



NGS

高通量测序技术 在肺癌领域的应用

罗洁◎著

Application of
High-through Put Sequencing
in Lung Cancer



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

NGS

高通量测序技术 在肺癌领域的应用

罗洁◎著

Application of
High-throughput Sequencing
in Lung Cancer



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

肺癌肿瘤组织样本较难获取与肿瘤异质性等相关信息,导致用传统的检测技术很难获得患者全面和实时的信息。近年来,高通量测序技术又称“下一代”测序(next-generation sequencing, NGS)以其快速、信息全面等优势在肺癌领域得以广泛应用。本书系统地从理论上介绍了高通量测序、全基因组测序、全外显子测序、靶向区域测序及甲基化测序的概念和发展。从临床实际病例入手,讨论了 NGS 技术在肺癌领域的应用,例如,怎样实现肺癌的精准分子分型,指导靶向治疗;如何鉴定耐药机制,调整治疗策略;如何发现罕见驱动基因及指导免疫治疗等。随着技术的不断进步,NGS 通过对肺癌患者外周血循环肿瘤 DNA 热点突变的高灵敏度检测,甚至可以发现影像学未能检出的微小病灶、耐药及休眠性克隆等。

本书可供医学专业人士参考、阅读。

图书在版编目(CIP)数据

高通量测序技术在肺癌领域的应用 / 罗洁著. — 上海: 上海交通大学出版社, 2018
ISBN 978-7-313-19407-7

I. ①高… II. ①罗… III. ①基因组—序列—测试—应用—肺癌—研究 IV. ①R734.2②Q343.1

中国版本图书馆 CIP 数据核字(2018)第 100199 号

高通量测序技术在肺癌领域的应用

著 者: 罗 洁

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 谈 毅

印 制: 上海景条印刷有限公司

开 本: 710 mm × 1000 mm 1/16

字 数: 174 千字

版 次: 2018 年 6 月第 1 版

书 号: ISBN 978-7-313-19407-7/R

定 价: 58.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 11.25

印 次: 2018 年 6 月第 1 次印刷

版权所有 侵权必究

告读者: 如发现本书有印装质量问题请与印刷厂质量科联系

联系电话: 021-59815625

肺癌是对人群健康和生命威胁最大的恶性肿瘤之一,在我国,肺癌发病率和病死率位居所有癌症首位,并有逐年上升的趋势,其5年生存率仅有16.1%左右,国内抗击肺癌的形势严峻。因此,对肺癌进行精准诊断和精准治疗成为改善肺癌预后、延长患者生存的关键。

由于肺癌的肿瘤组织样本较难获取与肿瘤异质性相关等问题,导致用传统检测技术难以获得患者全面和实时的信息。近年来,高通量测序技术又称“下一代”测序(next-generation sequencing, NGS)以其快速、信息全面等优势在肺癌的诊断、治疗和预后等方面得以广泛的应用。

本书系统地阐述了高通量测序、全基因组测序、外显子测序、靶向区域测序及甲基化测序的基本概念和应用进展,并从多个维度介绍了NGS在肺癌领域的应用:如何实现肺癌的精准分子分型,指导靶向治疗;如何鉴定耐药机制,调整治疗策略;如何发现罕见驱动基因,寻找新的治疗靶点;如何联合运用肿瘤突变负荷、微卫星不稳定性、免疫组库、肠道菌群基因组等技术确定免疫治疗的优势人群,指导免疫治疗;肺癌单细胞NGS测序的意义及进展。

本书还总结了6个经典的临床案例。从这些案例我们不难看出,NGS在当前肺癌的诊断和治疗中扮演着越来越重要的角色,能帮助临床医生探寻新的治疗靶点,真正地实现了精准诊断和精准治疗。

我们相信,随着科技的不断进步,NGS技术也将得到进一步发展。如通过对肺癌患者外周血循环肿瘤DNA热点突变的高灵敏度检测,去发现影像学未能检出的微小病灶、耐药及休眠性克隆等。

当然 NGS 作为一门新兴的技术还存在不少问题：如何在 NGS 产生的海量数据中发掘出有用的信息及其临床意义；NGS 从实验室走进临床的应用目前缺乏统一的质量控制标准；如何优化测序技术，减低成本，缩短检测时间等。

本书从理论到实践，在介绍 NGS 相关技术要点的同时，着眼于其在肺癌领域的临床应用，内容科学，观点新颖。也许最后进步的并不是 NGS 技术本身，而是凭借 NGS 这个工具，让我们对肿瘤复杂性有更深入的理解。

