

## 第三代测序技术及其应用\*

张得芳<sup>1</sup> 马秋月<sup>2</sup> 尹佟明<sup>1</sup> 夏 涛<sup>1,2\*\*</sup>

(1 南京林业大学 林木遗传与生物技术省部共建教育部重点实验室 南京 210037)

(2 南京林业大学森林资源与环境学院 南京 210037)

**摘要** 第二代测序技术发展到现在一定阶段以后其缺陷逐渐显现,而第三代测序技术的出现在一定程度上弥补了第二代测序技术在应用方面的缺点。就目前正在发展的5种第三代高通量测序技术的原理进行了阐述,并比较了这几种测序技术的优缺点,最后对三代测序技术在基因组测序,甲基化研究, RNA 测序以及医学方面的应用作了简单介绍。

**关键词** 第三代测序 单分子测序技术 SMRT tSMS

**中图分类号** Q523+.8

随着测序技术的不断发展,第二代高通量测序技术已经广泛地应用于各项研究领域中,但其不足之处也日益凸显。例如,第二代测序读长较短,对后续的序列拼接、组装以及注释等生物信息学分析带来困难<sup>[1]</sup>。该技术原理是建立在 PCR 的基础上,扩增后得到的 DNA 分子片段的数目和扩增前 DNA 分子片段的数目比例有相对偏差,而对基因表达分析有很大的影响<sup>[2]</sup>,尤其是对大量表达的基因影响会更大<sup>[3]</sup>。这些缺点在一定程度上制约了第二代测序技术的应用和发展,因此第三代单分子测序技术应运而生<sup>[4,6]</sup>,因其采用单分子读取技术,有着更快的数据读取速度;同时不需要 PCR 扩增步骤,进一步降低了测序成本。

现有的第三代测序平台主要有: Helicos biosciences 公司的 tSMS<sup>TM</sup> (true single molecular sequencing) 技术平台, 美国 Pacific Biosciences 公司的 SMRT (single molecule real-time) 技术平台; 美国 Life Technologies 公司的基于 FRET 测序技术和美国 Ion Torrent 公司、英国 Oxford Nanopore Nechnologies 公司的纳米孔单分子技术。本文就该五种第三代单分子测序平台和技术的原理、优缺点和各方面的应用进行综合介绍。

