



评述

下一代测序技术: 技术回顾与展望

周晓光^{①*}, 任鲁风^①, 李运涛^②, 张猛^①, 俞育德^②, 于军^{①*}

① 中国科学院北京基因组研究所, 基因组科学与信息重点实验室, 北京 100029;

② 中国科学院半导体研究所, 北京 100083

* 联系人, E-mail: junyu@big.ac.cn; joezhou@big.ac.cn

收稿日期: 2009-12-08; 接受日期: 2009-12-18

中国科学院科研装备研制项目(批准号: YZ200823)资助

摘要 在过去的30年中, 作为最重要的生物医学研究手段之一, DNA测序技术的数据产出能力呈指数增长, 而且这一技术本身也演变成为一个面向工程学和物理学的新技术领域. 本文分析了下一代测序仪的技术特点, 并对其未来的发展以及应用进行了前瞻性的展望. 预期在刚刚出现的技术中, 有些技术假以时日能够发展成熟, 实现1000美元基因组和100美元基因组的目标. 同时建议中国科学家在这场对科学研究以及社会医疗保健体系都将产生深远影响的运动中发挥积极的作用.

关键词

基因组学

DNA测序

下一代测序技术
测序仪

DNA测序技术自发明以来就一直在推动分子生物学发展方面起着至关重要的作用^[1]. 从早期Frederick Sanger的手工测序, 以及基于Sanger法开发的第1代自动化测序仪, 到目前的下一代测序平台, 这一领域已经发生了巨大的变化^[2]. 有人甚至将基因组测序技术的发展与半导体技术的发展相提并论, 这也不无道理^[3]——在过去的数十年中, 每经过几年, 测序速度就呈指数增长, 与半导体工业发展的摩尔定律(Moore's law)非常相似^[4]. 这种高速的发展, 如图1所示, 从根本上改变了人们研究所有生命蓝图的方式. 并且推动了基因组学及其分支乃至其他密切相关学科的创立与发展, 诸如比较基因组学、生物信息学、系统生物学以及合成生物学. 某种程度上, DNA测序技术的进展已经使生命研究的基本元素发生了转变——从单一、局部的基因或基因的片段转变成整个基因组. 反过来, 这种转变又需要更加强大的测序技术来支持. 测序技术与其应用之间的协同关

系使得两者的发展在可预见的未来内保持这种趋势, 并且由于其对个体化疾病诊断与治疗具有确定性的推动作用而加速. 本文回顾了测序技术的演进, 分析

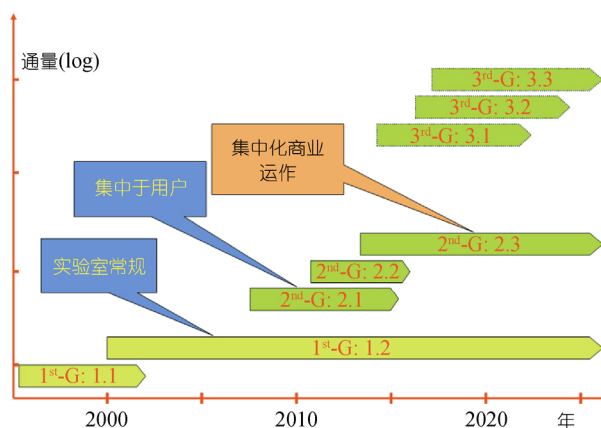


图1 测序技术发展时间轴

几代测序技术的优点与缺点, 并对这一领域可能的发展方向做出了预测. 为便于讨论, 根据技术的进展, 将测序技术分为具有不同亚代的 3 代(表 1). 尽管根据技术进展将其划为不同世代有些武断, 但是这种方法还是能够体现每个阶段的关键技术进展.

1 技术回顾与近期发展

1.1 第 1 代测序技术——荧光标记的 Sanger 法

在第一台全自动测序仪出现之前, 使用最为广泛的测序方法就是 Sanger 在 20 世纪 70 年代中期发明的末端终止法测序技术. Sanger 也因此获得 1980 年的诺贝尔化学奖^[5]. 他的发明第一次为科研人员开启了深入研究生命遗传密码的大门.

原来的方法主要依靠手工操作, 难以自动化. 例如, 它利用放射性同位素标记引物来进行 DNA 梯状成像, 操作十分不方便. 要使用双脱氧核苷酸分别做 4 个末端终止反应, 然后采用平板凝胶电泳技术, 用 4 条电泳道来分离 4 个反应所得产物, 费时费力, 试剂消耗也大, 这些都严重限制了测序的通量. 因此, 对于开发非放射性的第一代测序技术势在必行.

(1) G1.1. 最早版本的第 1 代测序仪是 20 世纪 80 年代中期在 Cal Tech 的 Leroy Hood 实验室发明的^[6]. 这一测序仪通过修改 Sanger 法得以实现. 最关键的改变是采用具有颜色的荧光染料代替同位素标记. 4 种双脱氧核苷酸终止子被标记上不同颜色的荧光基团. 另外, 与最初的 Sanger 法不同, 荧光基团是标记在终止子上, 而不是在引物上. 这种不同颜色标记的方案可以实现一个反应管中同时进行 4 个末端终止反应. 采用聚丙烯酰胺凝胶分离, 并通过计算机荧光检测系统分析梯状反应产物. 这些改进极大地提高了测序速度, 减少了测序过程中的人为干扰.

次年, 利用 Leroy Hood 实验室的技术, ABI 推出了第一款半自动 DNA 测序仪 ABI 370^[7]. 在随后的 20 年中, 测序仪的性能得到了极大的提升. 但基本工作原理直到最近才有所改变.

