新一代测序技术的研究进展*

韦贵将¹ 邹秉杰^{2,3} 陈之遥² 宋沁馨^{2,4} 李佑志¹ 周国华^{1,2,3△} (1中国药科大学生命科学与技术学院 江苏南京 210009;

2 南京大学医学院临床学院(南京军区南京总医院)药理科 江苏 南京 210002;

3 华东医学生物技术研究所 江苏 南京 210002 ;4 中国药科大学药学院 江苏 南京 210009)

摘要 :大规模 DNA 测序技术是揭秘人类和其它生物遗传密码的重要技术,在分子生物学和基础医学领域有广泛应用。第二代测 序技术的出现使 DNA 测序的通量大幅提高,测序的成本大幅下降,原来只有在大型测序中心才能完成的测序任务现在已经可以 在更多的实验室展开。但是,早期的第二代测序技术仍然存在诸如文库构建过程复杂、测序成本依然较高等缺点。为了克服上述 缺点,近三年发展了几种新的第二代和第三代测序技术,这些技术不仅继承了早期第二代测序技术通量高的优点,而且在文库构 建等方面取得了重要突破,进一步简化了测序操作,降低了测序成本,缩短了测序时间。本文就几种最新的大规模测序技术的原 理、特点与发展趋势进行简要介绍。

关键词:大规模 DNA 测序:单分子测序;第二代测序技术;第三代测序技术 中图分类号:Q331_R394-3, R446.1 文献标识码:A 文章编号:1673-6273(2012)19-3789-05

Advances in Next-generation Sequencing Technologies*

WEI Gui-jiang¹, ZOU Bing-jie²³, CHEN Zhi-yao², SONG Qin-xin²⁴, LI You-zhi¹, ZHOU Guo-hua^{1,23Δ}
(1 School of Life Science and Technology, China Pharmaceutical University, Nanjing 210009, Jiangsu, China;
2 Department of pharmacology, Jinling Hospital, School of Medicine, Nanjing University, Nanjing Jiangsu, 210002, China;
3 Huadong Research Institute for Medicine and Biotechnics, Nanjing Jiangsu, 210002, China;

4 School of Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu, 210009, China)

ABSTRACT: Large-scale DNA sequencing technology is a very important tool for us to know genetic information of human and other species, and is widely used in the fields of molecular biology and basic medicine. Because second-generation sequencing is high-throughput and low cost, the sequencing task, which only could be completed in the large-scale sequencing centers in the past, can be accomplished in many common laboratories now by the second-generation sequencing technologies. However, some drawbacks still exist in second-generation sequencing technologies such as laborious process of library preparation and still high sequencing cost. To overcome these drawbacks, a new second-generation sequencing technologies not only inherited the advantages of high-throughput of the early second-generation sequencing technologies, but also improved the process of library preparation, simplified the sequencing operations, reduced the sequencing cost and shortened the sequencing time. In this review, we briefly introduced the principles, features and prospects of the latest large-scale DNA sequencing technologies.

Key words: Large-scale DNA sequencing; Single-molecule sequencing; Second-generation sequencing technologies; Third-generation sequencing technologies

Chinese Library Classification (CLC): Q331, R394-3, R446.1 Document code: A Article ID:1673-6273(2012)19-3789-05

前言

在人类基因组完整图谱绘制完成后,人类进入了后基因组时代,基因组学、转录组学等新兴研究领域相继涌现,它们的研究成果大大加深了人类对生命遗传机制和疾病发生等复杂过程的认识。但是这些研究往往要对物种的全基因组或者所有的mRNA进行测序,因而需要简便快捷和成本低廉的大规模测序方法。虽然 Sanger 测序技术在测序准确性和读长方面已经相

当出众,但其测序通量与速度低、测序成本高等缺点导致它无 法满足大规模测序任务的要求,这促使新一代测序技术的诞 生。自2005年以来,454 测序技术^[1]、Solexa 测序技术^[2]和 SOLiD 测序技术^[3]等第二代测序技术相继出现,它们相对第一 代测序技术在测序通量和测序成本等方面取得了质的飞跃,并 已经在基因组从头测序和转录组测序等领域得到了广泛的应 用^[46] 极大地推动了生命科学的发展。但是第二代测序技术仍 然存在诸多亟待解决的问题,如前期文库构建过程复杂、测序

```
*基金项目 国家转基因生物新品种培育科技重大专项(2011ZX08012-005)
```

