美国 ABI 公司推出的 SOLiD 测序平台和 2010 年美国 Life Technologies 公司推出的半导体测序平台^[6]，各自占据一定的市场份额。其中 454 Life Science 后来被 Roche 收购，但是由于二代测序市场的竞争日趋激烈，以及较新的测序方法的出现，该测序方法被逐渐淘汰。2013 年 Roche 宣布关闭 454 测序业务，并于 2016 年全面终止相关服务，454 测序仪被市场淘汰。2008 年 invitrogen 和美国 ABI 公司合并成立 Life Technologies 公司，开始发展半导体测序，占领了部分测序市场，随之 SOLiD 测序也逐渐淡出市场。2006 年开始 Illumina 进入了二代测序市场，并且在此后的 10 年时间，Illumina 占领了大部分的测序市场，于 2010 年开始陆续推出的 HiSeq 系列测序仪，更是迅速成为二代测序平台中主流测序平台。

作为中国高通量测序的先驱，华大基因也于 2014 年推出首款二代测序仪——BGISEQ-1000，在 2016 年也陆续推出了 BGISEQ-500 等型号的测序仪。该测序平台在大规模 DNA 测序和 small RNA 分析中的能力已得到证明，但 BGISEQ-500 平台在转录组分析中的性能仍有待提升^[7]。每一种测序平台都有自身特点，在数据产出量、测序读长、测序准确率以及测序成本等多个方面各有不同的表现^[8]。

二代测序技术平台尽管在测序通量、数据产出量以及应用领域上相较于一代测序有显著优势，但仍然存在一定的短板，如测序读长较短导致在测序过程中会产生大量的高度碎片化的重复片段，尤其在进行大基因组测序时，测序拼接成为一个较大的挑战。且相较于一代测序而言，二代测序所需的测序时间显著增加，尚不能完全满足临床样本的快速诊断需要^[9]。在此背景下，可满足长读长和快速测序需要的三代测序（Third-generation Sequencing, TGS）平台应运而生，分别是 2008 年英国 ONT 公司首次推出了一款以纳米孔单分子测序为原理的测序仪器，但是当时该平台还不够稳定，无

法投入正常的使用。2014年，英国 ONT 公司推出了 MinION 测序仪，可供用户使用。2008年美国 Helicos Bioscience 公司的以单分子测序（Single Molecule Sequencing, SMS）技术为原理的 Heliscope 测序平台发布上市，2009年美国 Pacific Bioscience 公司推出了单分子实时测序技术（Single Molecule Real-Time, SMRT）。三代测序平台可以直接对给定的模板 DNA 或者 RNA 进行测序，实现真正意义上的实时测序，当核酸模板通过测序仪即可产生信号。相较于前两代测序平台，三代测序平台主要的改善有：1) 读长变长，可在一个反应内读取上千上万碱基的读长，理论上可以达到无限长；2) 测序流程简化，测序时间减少，在文库构建以及上机测序等流程上精简了流程，减少了样本测序时间；3) 避免了 PCR 扩增技术造成的扩增偏好；4) 可直接测定碱基上的修饰情况，三代测序平台可以直接鉴定碱基上的修饰情况，例如碱基甲基化^[10]。图 1 展示了基因测序发展历程中的里程碑事件。

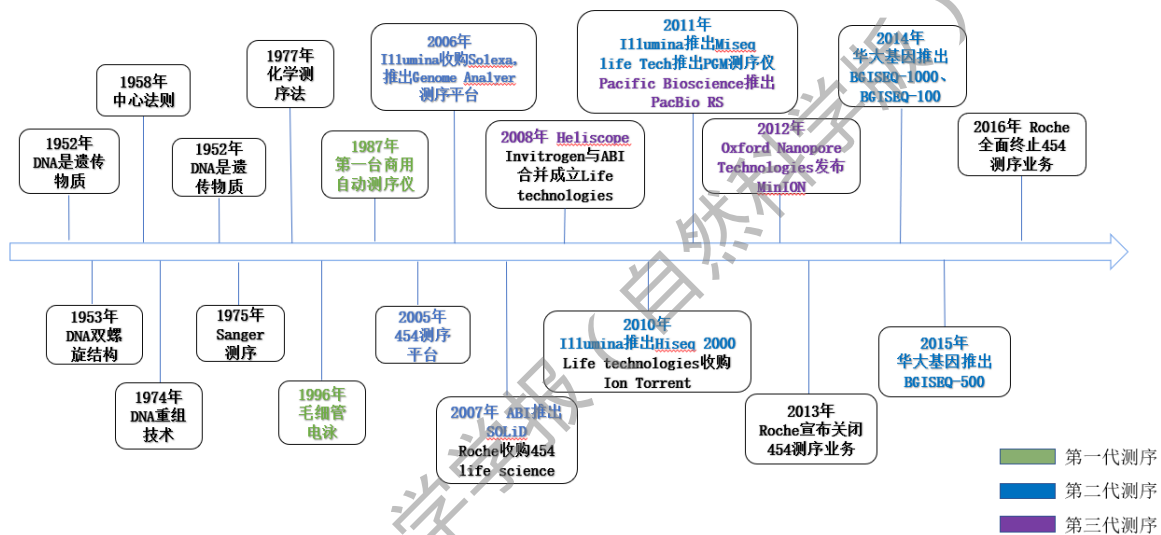


图 1 测序技术发展时间轴

Fig.1 Timeline of sequencing technology development

2 主流二代测序平台的发展概况

一代测序技术运用多年仍然蓬勃生机，二代测序技术更是在短时间内迅速崛起成为市场上最主流，这项基因测序技术在近年来也被广泛应用于临床病原体鉴定诊断中^[11]。不同于 Sanger 测序，二代测序将酶促 DNA 反应、碱基测序与数据收集同步进行，因此可以同时数千条到数十亿条 DNA 模板进行测序^[12]。以下三家测序平台因为不同的测序原理在各方面存在一定的差异，但是由于平台侧重点不同，都在不同时间段成为当时较为主流的测序平台，并且侧重应用于不同的领域。