### 1 测序平台介绍

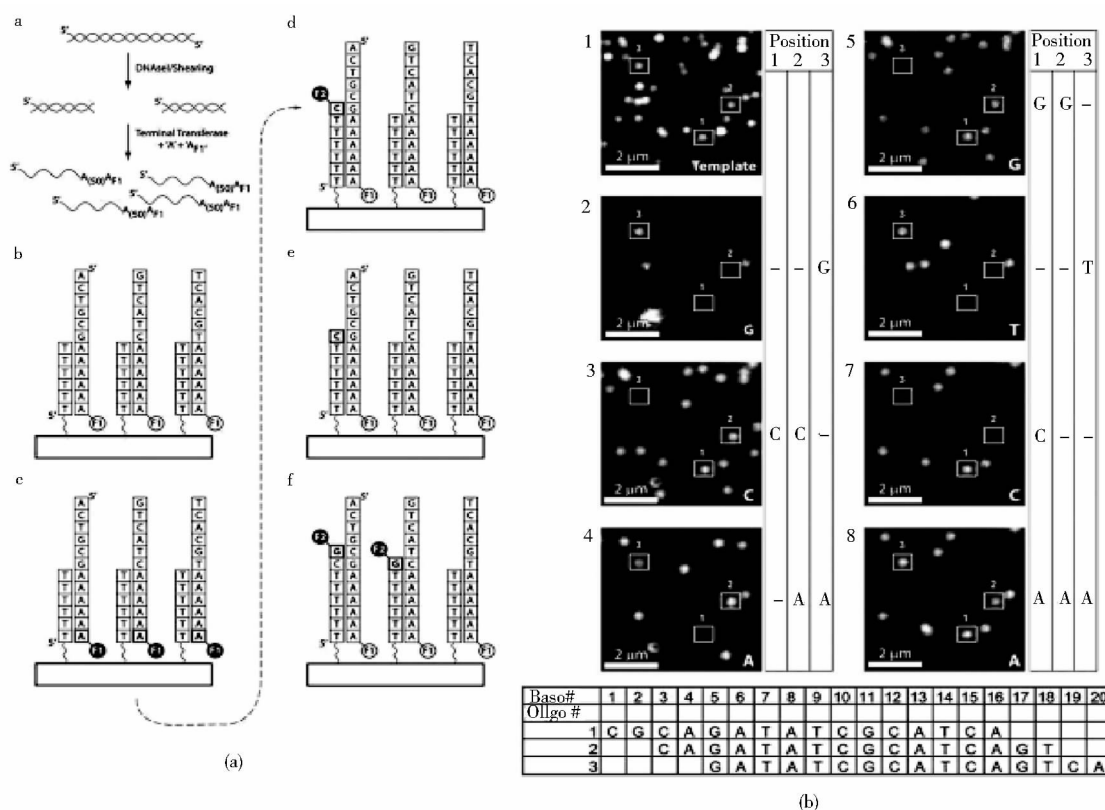
#### 1.1 并行单分子合成测序技术

Helicos Biosciences 是第一个设计开发单分子测序方法(tSMS<sup>TM</sup>)技术平台的公司<sup>[7-8]</sup>,其基础主要来自于 Braslavsky 等人的研究<sup>[9]</sup>。主要是利用合成测序理论<sup>[10]</sup>(图1),测序时首先将待测序列打断成小片段并在3'末端加上 polyA(图1a),并用末端转移酶阻断(图中标注为 F1),同时在玻璃芯片上随机固定多个 polyT 引物(其末端皆带有荧光标记),将小片段 DNA 模板与检测芯片上的 polyT 引物进行杂交并精确定位(图1b),通过成像来精确定位杂交模板所处的位置,建立边合成边测序的位点(图1c);逐一加入荧光标记的单色末端终止子及聚合酶的混合液孵育,洗涤,利用全内反射显微镜(total internal reflection microscopy, TIRM)进行单色成像(图1d),之后切开发光染料和抑制基团,洗涤,加帽,允许下一个核苷酸的掺入(图1e)。通过掺入、检测和切除的反复循环,就可以实现实时测序<sup>[10]</sup>。由于该技术采用了 Cy5(吡啶-5-菁)荧光基团(具有很好的光稳定性、高水溶性和高荧光效率,激发波长在 647nm 处)和灵敏的监测系统,能够直接记录到单个碱基的荧光,从而克服了其他方法须同时测数千个相同基因片段以增加信号亮度的缺陷。

收稿日期:2013-01-15 修回日期:2013-02-27

\* 林业公益性行业重大项目(201304102)国家自然科学基金(31270711)资助项目

\*\*通讯作者,电子邮箱:xiataonj@163.com



**图1 Helicos 单分子测序样本的准备及单核苷酸合成反应的图像采集**  
具体模板核苷酸碱基的加入、清洗及荧光位点的切除等示意图

**Fig.1 Helicos single-molecule sequencing sample preparation and imaging of single-nucleotide incorporation (a)**  
**Image series illustrating template-specific base addition, successful rinsing, and successful linker cleavage (b)**

该平台文库制备简单,不需要 PCR 扩增或连接酶,特别适合 RNA 直接测序的应用。Helicos 公司的研究人员利用 tSMS™ 进行 RNA 直接测序的可行性进行了验证。

该公司 HeliScope 测序仪的优势是样本通量非常高,2 个流动槽可同时运行,每个流动槽有 25 个独立通道,每个通道又可以运行最多 96 个标记分子条形码的样本,这样每次运行的样本数可高达 4 800 个;缺点是测序的平均读长相对较短,只有 35 bp,准确率较低,约为  $\leq 97\%$  [11],还有待于进一步的改进(表 1)。