作者简介 :韦贵将(1985-) 男 ,硕士研究生 ,主要研究方向 分子诊断学 ,Tel:13451890505 E-mail:cpuwgj@163.com △通讯作者 周国华 E-mail:ghzhou@nju.edu.cn (收稿日期 2012-04-06 接受日期 2012-04-30) 反应体系为多酶体系难以控制、需要荧光标记、"反应-拍照-洗脱"的步骤影响测序速度和测序读长等。针对这些问题 最 近发展了几种新一代的高通量测序技术,它们通过改变测序原 理或检测手段,进一步简化了测序过程,降低了测序成本,提高 了测序速度,推动了大规模测序的广泛应用。本文就这些新一 代测序技术的原理、特点和应用前景作简要介绍。

1 Ion torrent 测序技术

Ion torrent 测序技术摒弃了 454 测序技术中采用生物发光 检测延伸产生的焦磷酸的检测原理,通过检测 DNA 链延伸时 产生的氢离子实现边合成边测序,其原理如图 1 所示。首先通 过微乳液 PCR 制备测序模板,然后将制备得到的微球模板置 于一种半导体芯片上,加入含有 DNA 聚合酶的测序反应液,并 循环加入 4 种脱氧核糖核苷酸(dNTP),通过高灵敏的 pH 敏感 电极测定与模板互补的 dNTP 被掺入时释放的 H⁺,从而实现对 模板序列的测定^[7,8]。

Ion torrent 测序反应无需多酶体系,反应过程简单,也无需 特殊修饰 dNTP 和昂贵的光学设备,极大地降低了测序成本、 提高了测序速度。此外,由于 Ion torrent 测序技术利用了半导 体芯片,使其能达到极高的测序通量,并且芯片制备工艺更加 成熟简单。从第一张测序芯片 314Chip 到目前的 318Chip Jon torrent 测序芯片的通量在短短一年内由 10 Mb 提高到 1 Gb。 虽然目前它的通量还远比不上其它二代测序技术,但以半导体 芯片上集成密度预测,未来其通量将能达到现有的 SOLiD 测 序技术通量。在 2012 年 1 月的第 30 届摩根大通保健大会上, Life Technologies 公司宣布,他们将于年内推出一款新的测序 仪 --Ion Proton 测序仪。这款仪器旨在一天内完成人类基因组 的测序 耗资约为 1000 美元。

虽然 Ion torrent 测序技术能够实现快速、低成本基因组测 序,并已应用在基因组重测序等领域^[9,10],但是它依然存在一些 问题。首先是读序长度需要进一步增长,目前 Ion torrent 测序 的读序长度在 200 nt,这对序列拼接无疑造成了负担,但据 Life Technologies 公司发布的消息称,在 2012 年下半年,新一代的 测序试剂盒的读序长度将提升到单向读序 400 nt;其次,测序 准确度需要进一步提高, Ion torrent 测序的主要错误类型为同 质区碱基的插入与缺失,需要提高同质区测序的准确性,使 Ion torrent 测序技术更加精准;此外, Ion torrent 测序技术依然需要 对待测模板进行微乳液 PCR 扩增,操作繁琐,限制了测序速 度。



图 1 Ion torrent 测序技术的原理^[7] Fig.1 The principle of Ion Torrent sequencing technology^[7]

2 HeliScope 测序技术

为了简化测序模板的制备过程,发展起来了无需扩增待测模板的单分子测序技术。HeliScope 测序技术是单分子测序技术的一种,该技术无需 PCR 扩增测序模板,可直接对单个DNA分子进行序列测定,其原理如图2所示。基因组DNA经过剪切、多聚腺苷酸加尾并固定到测序芯片表面后A、C、G、T四种荧光修饰的dNTP 依次循环流加至芯片上,当碱基互补延伸后,采集荧光信号获得碱基信息,之后酶切去除荧光基团,再进行下一轮反应。数十个循环后,将测得的DNA序列拼接,得到完整的基因组序列。由于在检测单个荧光分子时,激发光源很容易激发周边的荧光物质,形成背景光源的干扰。对此,HeliScope 测序技术采用全内反射荧光显微镜,测序时激发光源只激发芯片表面上很薄一层空间内的延伸至延伸链末端的荧光基团,从而降低了背景光源的影响^[11,12]。