(2) G1.2. 第 1 代测序仪的第 2 个版本出现在 20 世纪末. 这一版本的测序仪, 其测序速度与质量得到了进一步的提高. 这主要归功于两方面的工作: 第一, 平板电泳分离技术被毛细管电泳所取代; 第二, 通过更高层次的并行化使得同时进行测序的样本数量增加. 使用毛细管替代平板凝胶取消了手工上样, 降低了试剂的消耗, 提升了分析的速度. 另外, 紧凑的毛

表 1 测序技术发展路线^{a)}

代次		第 1 代		第 2 代			第 3 代	
版本		1.1	1.2	2.1	2.2	2.3	3.1	3.2
平台	Sanger 法	ABI	ABI/GenoME MS					
	SBS			Illumina				
	SBL			ABI/Polonator G.007		Complete Genomics		
	SBP			Roche				
	FD					Helicos		
	SM-SBS					Pacific Biosciences/ VisiGen	?	
	FE							
	SM-SBL							
	SM-SBP							
	孔							PoC
纳米	刀							PoC
	石墨烯							PoC

a) SBS: 合成法测序(sequence-by-synthesis); SBL: 连接法测序(sequence-by-ligation); SBP: 焦磷酸测序(sequence-by-pyrosequencing); SM: 单分子(single molecule); FD: DNA 固定化(fixed DNA); FE: 酶固定化(fixed enzyme); PoC: 概念验证(proof-of-concept); ?: 预期可能出现的技术

细管电泳设备的形式更易于实现并行化, 可以获得更高的通量. ABI 3730 测序仪和 Amersham Mega-BACE 分别可以在一次运行中分析 96 个或 384 个样本. 这一代测序仪在人类基因组计划 DNA 测序的后期阶段起到了关键的作用, 加速了人类基因组计划的完成. 而且由于其在原始数据质量以及序列读长方面具有的优势, 这些测序仪今天还在使用之中.

通过几十年的逐步改进, 第 1 代测序仪的读长可以超过 1000 bp, 原始数据的准确率可以高达 99.999%, 测定每千碱基序列的成本是 0.5 美元, 每天的数据通量可以达到 600000 碱基. 不论这些数字如何令人印象深刻, 第 1 代测序技术在速度和成本方面都已达到了极限. 由于其对电泳分离技术的依赖, 使其难以进一步提升分析的速度和提高并行化程度, 并且难以通过微型化降低测序成本. 因此, 需要开发全新的技术来突破这些局限.

尽管如此, 第 1 代技术是不会很快消失, 它将与新的若干代测序平台并存. 这些久经考验的方法可靠、准确, 且已形成规模化, 特别是在 PCR 产物测序、质粒和细菌人工染色体的末端测序、以及 STR 基因分型方面, 将继续发挥重要作用.

1.2 第 2 代测序技术——循环阵列合成测序法

所谓下一代测序方法, 包括大量基于不同技术的方法. 尽管从模板文库制备、片段扩增到测序, 这

些方法所采用的技术与生物化学相当多样, 但是都采用了大规模矩阵结构的微阵列分析技术——阵列上的 DNA 样本可以被同时并行分析. 此外, 测序是利用 DNA 聚合酶^[8]或连接酶^[9]以及引物对模板进行一系列的延伸, 通过显微设备观察并记录连续测序循环中的光学信号实现的.

在一般性描述中, 下一代测序技术的几个关键特点是显而易见的: 第一, 通过有序或者无序的阵列配置可以实现大规模的并行化, 以提供高程度的信息密度. 理论上, 只有光的衍射极限会限制并行化的提高(即用来检测独立光学事件的半波长). 这极大地提高了总的测序数据产出通量; 第二, 不采用电泳, 设备易于微型化. 相对于第 1 代测序技术, 样本和试剂的消耗量得以降低.

(1) G2.1——下一代测序仪. 所有的下一代测序平台遵循了类似的工作流程, 如图 2 所示, 都要经过克隆扩增以加强测序过程中的光学检测灵敏度. 3 种广泛使用的商业化平台是 Illumina 的 Genome Analyzer, 罗氏 454 基因组测序仪以及 AB Life Technologies 的 SOLiD 系统. 它们基本都是在 20 世纪 90 年代末被发明和开发出来, 在 2005 年前后商业化. Polonator G.007 是最近刚刚实现商业化的新设备, 该仪器最初由哈佛大学 George Church 实验室开发, 现在由 Dover Systems 公司制造. Complete Genomics 公司最近推出了基于其专利技术的测序服务平台, 但该公

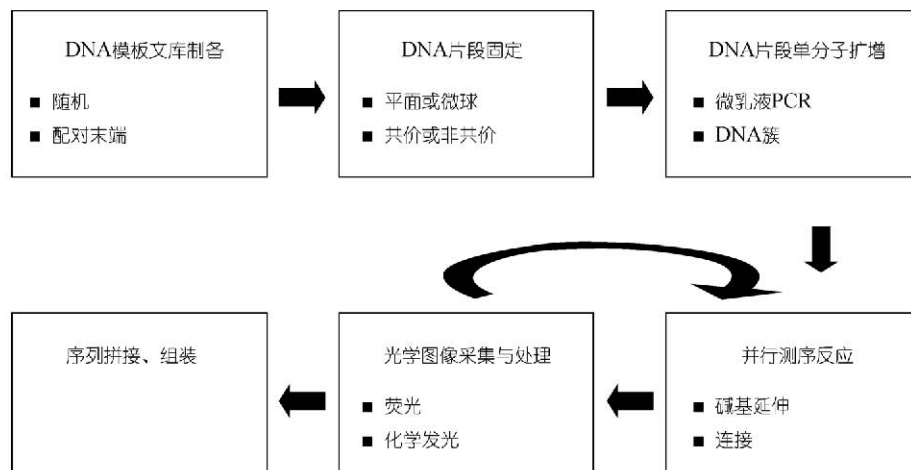


图 2 第 2 代测序技术工作流程

司并没有表示要在市场销售这一设备。这些仪器都采用了合成测序法, 只是在 DNA 阵列的排布、DNA 簇扩增, 以及基于酶的测序生化反应方面存在差异。

首先, 构建 DNA 模板文库。通过随机打断基因组 DNA 获得 DNA 文库片段(长度为数十到数百碱基), 或者构建控制距离分布的配对末端片段。在双链片段的两端连上接头序列, 然后变性得到单链模板文库, 并固定在固体表面上。固体表面可以是平面或是微球的表面。克隆的扩增通过以下几种方式之一进行, 如桥式 PCR^[10]、微乳滴 PCR^[11]或原位成簇^[12]。在芯片上形成 DNA 簇阵列的 DNA 簇或扩增微球, 利用聚合酶或者连接酶进行一系列循环的反应操作, 通过显微检测系统监控每个循环生化反应中产生的光学事件, 用 CCD 相机将图像采集并记录下来。对产生的阵列图像进行时序分析, 获得 DNA 片段的序列。然后按照一定的计算机算法将这些片段组装成更长的重叠群。

