借助生物传感器实现基因测序的一种设想

刘正浩 2019270103005 英才实验学院

【摘要】本文简要介绍了基因测序从诞生到现在所经历的几个发展阶段以及几个具有代表性的测序方法，并提出了一种借助生物传感器进行测序的方法假设，论证了该技术的可行性。

【关键词】基因，DNA，测序，生物传感器

1. 背景

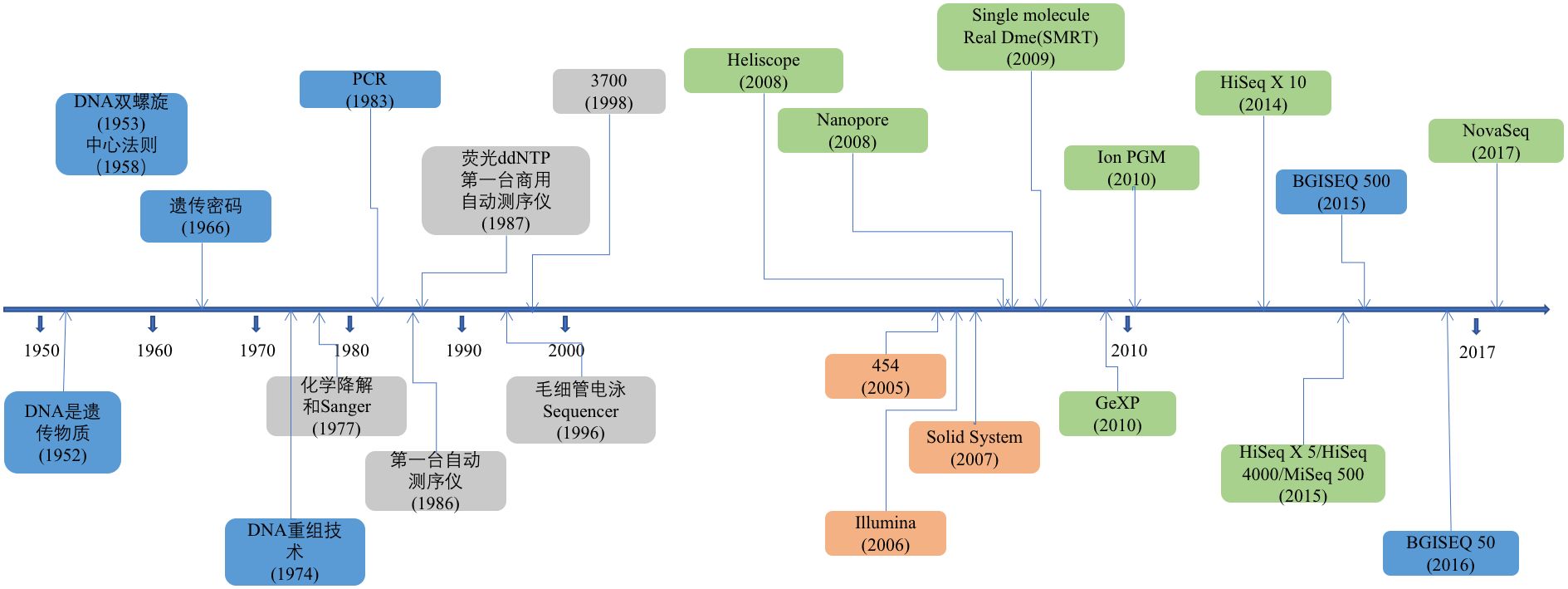
 基因测序技术又称为基因测序技术也称作DNA测序技术，即用不同手段获得目的DNA片段碱基排列顺序的技术。由于相当大的一部分生物都以DNA作为遗传物质来存储基因中的信息，获得目的DNA片段的序列就成为了进一步进行分子生物学研究和基因改造的基础。

图 1 基因测序技术的发展历程

1.1 第一代测序系统

第一代DNA测序技术所使用的是1975年由桑格（Sanger）和考尔森（Coulson）开创的链终止法或者是1976-1977年由马克西姆（Maxam）和吉尔伯特（Gilbert）发明的化学法（链降解）。在1977年，桑格测定了第一个基因组序列——噬菌体phiX-174，它的全长只有5,375个碱基。虽然与今日的技术比起来这五千多个碱基根本不算什么，但这次测序的成功进行标志着人类获得了窥探生命本质的能力，并以此为开端真正步入了基因组学时代。

Sanger法的核心原理是：由于ddNTP（4种带有荧光标记的A,C,G,T碱基）的2’和3’都不含羟基，其在DNA的合成过程中不能形成磷酸二酯键，因此可以用来中断DNA的合成反应，在4个DNA合成反应体系中分别加入一定比例带有放射性同位素标记的ddNTP（分别为：ddATP, ddCTP, ddGTP和ddTTP），然后利用凝胶电泳和放射自显影后可以根据电泳带的位置确定待测分子的DNA序列。Sanger法成为了一种相当经典的基因测序方法，许多研究人员在它的基础上改进出了许多测序的方法。这其中最著名的就是2001年完成的首个人类基因组图谱计划。