2.1 罗氏 454 焦磷酸测序平台

罗氏公司的 454 焦磷酸测序平台是国际上第一台相对较为成熟的二代测序平台，属于循环微阵

列法平台，其测序技术基础是边合成边测序（sequencing by synthesis, SBS）技术^[13]。该项测序技术的测序原理主要依靠荧光信号的生物发光，是将模板进行 PCR 扩增之后，与相应的引物杂交之后，并且与三磷酸腺苷双磷酸酶、DNA 聚合酶、ATP 硫酸化酶、荧光素酶、底物荧光素酶和 5'磷酸硫腺苷共同孵育，然后进行相应的酶促反应，在每一次的实时测序实验中，模板只会与一种 dNTP 进行配对反应，在此酶促反应中，DNA 聚合酶以该 dNTP 作为原料在合成互补链的过程中，会释放出等摩尔数的焦磷酸基团（PPi）^[13]。

454 测序技术的主要优势在于测序时间较短，且准确率较高，可达 99%，在单位时间内产生的片段数量多。该测序平台在一次测序工作中可以产生 100 万条序列，序列的平均长度 400 bp，数据总量约 500 Mb。454 测序平台已经被应用到多个方面，均取得了较为理想的结果。454 Life Sciences 后来被 Roche 收购，Roche 在 454 测序平台的基础上又相继推出 GS FLX^[14]，由于后期二代测序市场的新的测序方法不断更迭以及该平台的测序成本较高等原因，该测序平台已于 2016 年宣布全面停止相关的测序服务，被淘汰出二代测序市场。

2.2 Illumina 测序平台

Illumina 测序仪也称为 Solexa 分析仪，测序原理是检测的原理与 Sanger 测序法类似，将合成核酸的原料 dNTP 用 4 种不同的荧光进行标记，并且藕联上可逆的终止剂，固相基质上可以容纳数百万的模板克隆，每个固相基质上可以同时读取 10 亿个碱基。2005 年，Solexa 收购合并了一家仪器公司 Lynx Therapeutics，新的公司成功地将 Solexa 原型转化为商业测序仪器，2006 年，推出第一个 Solexa 测序仪——Genome Analyzer。该测序平台以其高精度度、高灵敏度、高特异性以及相对较低成本的特点，并且在 2010 年成功推出 HiSeq 系列测序仪。目前在遗传疾病分析、肿瘤癌症检测以及功能基因组测序等领域占据主要的测序市场。Illumina HiSeq 系列测序仪具有 PE150 的较长读长，Illumina 测序平台的优势主要在于其测序精准度最高可达 99.9%，而且相较于其他二代测序平台测序成本也较低，但是其也有相应的缺点——序列读长较短^[15]。

2.3 SOLiD 高通量测序仪

2006 年 7 月，美国 ABI 公司推出 SOLiD 测序平台，该平台的基本原理的特点在于每一步测序反应都是通过连接反应完成的，通过 PCR 反应进行平行扩增测序。SOLiD 测序平台支持两种测序文库，一种是与 Illumina 测序平台的文库构建类似，均先将 DNA 模板片段化，在片段的 DNA 模板两端加上接头，即成功构建文库；第二种是配对末端文库，依靠酶切反应加上接头，成功构建文库^[16]。2010 年年末 ABI 发布了第五代测序系统——SOLiD 5500x1 测序系统，SOLiD 5500x1 测序系统在读长、精准度以及数据产出量都实现了较大的进步，分别达到了 85 bp、99.99% 和 30 G。在该测序平台未退出市场之前是二代测序平台中精准度最高的平台^[16]。

2008年在Life Technoloies公司收购Ion Torrent之后,开始陆续推出Ion PGM和Ion Proton系列测序仪,由于公司的发展侧重点需要,该公司目前主推Ion PGM和Ion Proton系列测序仪,因此SOLiD测序平台逐渐淡出市场。2010年Life Technoloies公司发布Ion PGM测序仪,目前有三款芯片,芯片在不断改进,测序通量也在不断增加;2012年发布Ion Proton测序仪,拓展了该系列测序仪更多领域的应用;2013年Life Technoloies公司又被Thermo Fisher公司收购;2015年9月1日,发布了新产品S5系列Ion S5/the S5 XL,是Proton和PGM相续合而产生的产品,相比较于Proton, PGM.新系列更加容易操作,且节省了较多的测序时间^[3]。

2.4 不同二代测序平台比较分析

多种不同的二代测序平台在不同方面也有不同的表现,其中罗氏454测序仪作为最早问世的二代测序仪在读长上较长,运行较为快速,然而检测成本较高,且设备较大,可及性不高,且错误率也相对较高,目前该测序仪在测序市场中已经停产;Illumina测序仪作为目前市场上最广泛应用的二代测序仪读长较短,运行较慢,成本也相对而言较高,设备较大占据空间,对于实验人员要求较高,但是测序错误率较低,适用于全基因组测序、宏基因组测序,应用领域多^[7];ABI的SOLiD测序仪也是曾经较为常见的二代测序仪,测序准确率较高,然而读长较短,运行较慢,常用于外显子测序,基因突变测序,目前该测序平台也已经淡出市场^[12]。