## 1.2 单分子实时合成测序技术

Helicos Biosciences 公司虽然第一个成功的开发了单分子测序技术,但是将该技术成功进行商业化的是 Pacific Biosciences 公司推出的单分子实时 DNA 测序仪 (SMRT) <sup>[12]</sup>。

SMRT 技术是基于边合成边测序的思想,以 SMRT 芯片为载体进行测序反应。SMRT 芯片是一种带有很

多零模式波导孔 (zero-mode waveguides, ZMW) (厚度为 100 nm) 的金属片, ZMW 是一种直径只有几十个纳米的孔, 由于其底部上的小孔短于激光的单个波长, 导致激光无法直接穿过小孔, 而会在小孔处发生光的衍射, 形成局部发光的区域, 即为荧光信号检测区 (图 2a 中的白色区域), 该区域内锚定有 DNA 聚合酶。测序时将基因组的 DNA 打断成许多小的片段, 制成液滴后将其分散到不同的 ZMW 纳米孔中。当 ZMW 孔底部聚合反应发生时, 被不同荧光标记的核苷酸会在小孔的荧光探测区域中被聚合酶滞留数十毫秒, 荧光标记会在激光束 (Excitation) 的激发下发出荧光, 根据荧光的种类就可以判定 dNTP 的种类, 反应完成后荧光标记会被聚合酶切除而弥散出 ZMW 小孔, 其它未参与合成的 dNTP 由于没进入荧光信号检测区而不进入荧光信号检测区<sup>[13-14]</sup>。

SMRT 测序时,样品准备过程涉及到样品 DNA 的打断、末端补齐、连接接头、测序这几个步骤<sup>[15]</sup>。测序

中需要的样品量很少,样品准备中所用的试剂也很少,而且测序过程中省去扫描和洗涤的过程,所以测序所花的时间较少<sup>[16]</sup>。

Pacific Biosciences 公司第三代单分子实时测序系统最初的 C1 试剂测序读长为 1 300bp,数据量也较少。新版 C2 试剂推出后,片段平均读长提升至 3 000bp,数据量也随之提高了 3~4 倍。目前公布出来的 C2 试剂的参数是:在 Long 模式下,平均读长可达到 2 500bp,95% ile 的读长可达到 6 500bp;在 X-Long 模式下,平均读长可达到 3 000bp,95% ile 的读长可达到 8500bp。使用 C2 试剂,每个 SMRT 孔可以获得 90M 的可定位数据(Mappable data)。现阶段每天最多可运行 12 个 SMRT 孔,因此每天可获得  $12 \times 90 = 1\,080\text{Mb}$  的可定位数据量。

SMRT 技术对于聚合酶的聚合速度和持续合成能力达到一个较好的平衡,所以读长较长,这一点在很大程度上优于二代测序<sup>[16-17]</sup>。所以在序列拼接、定位以及需要跨越重复区域的应用中有着极大优势。另外,随着读长的增加,拼接过程中需要测序覆盖深度也会随之下降。长读长(Long reads)可以帮助研究者对变异进行更准确的定位(表 1)。另外,该平台通量高、费用较低、耗时短<sup>[18-19]</sup>。

SMRT 技术的缺陷是会出现插入和缺失错误。缺失错误源于有时候碱基掺入速度过快,超过了相机的拍摄帧数;插入错误源于有时候酶随机的选择一些碱

基,但并未真的将这些碱基掺入合成链中。但这些错误是随机的,并不会随着读长的增加而提高,随着测序覆盖深度的增加会逐渐被消除。未来随着试剂的不断优化以及每个 SMRT 孔可获得的数据量的增加,单分子测序的原始准确度会逐步提高,且也会进一步提高。

1.3 纳米孔单分子技术

Oxford Nanopore Technologies 公司研发的测序技术(图 2b)完全不同于前两种测序技术,其核心就是将在某一面含有一对电极的特殊的脂质双分子层置于一个微孔之上,该双分子层中含有很多由  $\alpha$  溶血素蛋白组成的纳米孔(图 2b 中紫色区域)中结合一个核酸外切酶。当 DNA 模板进入孔道时,孔道中的核酸外切酶会“抓住”DNA 分子,顺序剪切掉穿过纳米孔道的 DNA 碱基,每一个碱基通过纳米孔时都会产生一个阻断,根据阻断电流的变化就能检测出相应碱基的种类,最终得出 DNA 分子的序列<sup>[20-22]</sup>。