由于笔者能力有限，本书中可能出现纰漏或错误，恳请广大读者和同道给予批评与指正。

第一章 高通量测序的概念和发展	> 001
第一节 NGS 的发展	> 002
第二节 长读长的 NGS 测序	> 005
第三节 NGS 测序在临床和科研中的应用	> 009
小 结	> 014
参考文献	> 014
第二章 肺癌流行病学和靶向治疗	> 019
第一节 肺癌的诊断	> 020
第二节 肺癌的常见靶向治疗	> 021
第三节 其他靶向治疗	> 029
小 结	> 031
参考文献	> 032
第三章 NGS 在肺癌分子病理及其诊断 决策中的应用	> 035
第一节 NGS 对肺癌分子病理学及诊断的检测分析	> 036
第二节 临床应用对 NGS 检测的要求	> 040
参考文献	> 043
第四章 NGS 研究肺癌异质性和肿瘤进化	> 045
第一节 肿瘤异质性和肿瘤进化	> 046

第二节 研究肺癌异质性及进化的意义	> 051
参考文献	> 053
第五章 NGS 发现罕见突变基因	> 055
第一节 肺腺癌罕见突变基因	> 056
第二节 肺鳞癌罕见突变基因	> 059
第三节 小细胞癌罕见突变基因	> 060
第四节 香烟暴露与肺癌基因组改变	> 061
参考文献	> 062
第六章 NGS 发现肺癌基因表达异常	> 065
第一节 肺鳞癌基因表达异常	> 066
第二节 肺腺癌基因表达异常	> 069
第三节 小细胞肺癌基因表达异常	> 069
参考文献	> 070
第七章 NGS 指导治疗决策和鉴定肿瘤耐药机制	> 073
第一节 肺癌基因组学研究的主要内容及其临床意义	> 075
第二节 NGS 鉴定肺癌耐药机制	> 078
第三节 NGS 促进肺癌的研究和治疗	> 087
附：NCCN 关于 NSCLC 指导用药的参考	> 091
参考文献	> 094
第八章 NGS 指导肺癌免疫治疗	> 099
第一节 免疫治疗概述	> 100
第二节 过继性 T 细胞治疗和 CAR-T	> 101
第三节 使用 NGS 助力肿瘤反应性 T 细胞鉴定肿瘤 特异性抗原	> 104

第四节 NGS 指导免疫治疗	> 109
小 结	> 111
参考文献	> 112
第九章 NGS 和 cfDNA 辅助肺癌诊断和治疗	> 115
第一节 cfDNA 的基本概念和特性	> 116
第二节 cfDNA 的 NGS 测序方法	> 118
第三节 cfDNA 突变检测技术	> 123
第四节 cfDNA 在肺癌领域的最新进展	> 124
参考文献	> 129
第十章 肺癌单细胞 NGS 测序的意义和进展	> 133
第一节 单细胞测序方法	> 134
第二节 单细胞测序的临床研究	> 139
小 结	> 140
参考文献	> 140
第十一章 人体微生物测序在肺癌诊疗中的 研究和发展	> 143
第一节 NGS 和微生物研究	> 144
第二节 肠道微生物、口腔微生物和肺部微生物	> 146
第三节 微生物对肺癌发生和治疗的影响	> 148
参考文献	> 149
附录 经典病例	> 151
关键词索引	> 169

第一章

高通量测序的概念和发展

测序技术诞生于 20 世纪 50 年代,经过 30 年的发展又诞生了第二代测序技术——高通量测序,又称“下一代”测序(next-generation sequencing, NGS)。与一代测序相比,NGS 具有通量大、精准度高和信息量丰富等优点,可以在短时间内对感兴趣的基因进行精准定位,也可以对未知的序列进行检测。目前,NGS 在无创产前筛查、肿瘤基因检测、遗传性疾病诊断和用药指导等多个临床领域得到广泛应用,极大地推动了肿瘤精准医学的发展。

第一节 NGS 的发展

1. DNA 和 RNA 的概念

核酸包括脱氧核糖核酸(deoxyribonucleic acid, DNA)和核糖核酸(ribonucleic acid, RNA),是生命信息的载体,主要功能是信息储存,其中包含生命生长发育所需的全部信息。DNA 是储存、复制和传递遗传信息的主要物质基础。RNA 在蛋白质合成过程中起重要作用——信使核糖核酸(messenger RNA, mRNA)是合成蛋白质的模板;转运核糖核酸(transfer RNA, tRNA)起携带和转移活化氨基酸的作用;核糖体核糖核酸(ribosomal RNA, rRNA)是蛋白质合成的主要场所。