3 三代测序平台的发展现状

随着二代测序平台的成功应用,三代测序平台开始陆续推出,不同于二代测序平台的部分特点,三代测序在测序读长上作了更大的改善,测序时间也相应的减少,测序流程变得更为简便,测序设备更为便携,测序成本更小。三代测序不依赖于PCR扩增技术,这一测序平台的测序技术最大特点就是单分子测序^[7]。市面上出现的上单分子测序平台,分别为美国Heliscope BioScience公司的SMS技术^[18]和美国Pacific Bioscience的SMRT技术^[17],VisiGen Biotechnologies公司的FRET(Fluorescence resonance energy transfer)技术^[19]以及英国ONT公司所推出的纳米孔单分子测序技术^[10]。

3.1 SMS 测序平台

2008年,美国Helicos Bioscience公司推出HeliScope单分子测序平台,这是继二代测序平台之后出现的第一个可以商品化应用的三代测序仪,其测序的主要原理是单分子测序技术,是一种基于光学信号的边合成边测序技术,但是不同于二代测序的一点是该方法不依赖于PCR扩增技术,先随机将待测模板进行打断与筛选,在对片段化模板进行末端修复之后在片段3'末端连接上50 bp结合有荧光标记的poly(A)尾巴,经过连接接头的文库可以通过末端poly(A)尾巴结合固定在固相基质的Oligo d(T)探针上,类似于Solexa测序,该方法也需要将荧光染料标记的四种dNTP依次加

入微反应中，在 DNA 聚合酶的催化反应下，通过碱基互补配对，释放出相应的荧光信号，最后依靠 ICCD 相机进行光学信号的收集^[16]，在测序上也避免了扩增时引入的碱基错配以及扩增偏好性。该测序方法也存在相应的不足，就是对于光学信号收集的设备要求较高，并且在测序过程中由于信号较弱容易产生测序误差，导致准确性降低；因此该平台为了提高精准度，采取了两次测序（two-pass sequencing），增加了测序成本^[20]，然而由于该平台初始读长较短，约为 32 bp，且测序成本较高，测序准确率较低，错误率高达 1%，因此该测序平台并未得到广泛的应用，2012 年底，Helicos 正式提出申请破产保护。

3.2 SMRT 测序平台

美国 Pacific Bioscience 公司推出的单分子实时测序技术是目前三代测序平台中最为广泛应用的一项测序技术。SMRT 测序技术相较于其他测序平台而言有非常大的优势，该方法同样是基于对单个 DNA 分子进行测序，采用了四种荧光标记的四种 dNTP 以及零级波导（zero-mode waveguides, ZMW）的纳米结构作为测序技术的主要基础。零级波导这种纳米结构是一种孔状纳米光电结构，光线在通过 ZMW 之后会呈现指数级衰减，被衰减的光线最终只能使得孔内靠近基质的部分被照亮。ZMW 作为测序的微反应器，会提前在微反应器中结合上测序反应所需要的 $\Phi 29$ DNA 聚合酶，在构建文库时，将待测模板与引物结合，混合四种荧光标记的 dNTP 一同加入到微反应器 ZMW 中。测序反应过程中，待测模板 DNA 在以四种荧光标记的 dNTP 作为原料进行合成时，所连接的 dNTP 会因为反应而在 ZMW 底部短暂停留，荧光收集设备则可以收集到配对 dNTP 的荧光信号，从而实现测序^[21]。该平台在读长上进行了实现了较大的突破，其中 PacBio RSII 测序平台最长读长能够达到 30 kb，平均读长约为 8.5 kb，且该平台也具有三代测序平台普遍共有的优势——测序流程更为简便，构建文库时间缩短，不依赖于 PCR 扩增技术。然而与 SMS 测序技术类似，该测序技术同样依赖于单分子产生的荧光信号进行测序，因此测序的准确性也偏低，最高仅可达到 87.5%，尽管通过增加测序次数以及后期数据分析矫正，准确率可以提高，但是相对于 Sanger 测序以及二代测序，准确率仍然相对来说较低^[21]。

3.3 纳米孔单分子测序平台

2014 年，英国 ONT 公司推出了第一个商用的测序平台——MinION，该测序平台的主要测序原理是基于待测模板通过生物纳米孔时不同碱基产生的不同电位差而实现电信号向碱基信号的转变，Nanopore 测序系统主要由纳米孔、薄膜以及马达蛋白组成，其中的马达蛋白是一种 DNA 解旋酶，在构建文库的时候，马达蛋白与接头会一同连接在待测模板的一端。当将制备好的文库滴加到纳米孔上时，马达蛋白会通过解旋作用将双链 DNA 变为单链通过纳米孔，ATCG 四个碱基通过纳米孔会产生不同的电位差，这种电信号会被传导电子元件（application-specific integrated circuit, ASIC）以及

MinKNOW 接受并初级处理^[22]。该测序平台的序列读长与 PacBio 测序平台相似，可达到 10 kb，理论上可以达到无限长。然而相较于 PacBio 测序平台，MinION 测序平台的错读率更高，准确率在 65%~88% 之间。

前期当使用 9.4 版本芯片或者其他版本芯片 Flow cell 进行测序时，测序准确率非常低，大约只有 90%。后续平台推出 9.5 版本 Flow cells 芯片并且采用 1D² 建库方式，在一定程度上提升了测序准确率^[23]。

3.4 FRET 测序平台

RET 测序技术主要是基于荧光共振能量转移（fluorescence resonance energy transfer）原理，该测序平台在对样本核酸进行测序时，测序过程中 4 种脱氧核苷酸分子会被四种不同的荧光受体所标记，随着测序引物延伸，4 种不同的荧光受体会发出特异的荧光，不同的荧光分别代表不同的 4 种脱氧核苷酸分子。该测序平台由 VisiGen Biotechnologies 公司研发并推出，读长较长，平均读长在 1 500 bp 以上，测序准确率相对于其他三代测序平台较高，并且测序时长较短，测序过程较短。但是该平台因为缺乏具体应用的技术参数，因此并未得到广泛的应用^[9]。