(i) Illumina Genome Analyzer. 单链文库片段的扩增是通过所谓“桥式扩增”过程实现的^[13]。单链 DNA 两端加上非对称的接头, 并利用两端的接头将片段固定在芯片表面形成寡核苷酸桥。芯片具有 8 个独立的道, 每条道的表面均可固定寡核苷酸。固定了寡核苷酸桥的芯片放置于流通池内。经过多个 PCR 热循环, 成千上万的复制产物制备出来, 每一簇都分别固定在芯片表面一个单一的物理位置上。流通池中芯片表面的 8 个道中, 每条道上都可以独立产生数百万这样的簇(这样, 可以在一个反应中对 8 个不同的文库并行测序)。然后, 测序引物杂交到扩增产物中的通用序列上, 开始测序反应。

Illumina 公司的测序仪采用合成测序法, 使用荧光标记的核苷酸以及可逆的终止子。在每一轮测序循环中, 标记不同荧光基团的 4 种核苷酸以及 DNA 聚合酶同时加入流通池通道中, 按照碱基互补配对的原则进行 DNA 链的延伸。每个核苷酸的 3'羟基是被封闭起来, 以防止额外的延伸。采集荧光图像, 碱基特异的荧光标记揭示了这一轮中新加入核苷酸是什么, 也就获得模板中这一位置的 DNA 序列。然后, 打开 3'端, 继续进行下一轮反应。这一过程重复多次, 到 50 个循环, 产生 50 个碱基的 DNA 序列。

该平台的测序通量是传统平台(第 1 代测序仪)的

数千倍。其主要的缺点是由于光信号衰减和移相的原因使得序列读长较短。由于要记录每个 DNA 簇的光学信号, 每一簇中所有 DNA 链的延伸保持同步至关重要。但是, 测序中每一步化学反应都可能失败, 例如不能将荧光标记物切掉, 或者未能去除封闭基团。这将导致一个簇中的一些 DNA 链过长, 而另一些 DNA 链可能没有同步延伸, 进而引起信号衰减或荧光信号相位移。此外, 错误率是累积的, 即 DNA 链越长, 错误率越高。这些都限制了读长的增加。

(ii) Roche 454 Genome Sequencer. 454 测序仪利用微乳滴 PCR(emulsion PCR, emPCR)来生成扩增产物^[14]。将固化引物的微球与单链 DNA 文库模板以及必要的 PCR 反应化合物一起混合, 微球与文库片段的比例适当, 以确保大多数微球结合的单链 DNA 分子不超过一个。水溶液与油混合形成油包水结构乳滴。每个乳滴都是一个进行后续 PCR 反应的微型化学反应器。经过多轮热循环, 每个微球表面都结合了数千个相同的 DNA 拷贝。然后富集微球, 转移并放置到刻有规则微孔阵列的微孔板上。每个微孔只能容纳一个微球。微孔板被安装成为流通池的一部分。其中一面可以通过测序反应的化合物, 另一面则与 CCD 光学检测系统的光纤部件相接触。

碱基测定采用边合成边测序, 利用焦磷酸法产生的光学信号来进行检测^[15]。通常所说的焦磷酸测序法是利用 ATP 硫酰化酶和荧光素酶。在三磷酸核苷结合到 DNA 链上的时候释放焦磷酸, 通过 ATP 硫酰化酶和荧光素酶产生一系列级联反应, 导致生物化学发光放出光信号。测序是顺次向流通池中加入 4 种 dNTP 中的一种。每个微孔之中有或是没有光信号释放出来分别表明 dNTP 连接到片段上或者不是互补的核苷酸, 这样也就确定了 DNA 模板上的互补碱基。

焦磷酸测序的主要优势是它的速度和读长——将近 500 碱基。与这里讨论的其他下一代测序技术不同, 除了 DNA 聚合酶反应所需化合物, 焦磷酸测序法并不需要额外的化合物用于 DNA 链的延长, 例如, 并不需要去掉标记基团或解除终止子的封闭, 这就降低了化学反应出现意外的几率。导致移相的主要原因如 DNA 链提前终止或者延伸不同步出现的几率较少。但是这种非同步的处理方式也赋予焦磷酸测序技术一个局限, 由于没有终止基团可以停止 DNA

链的延伸, 在测定同核苷酸聚合物区域时, 如一连串的 GGGGGG, 焦磷酸测序会遇到问题, 不得不依靠光信号的强度来推断同聚核苷酸的长度, 这就容易产生错误. 因此, 这一技术平台主要的错误类型就是插入-缺失, 而不是碱基的替换. 454 的另一个缺点是由于它依赖于包含一系列酶的焦磷酸检测, 与其他下一代测序技术相比, 其试剂价格相对较高.

(iii) Life Technologies SOLiD System. 与 454 的情况相同, SOLiD 系统也采用了微乳滴 PCR 与微球相结合的策略来扩增 DNA 模板. 打破微乳滴后, 扩增微球被收集、富集并固定在一个平的玻璃基板上形成一个无规则的阵列.