HeliScope 测序技术的测序通量可以达到 35 Gb,平均测序

读长为 35 nt,在测序深度为 20 倍覆盖率的情况下,测序准确 率为 99.995 %。并且由于无需 PCR 扩增待测模板,避免了 PCR 歧视性扩增的影响,简化了文库制备过程。目前 HeliScope 测序 技术已有所应用^[13,14],但是由于单分子测序设备价格十分昂贵, 使其推广应用存在一定困难。

3 SMRT 测序技术

SMRT 测序技术是另一种单分子测序技术,其原理如图 3 所示。预先将 DNA 聚合酶固定在测序芯片的反应孔底部,当加 入的 DNA 测序单链进入反应孔被 DNA 聚合酶捕获后,四种 不同荧光标记的 dNTP 加在反应孔的上端,只有进入反应孔底 部很薄一层空间的荧光标记 dNTP 才能被激发光激发荧光,由 于布朗运动,荧光标记的 dNTP 偶尔进入反应孔底部的荧光激 发区域,但由于停留的时间很短,不能被检测。当加入的 dNTP 与模板互补延伸时,DNA 聚合酶首先捕获与模板匹配的 dNTP,然后催化生成磷酸二酯键并切除修饰在磷酸基团上的 荧光基团,这一过程使荧光标记 dNTP 的停留时间显著增长, 能够被激发从而检测到相应的荧光信号,根据标记荧光基团发 射波长的不同,可以识别延伸的碱基种类^[15,16]。

SMRT 测序技术的单分子荧光检测设备采用零模式波导 技术,可以将激发光局限在反应孔底部很薄的空间内,价格低 于全内反射显微镜, 使测序成本比 HeliScope 测序技术有所降 低。由于荧光基团修饰在脱氧核糖核苷酸磷酸基团上,在荧光 基团切除后自动开始下一个碱基的测定,所以 SMRT 测序技术 的测序速度可以达到 1.5 nt/秒,且测序读长可以达到 3000 nt, 比 HeliScope 测序技术有明显提高。但是 SMRT 测序技术目前 仍处于试用阶段,仅有少数文献对其应用进行了报道^[17,18]。



Fig.2 The principle of HeliScope sequencing technology^[11]



图 3 SMRT 测序技术的原理^[15] Fig.3 The principle of SMRT sequencing technology ^[15]

4 Oxford 纳米孔测序技术

Oxford 纳米孔测序技术也是一种单分子测序技术。它利用 镶嵌于脂质双分子层中的经过基因工程改造过的 α- 溶血素蛋 白作为纳米孔道,并在 α- 溶血素蛋白孔道上部和中部分别连 接一个核酸外切酶和一个氨基化环糊精配体(如图 4 所示)。在 测序反应开始时,核酸外切酶逐个切割单链 DNA 分子上的脱 氧核糖核苷酸分子,切割后的脱氧核糖核苷酸分子进入纳米 孔 经过氨基化环糊精配体时,将会引起电流的变化。由于四种 脱氧核糖核苷酸与甲基化胞嘧啶脱氧核糖核苷酸引起电流变 化幅度不同,从而可以区分不同的碱基,进而测定 DNA 链的序 列^[19,20]。

相对其它测序技术, Oxford 纳米孔测序技术的样本处理极

其简单,并且无需 DNA 聚合酶或者连接酶,也无需 dNTPs,因 此其测序成本十分低廉,比其他测序技术更有可能实现 1000 美元基因组目标。由于无需 "反应 - 拍照 - 洗脱 " 的测序过程, Oxford 纳米孔测序技术的测序速度可以达到 20-400 nt/s,将来 有可能实现 15 分钟完成人类基因组测序。然而,目前的 Oxford 纳米孔测序技术还存在一些缺限,其中最大的问题是核酸外切 酶切割的脱氧核糖核苷酸通过纳米孔的速度过快,来不及检 测,所以经常造成测序结果发生碱基缺失。虽然有人试图使用 一些方法如降低测序溶液温度和增加测序溶液黏度等来解决 这个问题,但是这些方案尚不能把脱氧核糖核苷酸通过纳米孔 的速度控制到一个合理数值,因此,纳米孔测序的准确度依然 不尽如人意,目前其测序错误率在4%左右。2012 年 Oxford Nanopore Technologies 公司宣布了两款基于 Oxford 纳米孔测 序技术的测序平台,但还没有利用它完成全基因组测序等相关 应用的报道,而仅用来完成过合成 DNA 模板测序,因此,该技 术实际应用的性能还不得而知。