3.5 不同三代测序平台的比较分析

由于不同测序平台之间的测序原理与建库方式等有一定的不同，因此各个测序平台之间侧重应用的领域也不尽相同，并且不同测序平台的测序成本、测序时间以及测序准确率也有一定的差异，如表 1 所示。三代测序平台读长都较长，且不依赖于 PCR 方式构建文库，测序成本较低，测序时间较短；然而相对于二代测序平台，三代测序平台测序错误率都较高，后续数据处理分析非常依赖于处理软件与数据库的选择与使用。三代测序平台中英国 ONT 公司的 Nanopore 测序仪对实验人员要求较小，可及性极高。然而对于目前的三代测序而言，如何提高测序的准确性是较受关注的方面，目前主流三代测序公司均在测序设备稳定性以及后续数据处理分析上进行了非常大的改进与完善。

表 1 三代测序平台对比

Tab. 1 Comparison of third-generation sequencing platforms

测序平台	技术原理	测序成本	读长/bp	运行时间	单次运行读取最多序列	便携性	单次测序错误率 (%)	综合评价	应用领域
Helicos BioSciences HeliScope	单分子测序	低	32*	8 d	7.5×10 ⁸	差	1	读长较短，运行快速，成本低，测序错误率相对其他三代测序平台较低	较大基因组测序
Pacific Biosciences	单分子实时测序	低	964*	<1 d	3.5×10 ⁴ ~ 7.5×10 ⁴	差	13	读长较长，检测周期短，成本较低，测序错误率较高	转录组测序，较大基因组测序

Nanopore MinION	纳 米 孔 单 分 子 测 序	较低	> 150 kb	<1 d	1.1×10 ⁴ ~ 4.7×10 ⁴	好	38	读长长, 检测周期 较短, 成本较低, 测序错误率高	大基因组测序, RNA 测序, 基 因修饰检测
FRET	荧 光 共 振 能 量 转 移	-	>1 500	20 min	-	差	1	读长较长, 检测周 期短, 测序准确率 较高	应用极少

注: *平均读长, 部分数据参考文献[10, 20, 26]。

4 高通量测序平台的临床应用

正如前言, Sanger 测序作为传统检测方法中的较为典型的方法, 由于该平台的短板, 其应用的领域较为有限。而高通量测序作为近几年较为火热的测序技术, 在各大领域有着广泛的应用以及突出的效果^[24], 比如临床预测、诊断、治疗有关的领域。如表 2 所示, 本文将从与临床有关的不同领域对高通量测序的应用进行详细的介绍以及阐述。

表 2 高通量测序平台的临床应用

Tab. 2 Clinical application of high throughput sequencing platform

临床应用领域	应用方向	主要测序平台	应用实例
临床微生物病原学 鉴定	已知、未知病原体的种属鉴定; 病原体全基因组拼接	Illumina Miseq/Hiseq、 Pacbio 等	MERS、2019-nCoV、PRV 等 新型病原体的发掘、全基因 组拼接等
临床微生物耐药诊 断	微生物耐药性的研究、微生物与 药物间相互作用	Illumina Miseq/Hiseq 等	幽门螺旋杆菌克拉霉素耐药 基因靶向性检测、微生物组 学数据库的构建等
临床肿瘤诊治	新肿瘤靶标的发掘、肿瘤细胞来 源、肿瘤基因的低频突变	Illumina Miseq/Hiseq 等	癌症基因组图谱重大科研项 目的创立、国际癌症基因组 计划的进行等
遗传疾病诊断检测	遗传病诊断、产前筛查与诊断、 试管胚胎等植入性胚胎遗传学 诊断-	Illumina、Ion Torrent、 Nnanopore 等	智力障碍或者残疾畸形儿童 的产前筛查、21 三体综合征 产前筛查等

4.1 临床微生物病原学鉴定

细菌、真菌、支原体、衣原体、寄生虫、病毒等微生物与人的健康系统稳定有着息息相关的联系, 人体中有着由细菌、真菌、病毒等微生物组成的最为庞大与复杂的胃肠道系统, 且人体许多疾病的发生都与微生物系统的失调或者微生物的入侵有着极其紧密的关系, 而高通量测序技术的出现为庞大而又复杂的微生物菌群的鉴定检测与研究提供了有力的技术支持^[25]。针对微生物病原学检

测,基于测序的策略的不同,主要可以分为以下3项:全基因组测序(whole-genome sequencing, WGS)、靶向下一代测序(targeted next-generation sequencing, tNGS)和宏基因组测序(metagenomic next-generation sequencing, mNGS)。

1) 全基因组测序是对未知基因组序列的物种进行个体的基因组测序,目前二代测序和三代测序均支持对个体物种进行全基因组测序,对于微生物的全基因组测序,可以准确从科属种上对样本中的微生物进行鉴定分析,并且可以根据对耐药基因的分析比对、毒力基因的比对预测该种微生物的耐药情况以及预后情况,同时最重要的一点,在发掘出罕见或者未知的微生物方面,全基因组测序必不可少^[26]。

Zhou 等^[27]基于二代测序平台以及扩增子测序等技术对一种未知的新型的病毒进行了全基因组测序,该病毒是一种新型的与 HKU2 相似的冠状病毒,与蝙蝠体内寄宿的某种冠状病毒序列有 98.48% 的一致性。研究人员基于二代测序平台对该新型冠状病毒进行了全基因组测序,迅速掌握了该病毒的传染性与毒力的分子学基础,研究了该病毒的衍化过程,为该病毒的治疗与防控提供了有利的测序分析手段。