纳米孔单分子测序技术的一大优势就是仪器构造简单使用成本低廉;因为它不需要对核苷酸进行标记,也不需要复杂的光学探测系统(如激光发射器和 CCD 信号采集系统等),能直接对 RNA 分子进行测序。同时由于它是直接检测每一个碱基的特征性电流,因而能对修饰过的碱基进行测序,这一点对于表观遗传学研究具有极高的价值;缺点就是它采用的是水解测序法,不能进行重复测序,因而无法达到一个满意的测序精确度(表 1)。

表 1 第三代测序技术的各项参数的比较

Table1 The different parameters comparison of the third generation sequencing technologies

测序平台	读长	准确率	Reads/循环	耗时/循环	费用/百万个碱基	优点	缺点
Helicos (tSMS)	25 ~45bp	97%	10 亿	约 8d	\$ 1	产量高,文库制备简单,不需要 DNA 扩增或连接	错误率高,仪器贵
PacBio(SMRT)	1000 ~4500bp	87% (读长模式), 99% (准确性模式)	3.5-7.5 万	0.5-1h	\$ 2	长读长,速度快	在准确性模式下产率较低,仪器贵
Oxford Nanopore	100kb	96%	-	-	-	有潜力达到高读长;可以成本生产纳米孔;无需荧光标记或光学手段	切断的核苷酸可能被读错方向;难于生产出带多重平行孔的装置
FRET	>1500bp	>99%	-	20min	-	-	-
Ion Torrent	200bp	>98%	>500 万	2h	\$ 1	仪器便宜,速度快	同聚体错误

1.4 基于 FRET 的测序技术

基于荧光共振能量传递(fluorescence resonance energy transfer, FRET)的测序仪最初是由 VisiGen Biotechnologies 公司开发的。基本原理是先将供体荧光基团修饰的

DNA 聚合酶和 DNA 模板分子固定在固体表面,在被不同颜色的受体荧光基团标记(荧光标记于  $\gamma$  磷酸盐上,反应完成后会被切除)的核苷酸掺入时,会释放一个 FRET 信号,使 DNA 分子发光,而发光的颜色则提示了碱基的类

别(图 2c)<sup>[23-24]</sup>。该过程比较简单直观,就像放电影一样。但是该平台的应用较少,缺乏相应的技术参数。

### 1.5 半导体测序技术

半导体测序技术是由 Ion Torrent 研制开发。由于该技术中也用了 Emulsion PCR 技术,其实质是介于第二代和第三代测序技术之间。该技术使用了一种高密度半导体芯片,芯片上布满了小孔(图 2d 中的蓝色部分),每个小孔就是一个单独的测序反应池。实验时先

将芯片置于一个离子敏感层(ion sensitive layer)和离子感受器(ion sensor)之上,当 DNA 聚合酶在每一个单分子模板链上滑动时,发生聚合反应,释放出氢离子( $H^+$ ),最终离子感受器就会捕捉到这种信号,从而读出 DNA 序列(图 2d)。由于该硬件设备无需光学检测和扫描系统,并且使用天然核苷酸和聚合酶、无需焦磷酸酶化学级联,无需标记荧光染料和化学发光的配套试剂,具有制造成本低和测序费用低的优势。

