

Genome Sequencer FLX 引领快速基因组 测序时代的到来

许冠东*

(罗氏应用科学部 罗氏诊断产品有限公司 上海 200031)

摘要: 焦磷酸法最早是应用在检测 DNA 甲基化和 SNP 位点分析的一项技术, 在 2005 年底, 此技术被用来进行基因组序列的测定, 已经成为下一代高通量测序中最成熟的一种技术。本篇文章着重介绍了采用焦磷酸测序法的罗氏公司最新一代高通量基因组测序系统 GS FLX 的技术原理, 操作过程和广泛的应用范围。

关键词: GS FLX, 焦磷酸测序, 高通量, 读长, 应用

GS FLX, the Leader in Next Generation Sequencer

XU Guan-Dong*

(Roche Applied Science, Roche Diagnostics (Shanghai) Limited, Shanghai 200031)

Abstract: Pyrosequencing technology is famous for its application in detection of SNP site and DNA methylation. At the end of 2005, pyrosequencing technology was used to sequence the whole genome and it has become the most developed next generation high-throughput sequencing technology. In this article, we introduce the theory, operation process and wide application field of the second generation genome sequencer FLX system provided by Roche

Keywords: GS FLX, Pyrosequencing, High throughput, Read length, Application

2007 年 9 月 28 日, 在《Science》国际顶级学术期刊的网站上, 同时查询到 3 篇突破性的基础研究成果^[1~3]。这 3 篇文章的研究对象和领域完全不同, 分别涉及了对人类基因组变异的研究, 蜜蜂社会性的进化讨论以及灭绝的古生物线粒体全序列的测定。然而它们的共同点是, 它们所使用的分析数据和实验结果都来自罗氏 454 公司研发的高通量测序技术平台 Genome Sequencer™。在 1 周的时间里就有 3 篇 GS 系统在不同研究领域的应用文章发表在《Science》上, 累计已有近 100 篇 Genome Sequencer™的应用成果发表在世界知名期刊上, 说明 454 测序技术和 GS 系统在生命科学研究领域的广泛

适用性。

Genome Sequencer FLX (GS FLX)系统是罗氏 454 公司的第二代测序平台, GS FLX的命名正是来源于其多领域的灵活(flexible)应用。GS FLX系统在技术原理和性能方面的表现优越, 每个反应可以得到超过 40 万个序列读长, 单个序列的读长可达 300bp, 一个测序反应就可以产出超过 1 亿个碱基信息。目前, Genome Sequencer™系统已经被世界上几乎所有从事基因组测序和相关结构功能研究的顶级实验室配备使用, 相信随着该系统性能和应用领域的不断提升和扩展, 必将带来整个测序领域的技术发展, 对大规模基因序列研究的相关应用领域产生

* 通讯作者: Tel: 010-85181622; ✉: guan-dong.xu@roche.com
收稿日期: 2007-10-12

巨大的推动作用,例如基因组学、比较基因组学、转录组分析和基因调控等。

1 GS FLX 系统超高通量测序技术原理

与 GS 20 一样,GS FLX 系统的测序原理也是基于焦磷酸测序法,一种依靠生物发光进行 DNA 序列分析的新技术;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,将引物上每一个 dNTP 的聚合与一次荧光信号释放偶联起来(图 1)。通过检测荧光信号释放的有无和强度,就可以达到实时测定 DNA 序列的目的。此技术不需要荧光标记的引物或核酸探针,也不需要进行电泳;具有分析结果快速、准确、高灵敏度和高自动化的特点。