Wu 等^[28]研究人员从武汉新冠患者的样本中提取到了新型冠状病毒的 RNA,并且通过 Illumina MiniSeq 对提取到的 RNA 进行了测序并进行了全基因组的拼接。研究人员共得到了 56 565 928 个读取序列,对以上读取序列进行初步组装拼接之后,形成了 384 096 个重叠群,其中长度最长的一个重叠群(30 474 个核苷酸)具有较高的丰度,与一株从蝙蝠分离得到的 bat SL-CoVZC45 的基因序列有 89.1% 的一致性。科研人员基于二代测序平台从患者样本中测序筛查出新型病毒,并通过组装拼接比对发现该病毒与一株从蝙蝠分离得到的冠状病毒有较高的一致性,初步判断为一种新的冠状病毒。二代测序平台在新的病毒中发现上有十分巨大的潜力,对于流行性疾病的诊断与治疗都有十分重大的意义。

对于新发传染病未知病原体的发掘、微生物耐药性分析方面,Illumina 测序平台相对而言使用较为广泛并且表现良好,未知病原体的全基因组测序目前主要依靠 Illumina Miseq 等测序仪完成,该测序仪在迅速、精准、高效获取病原体基因以及基因比对分析方面有着较好的表现。

2) 靶向目标测序是指对某物种的某特定区域或某特定功能的基因进行靶向测序,而对于临床微生物病原学检测来说,靶向目标测序主要是集中在对细菌的 16S rRNA 基因进行靶向目标测序,16S rRNA 基因是原核生物所特有的基因片段,由于该基因片段在细菌中普遍存在,既具有相对保守的区域又有高度可变的区域,因此经常被用作细菌鉴定分类的标准^[29]。根据 16S rRNA 基因序列的保守区域设计相应的引物,对可变区进行靶向扩增,并基于高通量测序平台对可变区进行靶向目标测序,后期借助生物信息学分析手段对样本中细菌进行精准的种属鉴定,类似地针对真菌的内转录组间隔区(internal transcribed spacer, ITS)进行靶向目标测序也可以对真菌进行精准的鉴定与分类。针对于

16S rRNA 和 ITS 的靶向目标测序, 不仅对微生物可以进行准确的科属种的鉴定分析, 同样可以从序列信息中得到毒力信息, 对抗生素药物的耐药性、代谢学等对临床学以及流行病学具有一定的效用。

Schloss 等^[30]通过 PacBio 单分子测序技术对 16S rRNA 基因进行测序, 研究人员将从社区和自然环境中获取的人的粪便样本、老鼠粪便样本以及土壤样本进行混合, 并对样本进行相应的靶向测序, 研究人员主要获取分析了 V4、V3-V5、V1-V3、V1-V5、V1-V6 以及 V1-V9 等可变区域的测序数据。并且研究人员基于数据分析处理方式将该测序平台对 16S rRNA 中的可变区 V1-V9 的测序错误率从 0.69% 降低至 0.027%。研究人员对物种多样性、微生物组成和微生物进化开展研究, 对种属鉴定的精准度再次提升。

3) 宏基因组测序是指从临床样本或者环境样本中直接提取全部微生物的核酸, 构建宏基因组测序文库, 并且进行测序。该方法不需要进行菌株分离培养, 因此大程度地避免了分离效率低、灵敏度低的问题。针对环境样本, 例如土壤、海水等中复杂的微生物群落, 针对人体口腔、粪便、肠道等部位样本, 通过宏基因组测序发现样本中一些无法培养或者难以培养的微生物种类, 发掘复杂样本中未知的罕见的微生物种类^[31]。

Stephen 等^[32]研究人员发现一名男子临床症状为左下叶肺炎, 口腔分泌物增加, 有多个组织坏死。前期抗生素治疗方案效果不明显, 研究人员基于 Illumina Miseq 测序仪通过对患者肺泡灌洗液样本进行了宏基因组测序。研究人员对于测序数据进行了较为完善的处理分析: 使用 PANDAseq 对配对的短序列进行组装拼接; 通过软件 USEARCH v6 对拼接全长进行读取比对; 将所得的组装好的数据片段与核糖体数据库项目 (the Ribosomal Database Project) 中具有较为典型、代表性的数据库序列进行比对分类; 使用 DeeNuRP 和 Taxtastic 对数据进行过滤与注释。通过对患者样本的宏基因组测序, 研究人员发现样本中测序到核粒梭形杆菌、假单胞菌等, 与之后细菌培养结果一致。

4.2 临床微生物耐药诊断

临床微生物病原学的诊断检测具有十分重大的意义, 临床上对于微生物相关的研究与检测需求也十分巨大。除了上文提到的复杂样本的微生物种属鉴定, 高通量测序还广泛应用于微生物耐药性 (antimicrobial resistance, AMR) 研究中, 临床上医院获得性感染的细菌的用药治疗一直是医疗难点, 主要是由于在长期的抗生物用药筛选中, 许多病原体通过基因突变而获得了对不同抗生素的耐药性, 对临床的治疗造成了极大的影响。

例如, 目前通过高通量测序深度挖掘并组建的药学相关的微生物组学数据库, 根据该大数据库已经发现有超过 60 种药物与微生物之间存在相互作用, 后期高通量测序也可持续进行挖掘微生物的耐药性和微生物与药物间的相互作用与人类遗传变异的相关性^[33]。