图 4 Oxford 纳米孔测序技术的纳米孔结构^[21] Fig.4 The structure of Oxford nanopore^[21]

5 其他新一代测序技术

· 3792 ·

2011 年谢晓亮等发明了一种在 PDMS 制造的测序芯片上 进行荧光焦磷酸反应的测序方法,该方法检测的是反应中产生 的荧光修饰焦磷酸基团,不仅提高了检测的灵敏度而且降低了 大规模平行测序的难度^[22]。2008 年 VisiGen 公司开发了一种基 于荧光共振能量转移原理的单分子测序技术。该技术将连接了 能量供体的 DNA 聚合酶固定在测序芯片的表面,当荧光分子 标记的 dNTP 与模板互补延伸时,能量供体通过荧光共振能量 转移原理激发附近的荧光分子,再利用全内反射荧光显微镜检 测延伸信号。由于荧光分子标记在 dNTP 磷酸基团上,在延伸 后被切除掉,因而测序过程中无需停顿,并且实现了实时检测 ^[23]。2008 年 Henk 等设想在石墨烯表面制造纳米孔,当单链 DNA 通过纳米孔时,不同的碱基引起不同的电流变化幅度,从 而实现利用固态纳米孔对 DNA 序列的测定^[24]。预计未来推出 的其它新一代测序技术将以物理检测手段直接测定核酸序列 为主^[25,26]。

6 总结与展望

表 1 四种新一代测序技术的特点 Table 1 The characteristics of the four next-generation sequencing technologies						
Ion torrent 测序技术	The PGM Sequencer	1Gb	200 bases	99%	测序速度快 成本 低 原始数据准确 率高	有效反应孔数不足 故通量低 :文库需 要扩增
HeliScope 测序技术	HeliScope Single Molecule Sequencer	35Gb	35 bases	95%	文库无需扩增 试 剂消耗少	读长短 通量低 原 始数据准确率相对 低 20器价格昂贵
SMRT 测序技术	The PacBio RS system	10Mb	3000 bases	84%	读长长 ;文库无需 扩增 ,测序速度快	原始数据准确率 低 <i>:</i> 有效反应孔不 足故通量低
Oxford 纳米孔测序技术	The GridION system	未知	未知	96%	读长长 ;文库无需 扩增 测序速度快 ;	芯片上纳米孔数目 少 <i>·</i> 碱基通过纳米

新一代测序技术的发展均以提高测序通量与读序长度、降 低测序成本、简化测序步骤为目标,但以上各种新一代测序技 术都存在各自优势与不足(如表1所示)。Ion torrent 测序技术 测序速度快、试剂和仪器成本低,但是前期文库制备过程繁琐、 测序通量和测序读长不足,未来增加有效反应孔百分比和测序 读长是其改进的方向。HeliScope 测序技术无需繁琐的文库制 备过程和使用反应试剂少,但是读长短、通量低且仪器价格昂 贵,增加测序读长和测序通量也是其发展方向。SMRT 测序技 术和 Oxford 纳米孔测序技术读长长和测序速度快,但是它们 都面临着有效反应孔或者纳米孔数目不足的挑战,另外 SMRT 测序技术还有测序原始数据准确率不高的问题,解决这些问题 是它们未来的首要任务。 新一代测序技术仍在不断开发中,技术的进步推动测序成 本不断降低和测序时间不断缩短,相信不久千元基因组甚至百 元基因组的目标将会实现。届时,人们对生命的发育过程和疾 病的发生分子机制等重大问题将会有更深刻细致的了解,个体 化早期诊断和个体化医疗的时代必将来临。

成本低;

孔速度难以控制

参考文献(References)

- Margulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high-density picolitre reactors [J]. Nature, 2005, 437 (7057): 376-380
- [2] Bentley D R, Balasubramanian S, Swerdlow H P, et al. Accurate whole human genome sequencing using reversible terminator chemistry[J]. Nature,2008,456(7218):53-59
- [3] Valouev A, Ichikawa J, Tonthat T, et al. A high-resolution, nucle-