它的边合成边测序采用的是连接反应而不是上述平台所采用的聚合反应^[16]. 另外, 它采用了双碱基编码策略来协助检测错误. 一段与 DNA 文库模板连在微球上的接头序列互补的通用引物杂交到接头区域上, 然后进行一系列连接反应. 每个连接反应都发生在延伸链和有荧光标记的变性八核苷酸探针池中的探针(荧光标记在第 8 碱基上)间. 八核苷酸探针池设计成在核苷酸的碱基识别位点, 其第 1 和 2 位的碱基与特定的荧光颜色有明确的对应关系. 连接反应后, 获取荧光图像. 接下来, 在第 5 和 6 位碱基之间切开八核苷酸, 将最后 3 个碱基以及荧光基团去除. 接下来每一轮连接反应都可以获得延伸链上第 1 和 2 位碱基的信息(即模板序列的第 1~2, 6~7, 11~12, 16~17, 21~22, 26~27 以及第 31~32 位上的碱基). 7 轮连接反应后, 已经扩增的链变性脱落, 系统重置. 第 2 个引物结合到接头区域. 第 2 个引物在 DNA 模板上的起始位置与第 1 个引物相比提前一个碱基. 接下来有另外 7 轮上述连接循环. 这样就可以读取一组新的位置 0~1, 5~6, ……等的信息. 这一过程继续进行, 重置后所用的引物都比上一个引物提前一个碱基, 直到所有位置上的序列信息均被读取. 这种方法虽然听上去较复杂, 但实际上, 整个系统都是在计算机控制下自动运行. 由于每个碱基都被测定了两遍, 即在两个独立的连接反应中被测定, 这个方法使该测序技术具有可以确定错误识别碱基的优点^[16]. 该技术主要的缺点是序列读长相对较短. 这也是由于同一簇扩增产物中存在移相造成的.

(iv) Polonator G.007. Polonator 是另外一款使用

连接测序技术的下一代测序仪. 它采用的是单碱基探针, 而不是上述双碱基编码的策略. 测序是通过在结合到经微乳滴 PCR 扩增的 DNA 簇上的通用引物与九碱基探针之间的一系列连接反应进行的^[9]. 每次连接反应, 将一个九碱基探针池与 DNA 连接酶一起加入, 以进行引物-探针连接. 九碱基探针池中包括很多荧光标记的变性寡核苷酸探针. 荧光标记与每个读取位置对应(即荧光颜色与读取位置的碱基相对应). 每次连接之后获取荧光图像. 然后, 延伸的引物——探针链经变性进行系统重置. 接下来, 对下一个读取位置进行引物与第 2 个九碱基探针池之间的连接. 这一重置-连接-获取图像的过程重复进行, 直到所有位置的信息被读取.

这一系统中, 系统重置后, 并不需要进行连串的连接反应, 因此测序错误也不累积, 这是该系统的一个优点. 但是, 这就使引物间可能的读取位置受到限制, 读长更短. 这一缺陷在某种程度上可以通过在文库序列中使用多重锚定位置来扩展读取区间. 与其他商业化的下一代测序仪相比, Polonator 的价格明显低得多, 而且是一个开源的技术平台, 就是说它允许最终用户变更并且改进测序操作或化学试剂.

(v) Complete Genomics. Complete Genomics 所采用的连接测序方法基本与 Polonator 相同. 只不过它采用了独特的设计, 增加芯片表面 DNA 簇的密度, 以降低试剂的消耗^[17]. 为了增加读长, 在基因组片段两侧加上了多个(4 个)接头以形成环状 DNA 模板^[18]. 模板序列通过环形 PCR 扩增, 以获得包含源于模板序列的两百个拷贝的串联体. 这种串联体折叠成球状结构, 被称为 DNA 纳米球(DNB). 每个球自动聚集在水平基质表面的黏性(活化)点上, 形成高密度的纳米球阵列. 纳米球不会黏在芯片上活化点中间的区域, 这样就形成了一个规则排布的纳米球矩阵. 因为纳米球可以更加有效的利用三维空间, 与 DNA 簇或者微球相比, 纳米球可以形成密度更高的阵列. 这一技术还取消了在其他下一代测序仪中都使用的流通池^[19].

该测序技术采用的方法与 Polonator 相近, 即连接测序, 使用单碱基读取特定位置信息. Complete Genomics 创造了组合探针锚定连接(combinatorial probe-anchor ligation, cPAL)的方法. cPAL 的变性寡核

核苷酸探针库中的探针用荧光标记, 4 种不同颜色的荧光分别与给定位置上 4 种不同的碱基相对应。读取每一个位置, 都有一个单独的探针库。根据碱基互补配对原则, 一个特定的探针库与锚定序列在读取位置处连接起来, 通过荧光颜色对应读取相应的碱基信息。每次读取后, 探针与锚定序列复合体被洗去。另一个锚定序列与模板杂交, 针对另外一个位置的探针库加入循环。这一过程重复进行直到读取所有位置的信息。最近 Complete Genomics 通过测定 3 个人类基因组, 展示了其测序的准确度以及成本效益^[20]。

由于没有使用 AB SOLiD 系统的连续性探针, 使得该技术具有一些优点。首先, 没有记忆效应, 前面连接循环中产生的错误不会带到后面的循环中, 具有更高的容错能力; 另外, 每个循环中连接产量不高, 降低了如探针以及锚定序列等试剂的用量。但是, 由于寡核苷酸探针的长度限制(9 个碱基), 从每个锚定序列位置获得的读长依然很短。对每个文库片段, 使用 4 个锚定位置以获取全部序列。

与第 1 代测序仪相比, 以合成测序为基础的下一代测序平台速度显著提高, 成本明显降低。每台设备每天产出千兆碱基的序列不足为奇。但是, 除了罗氏的 454 平台之外, 读长短成了下一代测序平台的致命伤。这主要是由于 DNA 簇中存在的光学信号移相造成的。应运而生的单分子测序技术是解决这一问题的一种方案。

(2) G2.2——单分子测序(SMS)。为克服下一代测序平台读长较短这一主要缺点, 经过努力, 开发出了单分子测序平台, 通过在单一 DNA 分子组成的阵列上进行合成测序。在一个限定的表面上, 使用单个分子可以增加独立分析的 DNA 片段的数量。因此, 可以使数据产出量更高。当然, 这也意味着不再需要昂贵的 DNA 簇扩增步骤了。这将进一步降低测序的成本。但同时也带来了一些新的挑战, 主要是集中在单分子水平光学信号的检测方面。主要的问题是要降低非检测特异性的背景干扰, 例如没有参与到实际化学反应中的游离荧光分子。有几种不同的办法试图解决这一问题。基本的原则都是将检测局限在测序反应发生的实际位置附近, 例如使用消逝波。下面介绍一些已经开发或正在开发中的平台。