2. 测序的发展

由于 DNA 是生命信息的蓝图,解码 DNA 信息就可以加深我们对生命过程的理解。从 DNA 的双螺旋结构被解析开始,人们就一直在努力探究健康与疾病的基因奥秘。DNA 测序作为一种重要的实验技术,广泛应用于生物研究领域。第一代 DNA 测序技术包括 1975 年由 Sanger 和 Coulson 开创的链终止法

(也称 Sanger 法)以及 1976 年—1977 年由 Maxam 和 Gilbert 发明的化学法(链降解法)。1977 年, Sanger 首次测定了一个噬菌体 X174 的基因组序列, 全长 5 375 个碱基^[1]。自此, 生命科学研究步入了基因组学时代。2001 年, 研究人员在改进的 Sanger 法基础上完成了首个人类基因组图谱。

Sanger 法巧妙地利用了 DNA 合成的原理。在 DNA 的复制过程中需要: DNA 聚合酶、DNA 模板、引物和 4 种脱氧核糖核苷三磷酸(deoxyribonucleoside triphosphate, dNTP)[脱氧腺苷三磷酸(deoxyadenosine triphosphate, dATP)、脱氧鸟苷三磷酸(deoxyguanosine triphosphate, dGTP)、脱氧胸苷三磷酸(deoxythymidine triphosphate, dTTP)和脱氧胞苷三磷酸(deoxycytidine triphosphate, dCTP)]。聚合酶根据碱基互补配对原则将 dNTP 加到引物的 3'-OH 末端, 使引物延伸, 合成出新的互补 DNA 链。如果加入双脱氧核苷三磷酸(dideoxyribonucleoside triphosphate, ddNTP), 由于它在脱氧核糖的 3' 位置缺少一个羟基, 不能与后续的 dNTP 形成磷酸二酯键, 会导致延伸终止。例如, 在存在三磷酸双脱氧胞嘧啶核苷(ddCTP)、dCTP 和 3 种其他 dNTP(其中一种为 α -³²P 标记)的情况下, 将引物、模板和 DNA 聚合酶一起延伸, 即可形成一种长短不一、具有相同的 5'-引物端并以 ddC 残基为 3' 端结尾的片段混合物。然后利用变性聚丙烯酰胺凝胶电泳, 可以将各个片段按其链长的不同进行条带分离, 最后获得相应的放射性自显影图谱, 即可从所得图谱直接读取 DNA 的碱基序列了。这个方法是先合成再测序, 准确率高, 但费用也高。

Sanger 法作为测序的“黄金标准”, 为生物医学的发展做出了突出贡献。随着技术的发展, 特别是人类基因组计划的推荐, 二代测序逐渐进入人们的视野。

3. 二代测序的基本概念和平台发展

随着人类基因组计划的进行和完成, 人们认识到通过测序技术与数据分析可以解答诸多的生物学问题。然而, 测序通量的限制以及高昂的成本阻碍了人们对生命活动和疾病的深入了解^[2]。2000 年之后推出的高通量测序平台很好地解决了这些问题, 人类基因组的测序成本直接下降了 50 000 倍, 并且由此产生了一个新名词: 二代测序(也称下一代测序, NGS)。在过去的十年中, NGS 得到不断发展——测序的数据量较前增加了 100~1 000 倍^[3]。技术方面, 研究

人员甚至可以在一条读长(read)上读出整条基因组的序列。Veritas Genomics的数据显示,人类基因组的测序成本已下降到1 000美元/人。目前,NGS技术也被广泛应用到临床诊断方面。

NGS是通过高通量技术来实现大规模测序,所以也叫作高通量测序(high-throughput sequencing, HTS),主要包括如下几种方法:①边合成边测序(sequencing by synthesis, SBS),代表性的有Roche公司的454焦磷酸测序、Illumina公司的Solexa合成测序;②边连接边测序(sequencing by ligation, SBL),如ABI公司的SOLiD连接法测序。

454焦磷酸测序的原理是借助生物发光来对DNA序列进行检测。在DNA聚合酶、ATP硫酸化酶、荧光素酶和双磷酸酶的协同作用下,系统将引物上每一个dNTP的聚合都与一次荧光信号相偶联。通过检测荧光信号释放的有无和强度,就可以实时测定DNA序列了。此技术不需要荧光标记的引物或核酸探针,也不需要电泳,具有分析快速、结果准确、高灵敏度和高自动化的特点。