在测序技术起步发展的这一时期中，除了Sanger法之外还出现了一些其他的测序技术，如焦磷酸测序法、连接酶法等。其中，焦磷酸测序法是后来Roche公司454技术所使用的测序方法，而连接酶测序法是后来ABI公司SOLID使用的测序方法，但他们的核心手段都是利用了Sanger中可中断DNA合成反应的dNTP。

1.2 第二代测序系统

第一代测序技术准确度很高，可以达到99.999%，但是其成本高、通量低的缺点严重阻碍了它真正大规模的应用。经过不断改进，以Roche公司的454技术、Illumina公司的Solexa/HiSeq技术和ABI公司的SOLID技术为标记的第二代测序技术诞生了。第二代测序技术在大幅提高了测序速度的同时，还大大地降低了测序成本，并且保持了高准确性，以前完成一个人类基因组的测序需要3年时间，而使用二代测序技术则仅仅需要1周的时间。

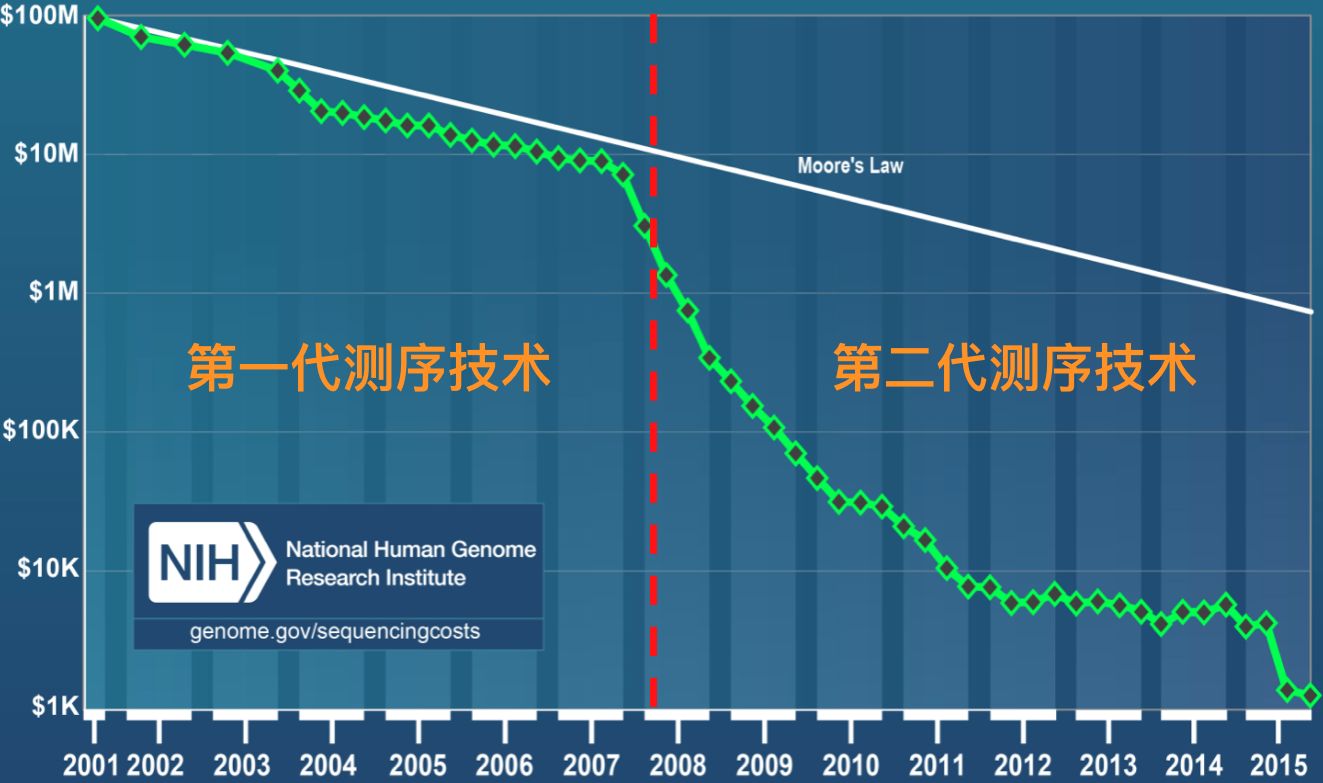


图 2 测序成本断崖式下降

从图二可以看出，自从第二代测序技术开始发展后，测序的成本开始断崖式下降，呈现“超摩尔定律”的现象，也就是性价比以比指数增长还快的速度上升。下面以illumina公司的技术为例简要介绍第二代测序技术的原理和特点。