(i) Helicos HeliScope. Helicos Biosciences 公司开

发的 HeliScope 遗传分析系统是近来市场上最早出现的单分子测序仪器。它以 Quake 集团的工作为基础^[21], 在单个分子上边合成边测序。构建的单链 DNA 文库未经扩增, 没有规律地排列在平面基板上。每个测序循环中, DNA 聚合酶和 4 种荧光标记的核苷酸中的一种流入, 按照模板序列延伸 DNA 链, 阵列中发生了碱基延伸反应的 DNA 链就会发出荧光, 并通过 CCD 记录下来。经过洗涤, 延伸了的 DNA 链上的荧光物质被切除并被移走, 便可以进行下一轮单个碱基的延伸, 荧光标记的切除以及图像的获取。在焦磷酸测序中, 每个重复的循环是不同步的, 阵列中的一些链可能延伸到前边, 一些可能落在后面, 甚至根本没有延伸。而在这里, 每条链都是独立操作的, 根本不用考虑移相的问题。但是这也意味着, 焦磷酸测序法所遇到的同核苷酸寡聚物对 HeliScope 也是问题。与 454 不同, 单分子操作可以让人们通过动力学控制酶的反应, 降低 DNA 链延伸的速度, 在 dNTP 被洗掉前, 减少两个连续碱基连接在链上的可能^[22]。

如前所述, 检测是单分子测序技术所面临的一个关键挑战。HeliScope 利用了一项被称为全内反射显微镜(total internal reflection microscopy, TIRM)的技术。只有靠近流通池反应表面很薄的一层空间内的荧光集团才能被消逝波所激发产生荧光^[23]。这有助于降低荧光背景。尽管有精密的光学仪器, 捕捉单分子事件仍然是个挑战。因此, 与成簇检测为基础的测序仪先行者相比, 该平台的原始数据准确度明显较低, 主要的错误类型是缺失。但是, 双向测序策略可以显著提升准确率。单分子意味着可以在测序结束后, 去掉延伸链, 将模板重置为最初的状态。从而可以利用远端的接头, 从相反的方向进行另一次测序, 生成同一模板的第 2 个序列信息。重复的序列可以用于去除缺失错误, 因而, 相对于单向测序显著提高了准确率。

(ii) VisiGen. VisiGen 生物技术公司, 目前是 Life Technologies 的一部分。该公司一直致力于研发一种单分子的合成测序仪^[24]。

概括地说, 他们加工了一种蛋白质纳米装置来实时观察和记录 DNA 聚合酶合成 DNA 的过程。这是通过荧光供体和受体之间的荧光共振能量转移(fluorescence resonance energy transfer, FRET)实现的。

在荧光共振能量转移过程中, 只有在附近存在能量供体时, 受体分子才会在激发状态下发出荧光. 在他们的设计中, 每一个 dNTP 分子的 γ 磷酸上携带一个具有特定颜色的荧光受体基团, 而 DNA 聚合酶经过修饰, 在其活性位点附近携带一个荧光供体基团. 在 DNA 碱基延伸的时候, 一个与模板配对的碱基被 DNA 聚合酶捕获, 并使荧光受体基团靠近供体基团. 荧光能量转移发生, 发出相应颜色的荧光. 一旦结束, 荧光基团作为焦磷酸的一部分被 DNA 聚合酶释放. 这样在本质上当核苷酸连接起来的同时协同地产生了一个荧光的爆发. 通过记录和分析时序的荧光爆发就构建了 DNA 序列信息. 值得注意的是在去除荧光标记或是切除封闭基团时, 整个过程是没有停顿的, 这确实是一个实时的过程, 意味着其工作速度极快, 要考虑到光学记录装置能否跟上这一速度. 为了进一步降低背景的干扰, 他们也采用全内反射显微镜作为荧光检测设备. 而且, 与 Helicos 的平台不同, 这套系统将 DNA 聚合酶固定在基质表面上, 而不是固定 DNA. 这样 DNA 链的延长就不受限制. 固定酶而不是 DNA 的另一个好处是当 DNA 延长时, 可以将核苷酸连接控制在很小的检测空间范围内, 即 DNA 链增长时荧光不会超出检测范围. 尽管 Life Technologies 公司没有公布仪器的性能, 但估计它的速度很快, 且读长较长. 理论上, 读长只是由 DNA 聚合酶的处理特性决定的. 实际上, 其他因素, 如 DNA 聚合酶上携带的荧光供体的光漂白作用也会限制读长. 据报道, Life Technologies 公司正在开发一种量子点荧光标记以解决这一问题.

(iii) Pacific Biosciences. Pacific Biosciences 公司致力于开发一种新的测序技术——单分子实时技术 (single molecule real time, SMRT)^[25]. 这一单分子合成测序技术依赖于被称为零级波导 (zero mode waveguide, ZMW) 的纳米结构来实时观察 DNA 的聚合^[26]. 该结构在一片薄金属膜上蚀刻出数以千计直径数十纳米的亚波长小孔. 并将金属膜附着在透明的支持基质上. 由于每个小孔的尺寸低于光的波长, 所以当光线从透明一侧照射时无法透射. 并且在每个小孔的底部形成指数衰减的消逝波, 这样创造了一个很小体积的检测空间. 每个小孔底部固定一个聚合酶分子. 在测序过程中, 由固定的酶根据单链

DNA 模板合成双链. 每次加入一个碱基, 聚合酶捕获具有荧光标记的 dNTP (也是标记在 γ -磷酸上), 并将其带到检测区间, 产生荧光光曝. 光曝的荧光颜色就揭示了模板上的互补碱基. 通过连续实时的监控每个波导孔的荧光光曝, 就快速测定了每一个孔内 DNA 模板的序列.

PacificBio 的技术在高速测序、长序列产出和低成本方面有着巨大的潜力. 但是, 与其他单分子测序平台一样, 实时检测单个分子对于提高原始数据准确率是一个挑战, 并可能成为一个障碍. 如前所述, 可以通过对同一样品进行重置后进行多次测序来减少误读. 另外, 当前的 CCD 技术也限制了同时观察零级波导的最大面积. 聚合酶所占据的零级波导比例较低 (约 30%) 也限制了可用的波导数量^[27]. 所有这些都成为提高 SMRT 技术数据通量的限制. 尽管有这些限制, 预计在 2010 年推出的第一个版本的设备的序列读长将不低于 1500 个碱基, 一次测序只要 15 min, 每次测序的实际成本不超过 60 美元. 当这些技术问题都被解决的时候, 预计未来版本的设备每天可能产出 100G 的数据, 而序列读长将达到 10 万碱基.