SBL方法则是使用带有荧光基团的探针与DNA片段杂交,再连接邻近的寡核糖核酸,从而成像,然后通过荧光基团的发射波长来判断碱基或者互补碱基的序列。该方法包括5轮测序反应,每轮又含有多次连接反应(一般情况下,片段文库是7次,末端配对文库是5次。所以片段文库共有35次连接反应,而末端配对文库共有25次连接反应),第一次连接反应由与P1引物区域互补的“连接引物”介导。

绝大多数的SBL和SBS方法,都是在一个固定的表面进行DNA扩增,使特定区域内有成千上万个DNA片段拷贝,可以确保将方法信号与背景信号区别开来。同时,大量的平行碱基信息有助于大批量reads的读取。一个测序平台通常可以达到百万级的数据读取量,也就是能对上百万DNA分子进行同时测序。与一代测序相比,NGS具有通量大、精确度高和信息量丰富等优点,可以对目标基因进行迅速精准定位,也可以用来检测未知序列,还可以对组织在特定时间表达的mRNA进行测序^[4]。

不过NGS技术也有其不足之处,它虽然在数据量上有了大幅提升,但质量却有待提高。有报道称,NGS在序列拼接过程中的错误率为0.1%~1%。而且NGS生成的read普遍较短,每条read的长度在35~700bp之间^[5],比普通的

Sanger 法测序都要短,这意味着需要更严格且复杂的序列拼接。

第二节 长读长的 NGS 测序

1. 简介

基因组是一个复杂的复合物,其中包含了多种重复序列、拷贝数变化和结构变异。这些与进化、适应以及疾病密切相关^[6-8]。许多复合物元件过长,用短读长测序无法进行完整读取。而长读长测序的 read 可以达到数千个碱基,因此能对大的结构进行解析。长读长测序产生的单一长序列甚至可以跨越不同复合物或者重复序列。长读长测序在转录组的测序中也大有优势,因为长读长的 read 可以不需要拼接就跨越完整的 mRNA 转录本,能鉴定出更多的基因亚型。

长久以来,短读长测序都很难解决长链 DNA 中重复序列以及复合序列难以拼接的问题^[9-11]。而 Chaisson 等应用长读长测序在 GRC 人类基因组数据库中提交了超过 1 MB 的新序列^[12],这些长序列甚至弥补了曾经缺失的信息。不仅如此,Chaisson 等还鉴定了至少 26 000 个超过 50 bp 大小的插入缺失(insertion-deletion, InDel),GRC 人类基因组数据库也因此成为最具参考价值的基因组数据库之一。此外,长读长还能够为临床诊断提供更有依据^[13-15]。

近来,研究人员开发出了两种长读长测序技术,分别是单分子实时测序法(single molecule real time sequencing, SMRT)和利用短读长技术在体外构建长读长的合成法。单分子测序又称为第三代测序技术,与短读长测序完全不同,可以不经过聚合酶链反应(polymerase chain reaction, PCR)扩增就能对每一条 DNA 分子做单独测序,测序过程中也不需要轮番添加 dNTP。代表技术主要为单分子荧光技术,通过识别发光的单个核酸分子来进行实时读取。合成法则利用条形码拼接来获得长片段,有别于常规长读长测序产生的原始 read。

目前最常用的长读长测序是 PacBio Biosciences(PacBio)的单分子实时测序法^[16,17]。短读长 SBS 技术需要用聚合酶结合 DNA,并沿 DNA 链进行扩增。PacBio 采用了类似 SBS 的方法,不同的是先使用聚合酶捕获模板 DNA 并锚定在零模波导孔底部,然后让不同荧光标记的 dNTP 随机进入零模波导孔底部。