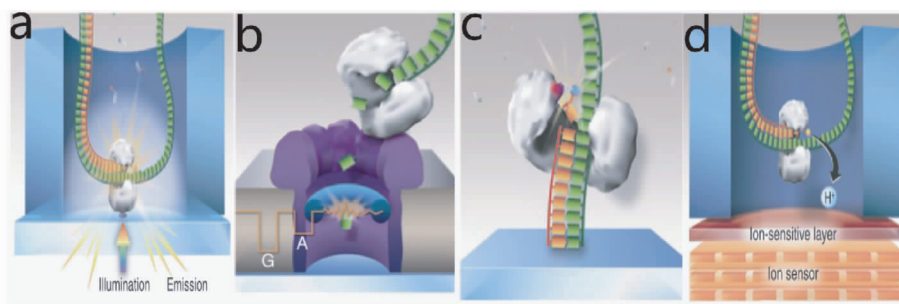


图 2 第三代测序平台示意图

Fig. 2 The third generation sequencing platforms

(a) Pacific Biosciences SMRT Real-time single-molecule DNA sequencing platforms (b) The Oxford nanopore sequencing platforms (c) Life Technologies co. FRET sequencing platforms (d) The ion Torrent sequencing platforms

## 2 第三代单分子测序技术的应用

由于单分子测序具有通量更高,仪器和试剂相对便宜、操作简单等的优势而比第二代测序技术有更广阔的应用空间。

### 2.1 基因组测序

由于具有读长长的特点,SMRT 测序平台在基因组测序中能降低测序后的 Contig 数量,明显减少后续的基因组拼接和注释的工作量,节省大量的时间<sup>[25]</sup>。Christophersen 等<sup>[26]</sup>仅仅用 0.5 × 的 PacBio RS 系统长度的数据与 38 × 的二代测序(NGS)的测序数据,对马达加斯加的一种指猴基因组进行拼装,大幅度提高了数据的质量和完整性,同时借助 Pacbio RS 的帮助将原有的 Contig 数量减少了 10 倍。David A. 等<sup>[27]</sup>利用 PacBio RS 平台和 C2 试剂通过全球合作几天内就完成了从德国大肠杆菌疫情中获得的大肠杆菌样品以及近似菌株的测序和数据分析,最终获得了 2 900bp 的平均读长以及 99.998% 的一致性准确度。在对霍乱病菌的研究中,第三代测序技术已初现锋芒。研究人员对 5 株霍乱菌株的基因组进行了测序研究,并与其他 23 株霍乱弧菌的基因组进行对比。结果发现海地霍乱菌株与 2002 年和 2008 年在孟加拉国分离得到的变异霍乱弧

菌 El Tor O1 菌株之间关系密切,而与 1991 年拉丁美洲霍乱分离株的关系较远。相对 NGS 的优势就是能更快获得结果,因此该系统在鉴定新的病原体 and 细菌的基因组测序方面得到很广泛的应用<sup>[28]</sup>。

在测序成本方面,Quake 利用一台 Helico 技术平台对自身的基因组进行了测序。实验仅用了 4 个星期,花费 48 000 美元。该结果覆盖了 90% 的人类参考基因组,覆盖度达 28 倍,序列读长为 24 到 70bp,平均读长为 32bp,预测出基因组中的 2 805 471 个 SNP,与 Illumina Human 610-Quad 的 SNP BeadArray 检测到的结果进行比较,一致性达到 99.8%<sup>[29]</sup>。Harris 等利用 Helicos 平台完成了 M13 细菌基因组的重测序,平均测序深度大于 150 ×,覆盖度达到 100%。他们用短小的 DNA 标签加在 M13 的基因组打断的小片断上,加入 DNA 复制酶和带有荧光标签的碱基或碱基对。当荧光 DNA 形成链时,就拍下每个新加上的碱基对。该方法成功的检测到单个碱基的荧光,实现了高通量、低费用的重测序技术开发<sup>[30]</sup>。

### 2.2 甲基化研究

SMRT 技术采用的是对 DNA 聚合酶的工作状态进行实时监测的方法,聚合酶合成每一个碱基,都有一个时间段,而当模板碱基带有修饰时,聚合酶会慢下来,

使带有修饰的碱基两个相邻的脉冲峰之间的距离和参考序列的距离之间的比值如果大于1,由此就可以推断这个位置有修饰。甲基化研究中关于5-mC和5-hmC(5-mC的羟甲基化形式)是甲基化研究中的热点。但现有的测序方法无法区分5-mC和5-hmC。美国芝加哥大学利用SMRT测序技术和5-hmC的选择性化学标记方法来高通量检测5-hmC。通过聚合酶动力学提供的信息,可直接检测到DNA甲基化,包括N6-甲基腺嘌呤、5-mC和5-hmC<sup>[31-32]</sup>,为表观遗传学研究打开了一条通路。