- [4] Thomas R K, Nickerson E, Simons J F, et al. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing[J]. Nat Med,2006,12(7):852-855
- [5] Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual[J]. Nature,2008,456(7218):60-65
- [6] Shen Y, Sarin S, Liu Y, et al. Comparing platforms for C. elegans mutant identification using high-throughput whole-genome sequencing[J]. PLoS One,2008,3(12):4012
- [7] Rothberg J M, Hinz W, Rearick T M, et al. An integrated semiconductor device enabling non-optical genome sequencing [J]. Nature,2011,475(7356):348-352
- [8] Howden B P, McEvoy C R, Allen D L, et al. Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR [J]. PLoS Pathog, 2011,7(11):e1002359
- [9] Rohde H, Qin J, Cui Y, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4 [J]. N Engl J Med,2011,365 (8):718-724
- [10] Miller W, Hayes V M, Ratan A, et al. Genetic diversity and population structure of the endangered marsupial Sarcophilus harrisii (Tasmanian devil)[J]. Proc Natl Acad Sci USA,2011,108(30): 12348-12353
- [11] Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome[J]. Science, 2008,320(5872):106-109
- [12] Bowers J, Mitchell J, Beer E, et al. Virtual terminator nucleotides for next-generation DNA sequencing [J]. Nat Methods,2009,6(8):593-5 95
- [13] Pastor W A, Pape U J, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells [J]. Nature,2011, 473(7347):394-397
- [14] Goren A, Ozsolak F, Shoresh N, et al. Chromatin profiling by directly

sequencing small quantities of immunoprecipitated DNA [J]. Nat Methods,2010,7(1):47-49

- [15] Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules[J]. Science, 2009, 323(5910): 133-138
- [16] Flusberg B A, Webster D R, Lee J H, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing [J]. Nat Methods, 2010, 7(6): 461-465
- [17] Uemura S, Aitken C E, Korlach J, et al. Real-time tRNA transit on single translating ribosomes at codon resolution[J]. Nature, 2010, 464 (7291):1012-1017
- [18] Grad Y H, Lipsitch M, Feldgarden M, et al. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011 [J]. Proc Natl Acad Sci U S A,2012,109(8):3065-3070
- [19] Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing [J]. Nat Nanotechnol, 2009,4(4):265-270
- [20] Stoddart D, Heron A J, Mikhailova E, et al. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore[J]. Proc Natl Acad Sci U S A,2009,106(19):7702-7707
- [21] Branton D, Deamer D W, Marziali A, et al. The potential and challenges of nanopore sequencing [J]. Nat Biotechnol,2008,26(10):1146-1153
- [22] Sims P A, Greenleaf W J, Duan H, et al. Fluorogenic DNA sequencing in PDMS microreactors[J]. Nat Methods,2011,8(7):575-580
- [23] Susan H, James B, Richard W. Methods for real-time single molecule sequence determination[P]. US patent 7329492,2008
- [24] Postma H W. Rapid sequencing of individual DNA molecules in graphene nanogaps[J]. Nano Lett,2010,10(2):420-425
- [25] Lagerqvist J, Zwolak M, Di Ventra M. Fast DNA sequencing via transverse electronic transport[J]. Nano Lett,2006,6(4):779-782
- [26] Tanaka H, Kawai T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope [J]. Nat Nanotechnol,2009,4 (8):518-522

(上接第 3727 页)

- [10] Blanco E, Farre D, Alba M M, et al. ABS: a database of annotated regulatory binding sites from orthologous promoters [J]. Nucleic Acids Res,2006,34:D63-D67
- [11] He X, Ling X, Sinha S. Alignment and prediction of cis-regulatory modules based on a probabilistic model of evolution [J]. PLoS Comput Biol,2009,5(3):1000299
- [12] Stormo G D. DNA binding sites: representation and discovery [J]. Bioinformatics,2000,16:16-23
- [13] PairoE, Maynou J, Marco S, et al. A subspace method for the detection of transcription factor binding sites [J]. Bioinformatics, 2012[Epub ahead of print]
- [14] Qian J, Ferguson TM, Shinde DN, et al. Sequence dependence of isothermal DNA amplification via EXPAR [J]. Nucleic Acids Res, 2012,40(11):e87
- [15] Li Q Z, Lin H. The recognition and prediction of σ 70 promoters in Escherichia coli K-12 [J]. J. Theor. Biol,2006,242:135-141