初亚男等^[34]等基于焦磷酸测序技术建立了一种对幽门螺旋杆菌克拉霉素耐药基因靶向性检测的方法, 在获知耐药情况的同时还可以基于 454 测序平台进行半定量测定, 并且可以根据半定量测定

的结果进行治疗效果的评估。该文章对 44 例临床样本进行了不同方法的检测, 比较了焦磷酸测序方法与快速尿素酶实验以及碳十三呼气试验的灵敏度。幽门螺旋杆菌对克拉霉素的耐药突变主要是由于其 23S 核糖体 RNA 基因中 A2142G 和 A2143G 两个单核苷酸多态性 (SNP) 位点的突变, 通过焦磷酸测序技术可以直接检测到这两个 SNP 的突变, 该文章用实验表明了焦磷酸测序技术具备灵敏度高、检测速度快、半定量的特点, 为临床诊断提供了一种供选择的高效的方法, 且可以对临床患者对不同病原体的耐药性情况进行分子学上的诊断以及耐药情况与治疗效果的半定量检测。

现阶段, 通过高通量测序可以靶向检测耐药基因的突变, 对临床用药提供及时的指导意见也对患者的预后情况提供了预估标准^[33]。同时, 高通量测序还广泛应用于与人类胃肠道菌群情况监测与重要疾病相关微生物感染的监控, 例如肺纤维化、肺衰竭等^[35]。高通量测序可以用于监测临床长期用药之后胃肠道的菌群的实时动态变化情况, 测序平台的高通量、高输出、高效率对于实时监测呈现了巨大的临床价值。

4.3 临床肿瘤诊治中的应用

目前, 高通量测序平台同样被广泛应用在临床肿瘤学相关的研究中, 针对 DNA 或 RNA 测序相关的肿瘤细胞来源或者肿瘤基因的低频突变以及寻找新的肿瘤靶标, 发挥了巨大的临床应用价值^[36]。临床肿瘤学的检测诊断主要涉及到基于高通量测序平台的全基因组测序以及外显子测序, 其中全基因组测序在前文已经有过介绍, 外显子测序则是与肿瘤学紧密相关的测序技术, 是利用序列捕捉技术将全基因组外显子区域 DNA 序列进行特异性的捕捉并富集扩增之后进行高通量测序, 外显子基因序列在全基因组中是非常重要的部分编码序列, 用于表达体内功能性或者结构性的蛋白, 与肿瘤的发生与扩散以及肿瘤的预测与治疗有着密不可分的联系^[37], 而外显子测序相对于全基因组测序更具有靶向性, 且耗费的时间成本以及经济成本也更低, 对于检测肿瘤细胞的低频突变以及基因组的 SNP、碱基插入或缺失有着巨大的优势。在癌症的预防上, 高通量测序平台可以用于肿瘤基因的突变筛查, 指导癌症的防控工作; 在治疗上, 通过高通量测序平台挖掘与癌症相关的关联基因, 不仅可以发掘出癌症相关的诊断靶标还可以发掘出于癌症治疗相关的治疗靶点, 对于临床而言, 提供了具体的个性化的用药指导^[38]。

癌症基因组图谱 (the Cancer Genome Atlas, TCGA) 重大科研项目的创立是旨在绘制出一万个肿瘤基因组景观图谱。目前该项目已经发现了近 1 000 万个癌症相关的基因突变。科研人员主要是通过高通量测序, 对肿瘤细胞的低频、中频突变进行监测与分析, 最终获得了上千万个与癌症发生相关的基因突变, 为癌症的预测与治疗提供了十分有意义的临床价值^[36]。国际癌症基因组计划 (the International Cancer Genome Consortium) 是利用测序及高通量突变检测方法识别与癌症发生发展相关的关键基因^[39]。该大规模项目有多个国家参与, 已经发掘 50 多种不同种类癌症, 通过高通量测序,

在基因组学、表观遗传学以及转录组学方面对超过 25 000 个癌症基因组进行了系统性的研究与分析，对癌症的治疗、预后情况预测有重要的意义。

4.4 遗传疾病诊断检测

高通量测序平台另一项较为重要的应用就是遗传性疾病的检测诊断，主要包括遗传病诊断、产前筛查与诊断、试管胚胎等植入性胚胎遗传学诊断。研究人员对产妇进行无创产前基因检测，然后对基因检测异常的产妇的羊水或者脐带血细胞进行染色体 G 显带检测和 FISH 检测以作为确诊标准^[40]。高通量测序平台为遗传性疾病的诊断、为新生儿疾病的早期诊断、产妇的无创检测都提供极其快速与高效的便利。高通量测序平台的检测诊断减少了智力障碍或者残疾畸形儿童的出生率，大大减少了家庭和社會的负担，也为遗传性疾病的治疗与预防方面呈现了巨大的临床应用价值。

史淑琼^[41]等研究人员对 4709 例孕妇进行血液的采集以及核酸提取，采用新一代高通量测序技术，结合生物信息分析，得出胎儿非整倍体及性染色体的风险率，对 21、18、13 高风险及可疑性染色异常的孕妇进一步行羊膜腔穿刺术获得染色体核型。经统计学分析，无创基因 21、18、13 三体检出率均为 100%，漏诊率为 0，假阳性率分别为 0.04% (2/4708)、0.08% (4/4708) 和 0.06% (3/4708)，阳性预测值分别为 97.1% (67/69)、78.9% (15/19)、50.0% (3/6)，阴性预测值均为 100%。

无创基因检测在临床检测诊断中表现出了巨大的优势^[42]，既减少了不必要的产前有创性检测操作，增加了检测诊断的灵敏性与特异性，也大大减少了有先天遗传病的儿童的出生率，展现了巨大的临床应用价值。