### 2.3 突变鉴定(SNP检测)

单分子测序的分辨率具有不可比拟的优势,而且没有PCR扩增步骤,就没有扩增引入的碱基错误,该优势使其在特定序列的SNP检测,稀有突变及其频率测定中大显身手。例如在医学研究中,对于FLT3基因是否是急性髓细胞白血病(AML)的有效治疗靶标一直存在质疑。研究人员用单分子测序分析耐药性患者基因,意外发现耐药性与FLT3基因下游出现的稀有新突变有关,重新证明了FLT3基因是这种最常见白血病—急性髓细胞白血病(AML)的有效治疗靶标,打破了一直以来对于这一基因靶标的疑惑。凭借PacBio平均3 000bp的读长,获得了更多基因下游的宝贵信息,而基于单核酸分子的测序能够检测到低频率(低至1%)罕见突变,正是这项成果的关键所在<sup>[33]</sup>。

### 2.4 RNA测序

根据三代测序技术实时测序的特点,可以直接对RNA进行测序<sup>[34]</sup>,做到了对特定组织和细胞内的表达差异的精确定位。运用Helicos操作平台对酵母的聚腺苷酸化的转录本进行精确定量<sup>[35]</sup>,用50个通道中的6个通道,共产生2.4亿个Reads。利用Helicos遗传分析平台可以将DNA聚合酶换成反转录酶对RNA直接测序,主要通过Poly(dT)包被的表面捕获聚腺苷酸化的RNA,成功完成对酿酒酵母RNA的直接测序,免除将RNA转变成cDNA的过程<sup>[36]</sup>。同时借助PacBio技术平台,也可完成对RNA的直接测序,Uemura等<sup>[37]</sup>就利用该技术进行实时测序观察核糖体中mRNA的翻译过程。单分子测序技术最大的优势,也是最直接的应用就是检测细胞和组织内基因表达水平,同时对基因的结构做出分析,如RNA的剪接,是否有碱基突变或异常表达基因等。而且还能检测到微量的基因表达子或罕见的非编码RNA<sup>[38]</sup>。PacBio RS还可对连续的A或者T区域测序,因为PolyA长度和RNA的半衰期有关,

所以对长PolyA的研究对RNA的代谢有重要意义。

### 2.5 重复序列和poly结构的测序

在所有人的X染色体上都有一段CGG三核苷酸重复序列,正常人的CGG重复次数为5-44次。过长的重复次数会对FRM1基因转录或翻译出FRM1蛋白不利,当重复次数超过200次时,就会导致脆性X综合征,所以检测CGG的重复次数非常有意义。但是这个重复长度的测序用Sanger测序和第二代测序技术都是难以完成的。美国UC Davis医学院利用PacBio技术环形比测序模式,对FMR1中的CGG三核苷酸重复区域进行了测序,获得了超过10kb的原始读长,覆盖了CGG重复超过750次的三核苷酸重复区域<sup>[39]</sup>。

### 2.6 医学领域的应用

加拿大安大略省癌症研究所通过PacBio RS系统进行临床样本的癌基因及癌症治疗敏感/抗性相关的遗传标记的测定,辅助病人的后续治疗方案。该研究所已经通过先期测试积累了大量的经验和数据,正在加快后续大规模临床测试及以此为基础的个性化治疗的进程<sup>[40]</sup>。

单分子测序技术的出现,对个体医学研究产生的影响是最大的。由于测序通量增加,所以在短时间内就可以完成人体染色体将近300亿个碱基对的测序。而费用降低,基因组测序费用将会接近甚至达到1 000美元基因组的目标。这就可以达到个体医学的初步任务,即建立个人基因组信息档案<sup>[41]</sup>。当得到个体的基因组信息档案后,个体的生命信息就一目了然,那么该个体是否具有某种疾病的易感性,或者在患有某种疾病以后对治疗方法的敏感性以及以后都可以在一定的范围内得到推断。这就使得医疗能以个人的基因组基本信息作为诊断、治疗和防止的手段。由此可见单分子测序技术将会对个体医学的发展产生巨大的影响<sup>[34]</sup>。