1）构建DNA测序文库。将DNA分子用超声波打断为300bp-800dp的序列片段并在两端添加接头，构建出单链DNA文库。

2）将建库后的DNA片段吸附到测序流动槽（flowcell）中。测序流动槽是用于吸附流动DNA片段的容器，也是核心的测序反应容器。所有的测序过程都会发生在测序流动槽中。

3）桥式PCR扩增与变性。桥式PCR以flowcell表面所固定的序列为模板，进行桥形扩增。经过不断的扩增和变性循环，最终每个DNA片段都将在各自的位置上集中成束，每一个束都含有原来单个DNA模板的很多分拷贝，这一过程的目的在于实现将单一碱基的信号强度进行放大，以达到测序所需的信号要求。

4）测序。测序过程采用边合成边测序的方法。向反应体系中同时添加DNA聚合酶、接头引物和带有碱基特异荧光标记的4中dNTP（如同Sanger测序法）。这些dNTP的3’-OH被化学方法所保护，因而每次只能添加一个dNTP，这就确保了在测序过程中，一次只会被添加一个碱基。同时在dNTP被添加到合成链上后，所有未使用的游离dNTP和DNA聚合酶会被洗脱掉。接着，再加入激发荧光所需的缓冲液，用激光激发荧光信号（图7），并由光学设备完成荧光信号的记录，最后利用计算机分析将光学信号转化为测序碱基。这样荧光信号记录完成后，再加入化学试剂淬灭荧光信号并去除dNTP 3’-OH保护基团，以便能进行下一轮的测序反应。

Illumina的这种每次只添加一个dNTP的技术特点能够很好的地解决同聚物长度的准确测量问题。它的主要测序错误来源是碱基的替换，目前它的测序错误率在0.7%-1%左右——这是很高的精确度。测序周期以人类基因组重测序为例，30x-50x的测序深度对于Hisq系列来说需要3-5天时间，而对于2017年初最新推出的NovaSeq系列则只需要40个小时

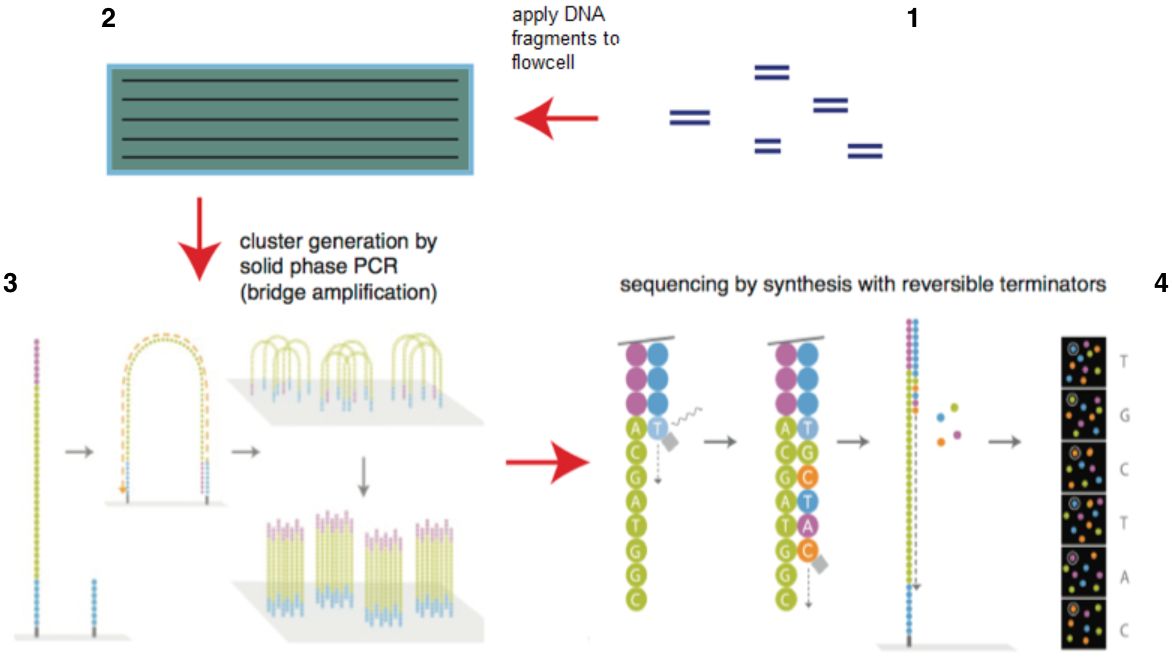
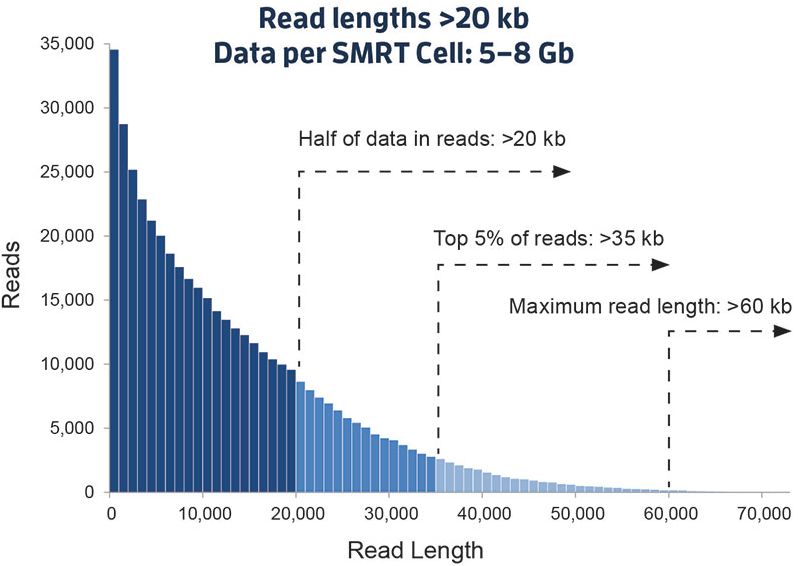