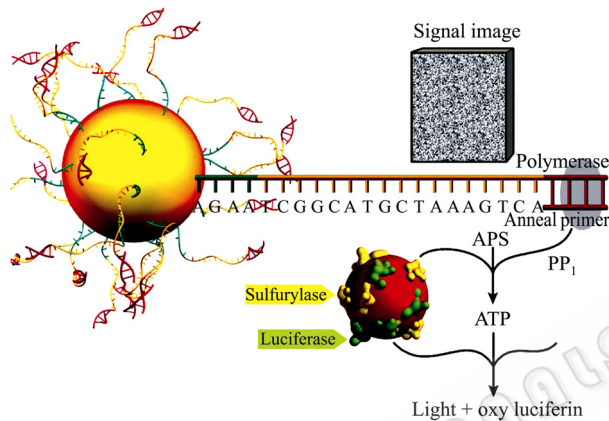


图 1 GS FLX 高通量测序方法原理示意图

Fig.1 Sequencing principle of GS FLX system

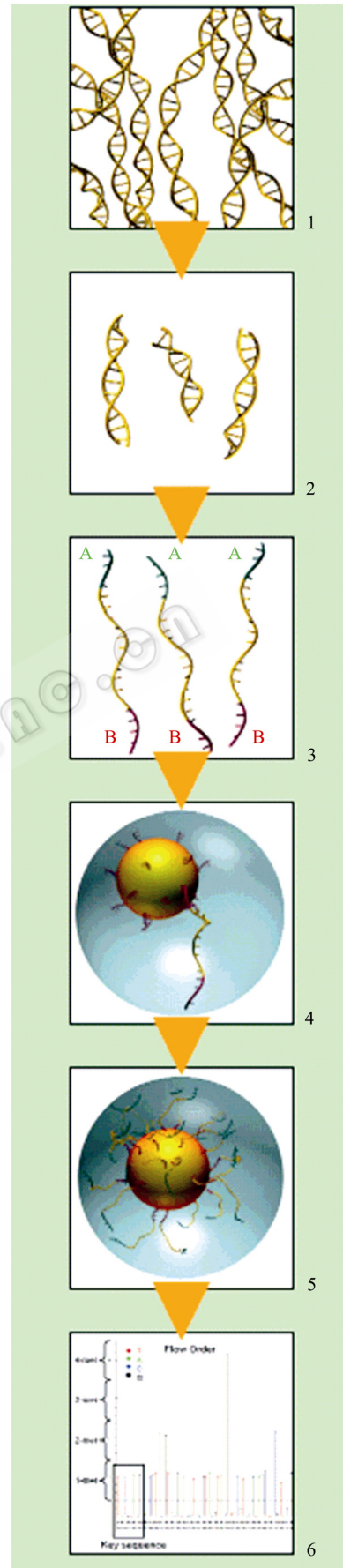
2 GS FLX 系统的操作过程

GS FLX 系统减少了传统测序样品制备过程中对大型自动站的依赖,完全不需要进行繁琐的建库过程,也不必进行克隆挑取、微孔板处理等工作;利用创新的磁珠和 emPCR 技术进行样品制备,同时提供完整的后续生物信息学分析解决方案。

1) 样品种类:GS FLX 系统支持各种不同来源的样品序列测定:包括基因组 DNA,PCR 产物,BACs,cDNA,及小分子 RNA 等等;不同类型的样品测序都可在一台仪器上完成。

2) 样品 DNA 打断:样品如基因组 DNA 或 BAC 等被打断成 300 bp-800 bp 的片段;对于小分子的非编码 RNA,这一步骤并不需要。短的 PCR 产物则可利用 GS 融合引物扩增后直接进行步骤 4)。

3) 衔接子连接:借助一系列标准的分子生物学技术,将 3' 端和 5' 端有特异性的 A 和 B 衔接子连接



到 DNA 片段上。衔接子也将在后继的纯化、扩增和测序步骤中用到。图中仅仅显示了后续步骤中要用到的单链的 DNA 片段。

4) 一条 DNA 片段 = 一个磁珠：衔接子使成百上千条 DNA 片段分别结合到一个磁珠上，此磁珠被单个油水混合小滴包被后，在这个小滴里进行独立的扩增，而没有其他的竞争性或者污染性序列的影响；所有 DNA 片段进行平行扩增(emPCR)。