5 展望

随着分子生物学和科学技术的不断提高，高通量测序技术在短短二十年内获得突飞猛进的发展。特别是“精准医疗”概念的提出，使高通量测序技术更是成为精准医疗的重要技术保障。尽管如此，高通量测序平台依然存在许多亟待改进和完善的地方：1) 二代测序平台序列读长较短，且依赖于 PCR 扩增技术，容易造成读取序列的误差与偏好性，对于后期生物信息学数据分析处理方面造成了较大的困难，为了增加测序的准确性，无疑增加了测序的时间成本以及经济成本，二代测序平台在后期需要提高序列的读长、精准性^[43]；2) 对于高通量测序平台，数据的分析处理是至关重要的步骤^[44]，目前测序市场上较为主流的 Illumina 测序平台和 Nanopore 测序平台等都有着各自个性化、独特的处理分析数据的标准流程，因此高通量测序平台对于实验人员的生物信息学基础有着十分高的要求，且由于数据处理分析方式的不同有时会产生不同的比对结果，往往没有一个“金标准”来进行校正与比较，因此后期需要进一步完善与改进数据分析处理的能力，提高平台的相对灵敏性与特异性，进一步构建相对准确与严格的标准化流程，提升平台的稳定性；3) 二代测序平台测序时间一般较长，

主要时间消耗在准备样品、构建文库以及测序分析上，相对较长的测序时间对于急性传染病的爆发的监控、临床样本的诊断鉴定尤其是高危患者的病原学鉴定是一个极其大的障碍，面对这一问题，三代测序平台的出现尽管大大缩短了测序时间，然而由于三代测序平台在精准性上比二代测序平台低，因此在广泛实际应用中依然有着障碍。总之，未来高通量测序平台还需要进一步缩短二代测序平台构建文库的时间并提高三代测序平台的准确性。

尽管高通量测序平台依然面临着十分巨大的挑战，在实际应用过程中也依然存在许多问题，然而不得不说，高通量测序在临床微生物病原学检测、感染相关病原体的诊断、肿瘤学研究、白血病等重大疾病诊断以及遗传性疾病的检测等方面都发挥了巨大的作用，也为相关的学科研究提供了新的思路与技术，高通量测序平台也将在更多的领域呈现更大的临床与研究价值。

参考文献:

- [1] CHURCH G M, GAO Y, KOSURI S. Next-generation digital information storage in DNA[J]. *Science*, 2012, 337(6102): 1628.
- [2] LI Y C, BAI W Z, HASHIKAWA T. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients[J]. *J Med Virol*, 2020, 92(6): 552-555.
- [3] HEATHER J M, CHAIN B. The sequence of sequencers: The history of sequencing DNA[J]. *Genomics*, 2016, 107(1): 1-8.
- [4] SANGER F, COULSON A R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase[J]. *J Mol Biol*, 1975, 94(3): 441-448.
- [5] MAXAM A M, GILBERT W. A new method for sequencing DNA[J]. *Proc Natl Acad Sci U S A*, 1977, 74(2): 560-4.
- [6] CHAISSON M J, HUDDLESTON J, DENNIS M Y, et al. Resolving the complexity of the human genome using single-molecule sequencing[J]. *Nature*, 2015, 517(7536): 608-611.
- [7] ZHU F Y, CHEN M X, YE N H, et al. Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants[J], 2018, 14: 69.
- [8] KILIANSKI A, HAAS J L, CORRIVEAU E J, et al. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer[J]. *Gigascience*, 2015, 4: 12.
- [9] SCHATZ M C, DELCHER A L, SALZBERG S L. Assembly of large genomes using second-generation sequencing[J]. *Genome Res*, 2010, 20(9): 1165-1173.
- [10] LU H, GIORDANO F, NING Z. Oxford nanopore MinION sequencing and genome assembly[J]. *Genomics Proteomics Bioinformatics*, 2016, 14(5): 265-279.
- [11] BOONHAM N, KREUZE J, WINTER S, et al. Methods in virus diagnostics: from ELISA to next generation sequencing[J]. *Virus Res*, 2014, 186: 20-31.
- [12] MCCOMBIE W R, MCPHERSON J D, MARDIS E R. Next-generation sequencing technologies[J]. *Cold Spring Harb Perspect Med*, 2019, 9(11): a036798.
- [13] 李继霞, 公衍文, 张月娥, 等. 用 VITEK 2 Compact 系统鉴定准确率低的临床少见病原菌评价焦磷酸

测序技术鉴定能力[J]. 生物医学工程与临床, 2016, 20(6): 626-630.

[14] MARGULIES M, EGHOLM M, ALTMAN W E, et al. Genome sequencing in microfabricated high-density picolitre reactors[J]. *Nature*, 2005, 437(7057): 376-80.

[15] DEURENBERG R H, BATHOORN E, CHLEBOWICZ M A, et al. Application of next generation sequencing in clinical microbiology and infection prevention[J]. *J Biotechnol*, 2017, 243: 16-24.

[16] LIU L, LI Y, LI S, et al. Comparison of next-generation sequencing systems[J]. *J Biomed Biotechnol*, 2012, 2012: 251364.

[17] RHOADS A, AU K F. PacBio Sequencing and its applications[J]. *Genomics Proteomics Bioinformatics*, 2015, 13(5): 278-89.

[18] BOWERS J, MITCHELL J, BEER E, et al. Virtual terminator nucleotides for next-generation DNA sequencing[J]. *Nat Methods*, 2009, 6(8): 593-5.

[19] ROY R, HOHNG S, HA T. A practical guide to single-molecule FRET[J]. *Nat Methods*, 2008, 5(6): 507-16.