图 3 illumina测序原理

1.3 第三代测序技术

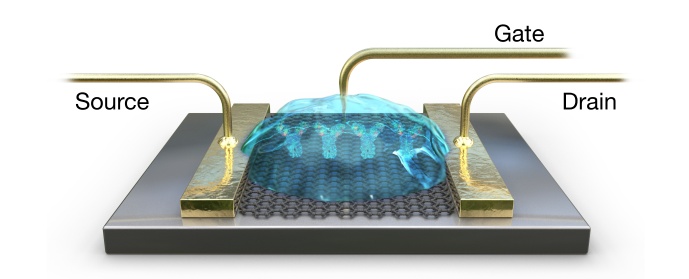
第三代测序技术是一个新的里程碑。第三代测序技术以PacBio公司的SMRT和Oxford Nanopore Technologies的纳米孔单分子测序技术为标志。与前两代相比，最大的特点就是单分子测序，测序过程无需进行PCR扩增，超长读长。下图是PacBio SMRT技术的测序读长分布情况，平均达到10Kb-15Kb，是二代测序技术的100倍以上。



2. 新方法简介

目前有很大一部分测序方法都是边合成边测序的，我们想到的这种方法也是采用了边合成边测序的测序方法。

在新方法中，我们使用了可以检测溶液中dNTP浓度的MOSFET传感器。在MOSFET源极与漏极中间的区域用化学方法固定特异性的某种dNTP抗体，形成一个有效的探针偶联区域。当dNTP分子在探针偶联区域上与探针发生识别效应并形成复合物，就会引起传感器表面的电荷浓度的改变，这就会改变源极与漏极之间材料的导电特性。因此，只需要知道导电特性改变的程度，就可以反映出探针偶联区域形成的复合物的浓度，进而可以反推出MOSFET所在环境中dNTP的浓度。



新方法的第一步依然是构建DNA文库。在构建好DNA文库后，将溶液倒入测试用的容器中。这个容器的底部贴有分别检测四种dNTP浓度的MOSFET传感器。

第二步是测序。向反应体系中同时添加DNA聚合酶、接头引物和4种dNTP（如同Sanger测序法）。这些dNTP的3’-OH被化学方法所保护，因而每次只能添加一个dNTP，这就确保了在测序过程中，一次只会被添加一个碱基。同时在dNTP被添加到合成链上后，所有未使用的游离dNTP和DNA聚合酶会被洗脱掉。在反应结束后，分别统计四种dNTP的浓度，有明显下降的那组就是反应到DNA链上的具体一位时，那一位所应有的dNTP种类。重复上述操作就可以进行整个DNA链的测试。

参考文献

【1】Sanger, F. & Nicklen, S. DNA sequencing with chain-terminating. 74, 5463–5467 (1977).

【2】Mardis, E. R. Next-generation DNA sequencing methods. Annual review of genomics and human genetics 9, 387–402 (2008).

【3】Shendure, J. & Ji, H. Next-generation DNA sequencing. Nature biotechnology 26, 1135–45 (2008).

【4】 Metzker, M. L. Sequencing technologies - the next generation. Nature reviews. Genetics 11, 31–46 (2010).

【5】 Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P. & Barron, A. E. Landscape of Next-Generation Sequencing Technologies. 4327–4341 (2011).

【6】 Rothberg, J. M. et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–52 (2011).