5) 一个磁珠 = 一条读长：经过 emPCR 扩增后，每个磁珠上的 DNA 片段拥有了成千上万个相同的拷贝。经过富集以后，这些片段仍然和磁珠结合在一起，随后就可以放入到 PicoTiterPlate 板中供后继测序使用了。

6) 数据读取和分析工具：GS FLX 系统提供三种不同的生物信息学工具对测序数据进行分析，适用于不同的应用：例如多达 3GB 序列的重测序，对比已知参考序列进行的扩增产物差异分析，及 120MB 的从头测序工作等。

3 GS FLX 系统的技术优势

3.1 基于其特有的技术原理和性能操作，GS FLX 系统具备诸多方面的卓越性能

- 7.5 个小时的一个测序反应，就可获得超过 1 亿的碱基数据
- 高通量测序产品中无可比拟的最长读长，可达 300 bp
- 通量更高，每个反应可以得到超过 40 万个序列读长
- 读长超过 200 bp 时，单一读长的准确性可以超过 99.5%
- 测序结果一致性超过 99.99%
- 无需克隆和挑取克隆的操作工程，生成的完整基因组库，避免了克隆偏差的影响
- 测序成本的大幅降低

3.2 GS FLX 系统具备完整的软件包是它的又一大特色

- 利用项目管理软件工具，可以把多次测序反应得到的数据进行成组分析，并且可以添加已完成的测序反应数据来进行其他方面的分析
- 同时可以整合 Sanger 测序结果进行拼接和作图，包括配对末端的拼接
- 支持将 GS FLX 的数据于第三方的软件工具整合，例如 Cosed
- 可自主选择拼接时的参数

- 采用新一代的算法，保证获得高度可信的结果——无论是碱基识别，拼接作图，还是差异分析

4 GS FLX 系统的广泛应用

这项技术的第一个“样本”就是来自有“DNA 之父”之称的 James D Watson，他向 454 公司提供了自己的血液样本并获得了全部的基因组序列。目前 GS 系统的用户在 Nature、Science、PNAS、Cell 等等世界顶级的期刊杂志上已经发表 100 篇的学术论文。(详细列表请查询

<https://www.rocke-applied-science.com/sis/sequencing/genome/index.jsp>)。与 GS 20 系统相比较，硬件配置和软件系统方面的革新改进，使得 GS FLX 系统有了更广泛的应用，包括：

- 基于末端配对(Paired-End)应用的从头测序(De Novo Sequencing)
- BAC 文库的重测序(Resequencing)
- 基于全基因组测序的比较基因组学研究
- 生物群总体基因组和微生物多样性研究
- 考古学
- 扩增产物的测序，识别高可信度的 SNP 位点
- 小分子 RNA
- 转录组分析
- 甲基化和基因表达调控的研究

当前全球众多从事测序技术研发的工作者都在为实现只需花费\$1000 就可获取个人基因组序列信息的目标而努力着，一旦这个目标得以实现，个性化的早期疾病诊断和医疗服务将不再只是梦想。罗氏 454 公司也将同样以此作为努力的方向，不断地提升现有 FLX 系统平台功能，包括读长的增加，改进提升数据输出和分析工具；对实验成本、覆盖率以及时间安排等方面的优化等。

参考文献

- [1] Jan O Korbelt, Alexander Eckehart Urban, et al. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science*, 2007, **318**(5849): 420-426.
- [2] Amy L Toth, Kranthi Varala, Thomas C Newman, et al. Wasp Gene Expression Supports an Evolutionary Link Between Maternal Behavior and Eusociality. *Science*, 2007, **318**(5849): 441-444.
- [3] Gilbert MTP, Tomsho LP, Rendulic S, et al. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*, 2007, **317**(5846): 1927-1930.