(iv) Mobious Nexus I. 除了宣布他们打算开发一款单分子测序仪之外, 到目前为止, Mobious Biosystems 公司并没有透露关于其 Polykinetic Sequencing 技术的很多细节^[28]. 该技术利用了聚合酶在 DNA 合成时的天然化学方式. 当聚合酶按照模板, 将碱基连接到 DNA 链上时, 它首先需要测试溶液中某一碱基以确定其是否与模板碱基相匹配. 如果不匹配, 则马上将该碱基释放. 如果匹配, 聚合酶将抓住它, 并继续进行耗时的步骤, 将核苷酸连接到 DNA 链上. Mobious 的合成测序方法就是利用匹配与不匹配碱基在这一步骤所需时间的差异进行检测. 按照顺序, 每次在反应体系中加入 4 种核苷酸中的一种, 通过测定 DNA 聚合酶 (固定在基质表面上) 抓住核苷酸以及完成聚合的时间, 可以推断出该核苷酸与模板是否匹配. 检测聚合酶的构象变化就可以来记录这一过程中的时间差异. VisiGen 所采用的使用配对的供体和受体的荧光共振能量转移策略也可用于这一检测. 荧光的主要问题是荧光基团的光漂白. 为解决这一问题, Mobious 利用酶的构象改变时电磁性质发生变化的特性, 通过离子体共振光谱, 核磁共振等进行检

测^[29]。除去单分子测序本身的原因, 这种方法所带来的一个优点是由于没有使用荧光标记核苷酸, 试剂成本降低的空间更大。

1.3 第3代测序技术——直接测序

在所有上述第2代测序技术中, 序列都是在荧光或者化学发光物质的协助下, 通过读取 DNA 聚合酶或 DNA 连接酶将碱基连接到 DNA 链上过程中释放出的光学信号而间接确定的。除了需要昂贵的光学监测系统, 还要记录、存储并分析大量的光学图像。这都使仪器的复杂性和成本增加。依赖生物化学反应读取碱基序列更增加了试剂、耗材的使用, 在目前测序成本中比例相当大。直接读取序列信息, 不使用化学试剂, 对于进一步降低测序成本是非常可取的。在一个正在发生突破瓶颈巨变的领域内, 很难准确预测未来将发生什么。但是, 最近几个领域大量的研究工作表明未来新一代的测序平台将在其中产生。

(1) 非光学显微镜成像。常言道“百闻不如一见”, 确定 DNA 序列的最直接方法之一就是将其核苷酸(主要是碱基)的空间线性排列方式可视化。如果一个 DNA 链的图片具有足够高的分辨率, 可以将 DNA 链上的 4 种碱基区分开来, 那么序列将非常容易被读出。这正是目前显微镜领域科研人员所努力实现的。这一设想是借助具有原子水平分辨率的非光学显微镜, 如扫描隧道显微镜(STM)、原子力显微镜(AFM)等^[30,31]。尽管非常有限, 但还是有一定的进展。最近, 日本大阪大学的研究小组发现, 在一条拉长的 DNA 链上, 利用扫描隧道显微镜的图像, 根据独特的电子指纹, 可以将鸟嘌呤与其他 3 种碱基区分开来^[32]。另外一些团队一直积极地致力于使用原子力显微镜来记录将每个碱基紧密连接的环拉开所需力的大小, 以区分不同的碱基^[33]。而 ZS Genetics 公司致力于使用电子显微镜直接测序。为了解决天然 DNA 分子在电子显微镜下对比度不足的问题, 他们在聚合酶合成新 DNA 链时加入更重的元素^[34]。在电子显微镜下观察就可以获得的更重的 DNA 并确定其序列。

(2) 纳米孔。本研究注意到另外一个非常活跃的领域是利用纳米孔结构进行测序。顾名思义, 纳米孔就是直径在纳米尺度的小孔(1~2 nm)。通常是利用固态物质或者生物分子制成的小孔。这种想法是在电

场驱动下, 当线状 DNA 分子通过小孔时, 通过一些物理手段来确定碱基的序列。在全球, 许多公司和组织, 如 Agilent, DNA Electronics, IBM, NabSys, Oxford Nanopore Technologies, Sequenom 等都在进行纳米孔测序的开发, 但采用的方法不同。

所有以纳米孔为基础的测序技术都面临两个关键的挑战^[35]: (i) 区分 4 种核苷酸的速度要与 DNA 运动的速度相称; (ii) 控制 DNA 通过纳米孔的速度。初步尝试测量离子电流的波动——单链 DNA 分子通过纳米孔时, 堵塞纳米孔而造成电压的波动。但迄今为止, 成效甚微。计算和实验表明, 仅仅测量通过纳米孔离子电导率是不可能提供识别 DNA 分子中每个核苷酸所需要的分辨率^[36]。纳米孔通道长度通常为 5 nm, 可以容纳十多个碱基, 这一尺寸对于测序所需要的分辨单个碱基引起的电流变化过长。

尽管离子电流测量还不能区分单个碱基, 但其可以很容易地辨别单链与双链 DNA^[37]。NABsys 公司与 Brown 大学的一个团队合作, 利用这一性能来开发一种杂交测序法——杂交辅助的纳米孔测序方法(hybridization assisted nanopore sequencing, HANS)^[38]。将基因组 DNA 随机切割成大约 100 kb 左右的片段, 制成单链并与六寡聚核苷酸探针杂交。然后驱动结合了探针的基因组文库片段通过可寻址的纳米孔阵列。通过每个孔的离子电流均可独立测量。追踪电流的变化确定探针杂交在每个基因组片段上的精确位置。利用基因组片段上杂交探针的重叠区域将基因组片段文库排列起来, 建立一组完整的基因组探针图。利用计算机算法, 获得完整的基因组序列。但确定杂交位置的精确度和一致性还需要进一步验证。

为了提高纳米孔的灵敏度以确定不同的碱基, 科研人员也在尝试其他方法, 包括在纳米孔内嵌入电子探针^[39]。他们期望仅靠纳米孔两侧的隧道电极在每个碱基被驱动通过纳米孔时能够记录特征性隧道电流。计算机模拟以及成功用于揭示原子水平特性的扫描隧道显微镜的经验都让本研究组对这一工作持乐观态度。但制作这种纳米尺度的装置并不简单。对这一问题, 另一个创造性的解决方案是发展化学功能的纳米孔, 而不是嵌入固态电极。Lindsa 及其同事提出了在核酸通过纳米孔时, 使用两种化学探针分别与磷酸基团和碱基基团形成氢键的设想。与

磷酸基团作用的探针起到捕获核酸的作用, 而与碱基作用的探针起到识别的作用^[40]. 针对 4 种不同的核苷酸, 需要 4 种不同的探针.