[20] METZKER M L. Sequencing technologies: the next generation[J]. *Nat Rev Genet*, 2010, 11(1): 31-46.

[21] EID J, FEHR A, GRAY J, et al. Real-time DNA sequencing from single polymerase molecules[J]. *Science*, 2009, 323(5910): 133-138.

[22] LAVER T, HARRISON J, O'NEILL P A, et al. Assessing the performance of the Oxford nanopore technologies MinION[J]. *Biomol Detect Quantif*, 2015, 3: 1-8.

[23] PETERSEN L M, MARTIN I W, MOSCHETTI W E, et al. Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore sequencing[J]. *J Clin Microbiol*, 2019, 58(1): e01315-19.

[24] ILYAS M. Next-Generation Sequencing in Diagnostic Pathology[J]. *Pathobiology*, 2017, 84(6): 292-305.

[25] COUTO N, SCHUELE L, RAANGS E C, et al. Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens[J], 2018, 8(1): 13767.

[26] MARINE R L, MAGAÑA L C, CASTRO C J, et al. Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses[J]. *J Virol Methods*, 2020, 280: 113865.

[27] ZHOU P, FAN H, LAN T, et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin[J]. *Nature*, 2018, 556(7700): 255-258.

[28] WU F, ZHAO S, YU B, et al. A new coronavirus associated with human respiratory disease in China[J], 2020, 579(7798): 265-269.

[29] PETTI C A, POLAGE C R, SCHRECKENBERGER P. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods[J]. *J Clin Microbiol*, 2005, 43(12): 6123-6125.

[30] SCHLOSS P D, JENIOR M L, KOUMPOURAS C C, et al. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system[J]. *PeerJ*, 2016, 4: e1869.

[31] GU W, MILLER S, CHIU C Y. Clinical Metagenomic next-generation sequencing for pathogen detection[J]. *Annu Rev Pathol*, 2019, 14: 319-338.

[32] SALIPANTE S J, HOOGESTRAAT D R, ABBOTT A N, et al. Coinfection of *Fusobacterium nucleatum* and *Actinomyces israelii* in mastoiditis diagnosed by next-generation DNA sequencing[J]. *J Clin Microbiol*, 2014, 52(5): 1789-1792.

- [33] KUNTZ T M, GILBERT J A. Introducing the microbiome into precision medicine[J]. Trends Pharmacol Sci, 2017, 38(1): 81-91.
- [34] 初亚男, 张婕妤, 陆瑶, 等. 焦磷酸测序法在幽门螺杆菌克拉霉素耐药基因检测及药物疗效评价中的应用[J]. 国际检验医学杂志, 2020, 41(3): 262-264, 269.
- [35] FRANCINO M P. Antibiotics and the human gut microbiome: dysbioses and accumulation of resistances[J]. Front Microbiol, 2015, 6: 1543.
- [36] HOADLEY K A, YAU C, HINOUE T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer[J]. Cell, 2018, 173(2): 291-304.
- [37] ARAVANIS A M, LEE M, KLAUSNER R D. Next-generation sequencing of circulating tumor DNA for early cancer detection[J]. Cell, 2017, 168(4): 571-574.
- [38] LI M M, DATTO M, DUNCAVAGE E J, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists[J]. J Mol Diagn, 2017, 19(1): 4-23.
- [39] HUDSON T J, ANDERSON W, ARTEZ A, et al. International network of cancer genome projects[J]. Nature, 2010, 464(7291): 993-998.
- [40] 杨勉, 宋丰杰, 姚岚. 无创产前基因检测在产前筛查和诊断中的应用效果观察[J]. 当代医学, 2020, 26(5): 122-124.
- [41] 史淑琼, 宋春林, 陈淑芬, 等. 无创基因检测技术在胎儿非整倍体及性染色体产前筛查的临床研究[J]. 中国优生与遗传杂志, 2019, 27(5): 542-544.
- [42] 彭红梅, 张浔, 宋清源, 等. 无创基因检测在高龄孕妇产前筛查中的应用价值[J]. 现代诊断与治疗, 2019, 30(9): 1463-1465.
- [43] VAN DIJK E L, JASZCZYSZYN Y, THERMES C. Library preparation methods for next-generation sequencing: tone down the bias[J]. Exp Cell Res, 2014, 322(1): 12-20.
- [44] BRIS C, GOUDENEGE D, DESQUIRET-DUMAS V, et al. Bioinformatics tools and databases to assess the pathogenicity of mitochondrial DNA variants in the field of next generation sequencing[J]. Front Genet, 2018, 9: 632.

Development of high-throughput sequencing technology and its application in clinical detection

WANG Yujing, LU Zicen, CHEN Junyu, CHEN Yixin*

(National Institute of Diagnostics and Vaccine Development in Infectious Diseases, School of Public Health, Xiamen University, Xiamen 361102, China)

Abstract: Nowadays, high-throughput sequencing technology has undergone different generations, and its performance and technology are becoming more and more mature and stable. It provides an efficient analytical and detection tool for precision medicine in the fields of pathogen identification, tumor detection, disease diagnosis disease detection in clinical samples. Scientist can explore more genomic data about diseases, tumors and microbes with the high-throughput sequencing platform. The clinicians can quickly grasp the causes of the disease and microbe related to the infection of patients, which is very significant for the quick and accurate detection of disease or prediction of treatment and prognostic factors. Here we review the development of sequencing technology, the introduction of different high-throughput sequencing technology platforms and their clinical applications.

Keywords: high-throughput sequencing; next-generation sequencing; the third-generation sequencing; clinical detection

厦门大学学报（自然科学版）