### 2.7 其他方面的应用

除了以上提到的几个应用方面外,第三代单分子测序技术在其他方面也有应用。在二代测序数据拼接中,第三代测序技术读长长的优势被用来补Scaffolds之间的缺口<sup>[42]</sup>。由于测序时对GC没有偏好性,该特点也可以被用于高GC含量的区域测序中。同时,三代测序技术也能被用于具有复杂二级结构的测序中,例如发夹结构,茎环结构等。

## 3 展望

第三代测序技术由于其特有的优势必将会在很多

研究领域替代一代和二代测序技术平台,使测序成为中小型实验室的常规分析手段。当然,单分子测序技术还处于初步发展的阶段,其缺点也是客观存在的,还有很多需要完善的地方。

同时第三代测序技术的进一步发展也面临着来自各方面的挑战,例如,记录数据所用的工程学方面和光学方面的挑战;测序反应时需要的化学和生物工程领域的挑战以及相比二代测序更多的数据处理能力等<sup>[43]</sup>。

### 参考文献

- [ 1 ] Pop M, Salzberg S L. Bioinformatics challenges of new sequencing technology. *Trends Genet.* 2008, 24(3):142-149.
- [ 2 ] Acinas S G, Samm R R, Klepac-Ceraj V, et al. PCR induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol*, 2005, 71(12):8966-8969.
- [ 3 ] Tortes T T, Metta M, Ottenwlder B, et al. Gene expression profiling by massively parallel sequencing. *Genome Res.* 2008, 18(1): 172-177.
- [ 4 ] Derrington I M, Butler T Z, Collins M D. et al. Nanopore DNA sequencing with Msp. A. *Proc. Natl Acad. Sci. USA*, 2010, 107: 6060- 6065.
- [ 5 ] Luan B, Peng H, Polonsky S, et al. Base-By-Base ratcheting of single stranded DNA through a solid-state nanopore. *Phys. Rev. Lett.*, 2010, 104: 238103.
- [ 6 ] Eid J, Fehr A, Gray J, et al. , Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, 323, 133-138.
- [ 7 ] Bowers J, Mitchell J, Beer E, et al. Virtual terminator nucleotides for next-generation DNA sequencing. *Nat. Methods*, 2009, 6, 593-595.
- [ 8 ] Tessler L A, Reifengerger J G, Mitra R D. Protein quantification in complex mixtures by solid phase single-molecule counting. *Anal Chem*, 2009, 81, 7141-7148.
- [ 9 ] Braslavsky I, Hebert B, Kartalov E, et al. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci USA.* 2003, 100(7): 3960-3964.
- [ 10 ] Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320 ( 5872 ): 106-109.
- [ 11 ] Ozsolak F, Platt A R, Jones D R, et al. Direct RNA sequencing. *Nature*, 2009, 461(7265): 814-818.
- [ 12 ] Pacific B, PacBio M P C A, USA on World Wide Web URL: <http://www.pacificbiosciences.com>.
- [ 13 ] Stephen Turner. Single Molecule Real Time (SMRT<sup>TM</sup>) DNA Sequencing, pacific biosciences, 1-12.
- [ 14 ] <http://www.pacificbiosciences.com/products/smrt-technology>
- [ 15 ] Travers K J, Chin C S, Rank D R, et al. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucl Acids Res.* 2010, 38(15): 159.
- [ 16 ] Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, 323(5910): 133-138.
- [ 17 ] Levene M J, Korlach J, Turner S W, et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 2003 299(5607): 682-686.
- [ 18 ] Shendure J A, Porreca G J, Church G M, et al. Overview of DNA sequencing strategies, current protocols in molecular biology, 2011, chapter7, 7.1.1-1.1.23.
- [ 19 ] Schadt E E, Turner S and Kasarskis A. A window into third-generation sequencing, *human molecular genetics*, 2010, 19 (2): 227-240.
- [ 20 ] Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol*, 2009, 4: 265-270.
- [ 21 ] Howorka S, Cheley S, Bayley H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.*, 2001, 19: 636-639.
- [ 22 ] Stoddart D, Heron A J, Mikhailova E, et al. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl Acad. Sci USA*, 2009, 106: 7702-7707.
- [ 23 ] Rahul Roy, Sungchul H, Taekjip H. A practical guide to single-molecule FRET. *Nature Methods*. 2008, 5: 507-516.
- [ 24 ] Hardin S, Gao X L, Briggs J, et al. America, Methods for real-time single molecule sequence determination. *US7329492*. 2008.
- [ 25 ] English A C, et al. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE*. 2012, 7(11): 47768.
- [ 26 ] Perry G H, Reeve D, Melsted P, et al. A Genome Sequence Resource for the Aye-Aye (*Daubentonia madagascariensis*), a Nocturnal Lemur from Madagascar. *Genome Biology and Evolution*, 2011, 4(2): 126-135.
- [ 27 ] David A R. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in germany. *N Engl J Med.* 2011, 365:709-717.
- [ 28 ] Ribeiro F, Przybylski D, Yin S, et al. Finished bacterial genomes from shotgun sequence data, *Genome Res.* 2012. 22 (11):2270-2277.
- [ 29 ] Pushkarev D, Neff N F, Quake S R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol*, 2009, 27:

- 847-852.
- [30] Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320: 106-109.
- [31] Song C X, Clark T A, Yu X, et al. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat. Meth.* 2012, 9: 75-77.
- [32] Flusberg B A, Webster D R, Lee J H, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Meth.* 2010, 7: 461-465
- [33] Catherine C S, Qi W, Chin C S, et al. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature*. 2012, 485: 260-263.
- [34] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009, 10: 57-63.
- [35] Lipson D, Raz T, Kieu A, et al. Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* 2009, 27: 652-658.
- [36] Ozsolak F, Platt A R, Jones D R, et al. Direct RNA sequencing. *Nature*, 2009, 461: 814-818.
- [37] Uemura S, Aitken C E, Korlach J, et al. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature*, 2010, 464: 1012-1017.
- [38] Philipp K, Fatih O, Kim S W, et al. New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. *Nature*, 2011, 466: 642-646.
- [39] Erick W, Loomis, John S Ei, et al. Sequencing the unsequenceable: Expanded CCG-repeat alleles of the fragile X gene. *Genome Res.* 2013, 23: 121-128.
- [40] Janet E, Dancey G, philippe L, et al. The genetic basis for cancer treatment decisions. *Cell*, 2012, 148(3): 409-420.
- [41] Wheeler D A, Srinivasan M, Egholm T, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 2008, 452: 872-876.
- [42] Zhang X, Davenport K W, Gu W, et al. Improving genome assemblies by sequencing PCR products with PacBio, *BioTechniques*, 2012, 53: 61-62.
- [43] Shen S Q. Single molecule sequencing and individual medical. *Progress in physiological sciences.* 2009, 40(3): 283-288.

## The Third Generation Sequencing Technology and Its Application

ZHANG De-fang<sup>1</sup> MA Qiu-yue<sup>1</sup> YIN Tong-ming<sup>1</sup> XIA Tao<sup>2</sup>

(1 Key Lab of Forest Genetics and Biotechnology Nanjing Forestry University, Nanjing 210037, China)

(2 Forestry Resources and Environment institute, Nanjing Forestry University, Nanjing 210037, China)

**Abstract** With the development and application of next generation sequencing technology (NGS), some disadvantages are emerging. The third generation single molecular sequencing technology, to a certain extent, may offset these disadvantages of the next generation sequencing in application. This study elaborated the 4 sequencing principles of the third generation single molecular sequencing technologies, compared the advantages and disadvantages with these sequencing technologies, and presented their application in genome sequencing, methylation research, RNA sequencing and medical research.

**Key words** The third generation sequencing technology Single molecular sequencing technology SMRT tSMS