另外一些研究团队正在开发一种固体纳米孔——生物孔(Biopore). 其中一个团队利用改造了 MspA 蛋白质构建了生物纳米孔, 用以分析单链 DNA 分子^[41]. 他们证实单链 DNA 分子可以通过这个生物孔. Oxford Nanopore Technologies 与牛津大学合作, 正设计另外一种基因工程蛋白质纳米孔. 通过遗传工程改造, 他们已经可以构建一种将氨基化环糊精(Aminocyclodextrin)配体共价连接到位于脂双层膜中的 α -溶血蛋白内生物纳米孔^[42]. 最近, 他们证实驱动 4 种核苷酸单磷酸(dNMPs)通过生物孔, 通过纳米孔的电流将分别减少到 4 种不同的状态, 每种状态都与一种核苷酸单磷酸相对应^[43]. 将这一机制与外切酶将核苷酸从 DNA 链上切除并释放出来相结合, 提供了另外一种纳米孔测序技术. 为了做到这一点, 至关重要是外切酶的固定方式, 以确保切除下来的核苷酸单磷酸能被严格单一地运送并通过纳米孔.

除了碱基的检测, 控制 DNA 的运动以及通过纳米孔的速度也是非常重要和具有挑战性的. DNA 高速移动通过纳米孔具备了开发超高速测序方法的可能. 但是, 如果 DNA 链通过孔隙速度过快, 用来确定每个碱基的时间就非常短. 在 DNA 随机运动以及 DNA 分子与纳米孔表面非特异作用的情况下, 这种情况还可能更加严重^[44]. 所有这些增加了 DNA 分子转位通过纳米孔速度的不确定性. 尽管可以通过降低温度、增加溶液黏度、降低纳米孔的偏好性等降低 DNA 通过的速度, 但速度的变化还是一个问题. 对于克服这一困难有种种设想. 其中之一是想法是将某种类型的进行性酶(Processive Enzyme)加入与通过的 DNA 链结合^[45]. 这将显著降低移动的速度. 最近, IBM 宣布了其纳米孔装置, 被称为 DNA 晶体管^[46], 将纳米孔嵌入金属层中, 形成可以通过调控将 DNA 分子捕获在纳米孔内的金属介质结构. 计算机模拟表明, 经过周期性的门电位的开关, 每次将一个碱基通过纳米孔是可行的, 这就给探测通过的核苷酸留出了充足的时间.

(3) 石墨烯和碳纳米管. 石墨烯是由碳原子构成的二维晶体, 碳原子排列与石墨的单原子层一样. 非

常稳定并具有非常良好的导电性, 非常适合制作核酸测序用电极的材料. 有种设想是在石墨烯上打出一个大约 1 nm 宽的缝隙^[47]. 引导 DNA 分子垂直通过缝隙. 当 DNA 通过时, 缝隙两边的石墨烯边缘可以作为电极来确定核酸的序列. 除了在石墨烯上打出这样一个缝隙是一种挑战之外, 控制 DNA 的运动、移动的方向以及通过缝隙的速度同样是不小的障碍.

由于其独特的电物理特性及其纳米尺度的结构, 虽然到目前为止还没有开发出任何装置, 但是碳纳米管(CNT)在高速 DNA 测序中已经表现出了巨大的潜力. 已经有工作表明, 碳纳米管表面同 DNA 分子之间可以高度的互相作用, 甚至是与序列特异性相关^[48]. 长的基因组单链 DNA 可以缠绕在一条单壁碳纳米管上, 形成一个稳定的 DNA-碳纳米管复合物^[49]. 计算机模拟表明, 引入的 4 种核苷酸表现出独特的局部密度^[50]. 这是碳纳米管可以单独或者与其他技术整合作为电测序的候选材料. 这些设想还停留在理论验证阶段.

2 展望

回顾了 30 多年来测序技术的进展以及可能产生的突破, 难免会产生这样的问题——未来几年内将发生什么? 对于测序技术的参数有什么样的展望? 千美元人基因组(TDG)和百美元人基因组(HDG)技术将在何处产生? 什么时候能产生? 千美元人基因组以及百美元人基因组这两个目标是要实现以现有序列为参照, 将测定一个人基因组的全部成本控制在 1000 美元和 100 美元之内. 本节将讨论这些问题.

2.1 技术融合

现在不能确定哪种改变游戏规则的想法或革命性的技术最终能实现千美元人基因组甚至是百美元人基因组的目标, 并将人们带到充满希望的基因组研究新领域. 回顾 3 代测序平台的技术进展, 发现有一点是非常明确的, 即固态技术与生物化学的联姻. 另外, 技术的融合将从生物化学或化学手段向物理手段发展(表 2). 相信这一趋势将持续下去. 下一代测序仪将可能根本不使用生物化学方法, 而纳米技术将可能发挥更大的作用.