荧光 dNTP 与 DNA 在酶作用下配对合成新的碱基,这时可根据荧光信号的颜色和存在时长来区分游离碱基与配对碱基,从而获得 DNA 序列。

2014 年,第一台纳米孔(nanopore)测序仪——MinION 诞生。与其他平台不同,Nanopore 测序仪并不是检测与模板 DNA 结合或杂交的核糖核酸,而是直接读取天然的单链 DNA 分子。该过程是让 DNA 通过一个特殊的纳米蛋白孔,此时会产生特定的电压改变,由于不同碱基发出的电流信号强度有差异,由此可对 DNA 序列进行读取。该方法最突出的优点就是降低了测序成本,几乎没有试剂的耗费,也不像其他测序方式那样需要用到核苷酸、聚合酶等。并且由于不需要克隆、扩增的步骤,也极大地节省了时间。还有一点就是纳米孔测序能测得更长的 read,降低了基因组组装的不确定性。但它也有诸多技术问题亟待解决。首先就是纳米孔的结构问题,理想情况应该是孔径仅容 1 个碱基通过,这样才能获得最大的信号区分度,但目前的技术水平尚无法做到这一点。另外就是 DNA 分子的穿孔速度过快导致了信号分辨率不高,这也是目前纳米孔测序准确率较低的原因之一。因此,虽然纳米孔测序相比二代测序有很多优势,但还不能完全取而代之。

合成法主要可通过两个系统来实施: Illumina 长片段合成系统与 10X Genomics 乳液系统。Illumina 系统不需要借助特殊仪器就能将 DNA 分隔到微孔板上。而 10X Genomics 乳液系统则需先使用微流体平台来进行测序前的准备工作,然后再用乳液分隔 DNA。在 DNA 浓度低至 1 ng 的情况下,10X Genomics 乳液系统仍能将 DNA 分子切割成任意长度的片段(最大达 100 KB)。

2. PacBio 测序技术

现在最常用到的长读长测序设备是 PacBio RS II。该设备可以生成超过 50 KB 长度的单个 read,长链建库的测序平均长度为 10~15 KB。这对基因组拼接以及基因组结构的大范围应用都很有帮助^[18,19]。不过其长链测序中单个碱基的错误率在 15% 左右^[20],又让人们对该仪器的使用有所顾虑^[21]。这些错误随机分布于每个 read,必须通过足够高的覆盖度才能消除错误率的影响^[22]。只要单个碱基的测序次数增多了,所得结果还是比较可靠的,实际上该方法的最高准确率可达到 99.999%^[23]。这点与 Sanger 法测序相似,因此该技术与

Sanger 法一样都被视为研究单核苷酸多态性(single nucleotide polymorphism, SNP)的方法^[20]。PacBio RS II 的运行时间与通量会受到测序读长的影响,长的模板需要更长的时间。举例来说,1 KB 的库运行 4 小时,每个分子可以产生约 30 000 个碱基,期间可重复测序 30 次左右。而 10 KB 的库运行 4 小时,同样产生 30 000 个碱基的情况下却只能重复测序 3 次左右。高成本(1 000 美元/GB)、低通量,加上需要较高的覆盖度,使得 PacBio RS II 在一些较小的实验室难以开展应用。

不过 PacBio 新推出的 Sequel 系统,通量比 RS II 高出了 7 倍,测序成本反而下降了一半。二代测序存在的读长较短的问题在 Sequel 上也得到了改善,其读长一般在 9 KB 以上,准确率超过 85%。虽然 Sequel 的通量已有大幅提升,但仍较二代测序要低得多,一个细胞才产出 5 GB 左右,用于临床检测的话成本还是太高。

现阶段 Pacbio 的产品主要应用在两个方向:一个是基因组的组装;另一个是全长转录组测序。

3. MinION 测序技术

MinION 纳米孔测序仪的主件是 2 048 个纳米孔,分成 512 组,由专用集成电路控制。测序原理:首先,待测 DNA 分子连接上引导接头(lead adaptor)、发夹接头(hairpin adaptor)和拖尾接头(trailing adaptor);由引导接头引领进入纳米孔,其后按照待测 DNA、发夹接头、待测 DNA 互补链、拖尾接头的顺序依次通过。经双序列比对后,待测 DNA 与其互补链可组合成 2D read。另外一种方法是不使用发夹接头,只测序待测 DNA,这样形成的是 1D read。1D 测序方法通量更高,但是准确性要低于 2D read。

ONT MinION 是一个小型的 USB 设备(3 cm×10 cm),可以在个人电脑上运行,这也是目前最小的测序平台。尽管一些相应的设备必须要有固定场所来安置,如做文库准备的恒温器,但本身小体积仍使其操作极具便利性。不过 MinION 对测序片段的大小有一定限制。理论上,该设备能测序任意大小的 DNA 分子,但实际上,对长片段进行测序时还是会出现一定的错误率^[24]。同样,如何有效地对核糖核酸复合物进行测序也是 ONT MinION 面临的一大

问题。由于通过纳米孔时的电流信号存在时间很短,加上一些修饰的碱基也会改变原先设定的电压变化,当核糖核酸复合物较长时,也没法准确鉴定其通过纳米孔的顺序。但最近一系列对试剂和算法的改进使其准确率提高了不少^[25]。

1) MinION 相对于其他 NGS 测序平台的优势

(1) 碱基修饰的检测。纳米孔测序技术可以检测 4 种胞嘧啶的碱基修饰,分别为 5-甲基胞嘧啶、5-羟甲基胞嘧啶、5-甲酰胞嘧啶和 5-羧基胞嘧啶,检测准确率为 92%~98%。

(2) 实时测序监控。临床实践中,传统的 NGS 测序要做到实时获取和分析 DNA/RNA 序列,是一件不容易的事情,但运用 MinION 则相对容易办到。首先是 MinION 体积小,方便携带;其次是当待测 DNA 分子穿过纳米孔时,其电流变化马上被检测并识别,能迅速给出结果。在不考虑错误率的情况下,可以说真正做到了实时监控,这对于一些特定目标的测序有着重要的作用。另外,还可利用这种技术特点实现目标序列的富集:当 DNA 片段通过纳米孔时,如电流变化与目标序列一致,则能通过;如呈现不同的电流变化,则片段不能通过纳米孔。这样能显著地减少测序时间,有利于野外操作和即时诊疗。

(3) 测得更长的 read。MinION 测序仪可以生成 300 KB 长的 1D read,以及 60 KB 长的 2D read。这种长 read,甚至帮助研究人员完善了人类基因组 Xq24 号染色体上一个长 50 KB 的间隔。

(4) 结构变异的检测。由于只能检测短序列的缘故,NGS 对结构变异的检测往往不准确,这个问题在肿瘤检测中尤为突出,因为肿瘤组织中充斥着各种结构变异。研究发现 MinION 通过几百个长 reads 测得的结构变异结果甚至比 NGS 所测的上百万 reads 的结果要更可靠。

(5) RNA 表达分析。NGS 平台测得的短序列都需要进行序列拼接才能得到转录本,通常情况下由于缺乏足够信息而无法区分形式各异的可变剪切。而 MinION 测序仪产生的长 read,可以更好地解决这个问题。以果蝇的唐氏综合征细胞黏附分子 1(down syndrome cell adhesion molecule 1, *Dscam1*)基因为例,其存在 18 612 种可变剪切形式,MinION 测序仪可以检测到其中 7 000 种以上,这样的结果是 NGS 短序列测序不可能获得的。

2) MinION 目前的应用领域

Nanopore 测序仪的具体功能定位仍在探索当中,像快速文库制备、实时数据生产以及小体积等优势都有望转变为实际价值。

(1) 即时检测传染源。纳米孔测序方法与 NGS 方法都可以用于院内传染源病菌的检测,而纳米孔技术在测序读长、便携性、检测时长等方面更具优势。文献记载 MinION 测序从样品准备到发现致病菌只要 6 小时,其中从样品放入机器到发现致病菌仅需 4 分钟。有英国的研究人员就将 MinION 用于监测沙门氏菌的爆发^[26]。在 2014 年的埃博拉病毒爆发事件中,MinION 测序仪也有出色的表现^[27]。

(2) 非整倍体检测。MinION 在胎儿非整倍体产前检测中也发挥了重要作用。此检测在 NGS 平台通常需要 1~3 周时间才能获得结果,而文献报道使用 MinION 测序只需要 4 小时。

第三节 NGS 测序在临床和科研中的应用

NGS 广泛应用于基因表达和调控研究。像蛋白-DNA 相互作用就可以通过染色质免疫共沉淀结合 NGS 测序来进行鉴定^[28]。NGS 还可用于修饰碱基的研究。例如,最初的甲基化测序虽然可以实现甲基化 DNA 的捕获与富集^[29],以及可以在酶作用下选择性区分甲基化与非甲基化区段^[30-32],但其修饰与捕获的过程不够理想。针对这个问题,Flusberg 等在 2010 年发表了一个概念性的研究方法,使用 PacBio 来区分甲基化与非甲基化的碱基^[33]。由于聚合酶在甲基化位点上会停留更多的时间,因此可以通过碱基上的信号改变来分辨是否存在甲基化修饰。同样,Nanopore 平台也能够监测到修饰的碱基,因为甲基化同样会引起纳米孔电压的变化。由此甲基化测序可以在不需要化学操作的条件下进行^[34]。

接下来主要介绍 NGS 在临床和科研中的应用。

1. 全基因组重测序

全基因组重测序(whole-genome resequencing, WGR)是对已知基因组序