表 2 测序技术的发展依赖于学科含量交叉增加并向微纳技术倾斜

关键技术	第 1 代	2.1~2.2 代	2.3 代	第 3 代
DNA 杂交		√	√	
DNA 聚合酶法测序	√	√	√	
PCR 扩增	√	√		
电泳	√			
光电子	√	√	√	
微流体		√	√	
微纳米工艺		√	√	√
单分子检测			√	√

2.2 测序通量与序列读长

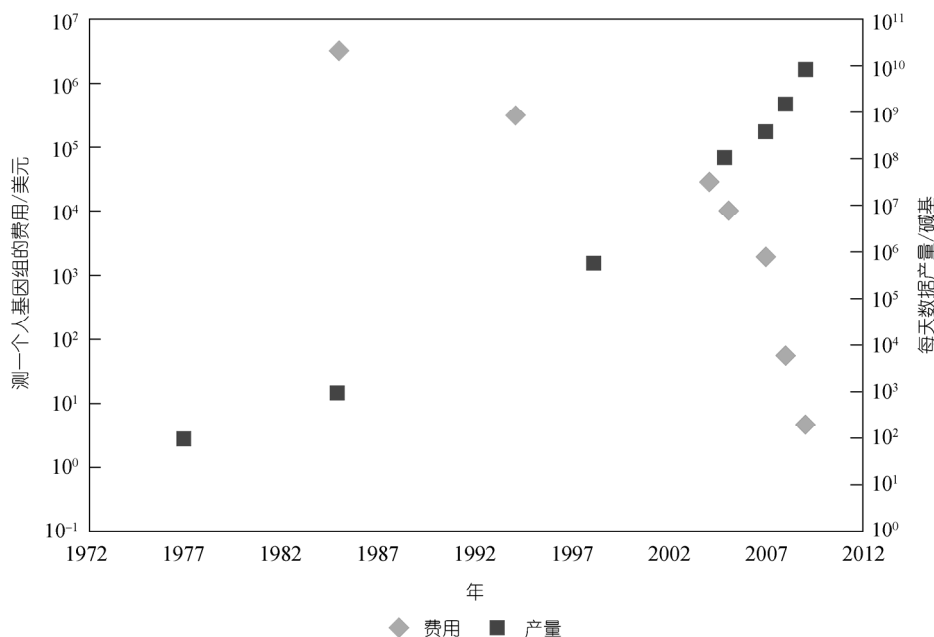
测序通量和数据读长是测序技术相对立的两个选择。从第一代到下一代测序技术, 可以看到测序通量显著改善, 但读长也显著缩短。正如前面章节所述, 这主要是由于同一 DNA 簇中, DNA 链的化合物延伸不同步所致。用户只能在低通量且具有较长的读长(如第一代测序仪)与高通量且具有较短的读长(如下一代测序仪)的测序平台之间进行取舍。这种情况在单分子测序中将得到改变。例如 Pacific Biosciences 的零基波导技术, 能同时实现超高的通量与较长的

读长。假如没有光学检测能力的限制, 通量和读长将仅受到 DNA 聚合酶的合成速度和连续性的限制。未来一代的基于纳米孔的测序技术, 当线性 DNA 通过纳米孔结构, 一个接一个的核苷酸被确定下来时, 也可以同时实现高通量和较长读长。原则上, 一个纳米孔结构单次测序读长仅仅受到可能通过纳米孔结构未经剪切的非常长的线状 DNA 链的长度限制。已经证实长达 5.4 kb 的线状 DNA 可以通过固态的纳米孔^[37]。因此, 预计新一代的测序仪在具有超高通量的同时, 其读长也将轻易超过 Sanger 设备。

2.3 测序成本与生产能力——经验曲线

过去 30 年中, 在测序成本急剧下降的同时, 测序通量(生产能力)呈指数提高。前面提过, 有人将这种巨变与 IT 工业相提并论。实际上在某些方面, DNA 测序的改变速度已经远远超过了半导体工业的摩尔定律的描述。图 3 以对数坐标显示了过去数年中测序成本(以美元每人类基因组为单位)以及生产能力(以每天每台仪器核苷酸数目为单位)随时间的变化。

观察这些曲线可以发现, 正如预期, 在过去 30 多年, 测序成本的降低以及通量的提高均呈指数变化。仔细观察可见, 另一个有趣的现象: 两条曲线的

**图 3 测序费用与产量的发展变化**

斜率都在 2005 年左右出现拐点. 这意味着从那时开始, 测序成本降低和通量提高的改变速度从那时开始加快. 这正是下一代测序仪开始进入实验室的时间. 拐点两侧的斜率都接近线性, 表明两个阶段有两种不同的因素驱动测序生产能力的提高: 一种是世代内的技术改进提高了生产能力并降低了成本, 而另一种是通过技术突破.

生产能力和成本的关系, 可以描述为经验曲线(又名学习曲线)效应的模型. 这种效应表明, 随着更多得执行某项任务, 任务的成本也随之降低. 这一模型在 20 世纪 30 年代中期, 由 Theodore Wright 最早量化, 并成功应用于提高飞机产量和降低成本的项目中^[51]. 从那时起, 该模型已在许多行业被用来研究生产能力和成本的关系. 本研究以测序成本与测序通量为坐标(在对数坐标中)绘制了从 2005 年以来两者随时间的变化, 获得了第 2 代测序技术的经验曲线(图 4). 近乎线性的曲线呈现了典型的经验曲线效应, 当生产能力增加一定倍数的同时, 成本也按照一定的百分比降低. 线性的经验曲线偶尔可能突然中断. 这种中断反映了技术过时或者已被更新的技术所取代的过程. 正如前面所讨论, 从 2005 年前后第 2 和第 2 代测序技术变化中所观察到的现象. 自 2005 年以来, 一个数量级的测序通量变化与 1.8 个数量级的测序成本降低相对应.

2.4 千美元基因组(TDG)和百美元基因组(HDG)

由图 4 的经验曲线推测可得, 第 2 代测序技术(包括目前正在开发的单分子测序技术)中, 千美元基因组与百美元基因组的目标将分别在每天每台测序仪测序通量达到 20 G 与 70 G 时实现. 如果这一趋势保持下去, 千美元基因组可能相当快, 在 1~2 年内实

现. 基于相同的模型, 百美元基因组可能在 2~3 年内实现.

当然, 这些预测都是基于本研究的经验曲线模型. 同时存在一些经验曲线所无法预测的其他因素, 例如革命性技术突破或缺乏突破. 如果现有的技术在达到目标之前就宣告失败, 那么所有的预测都将是错误的. 必须建立新的替代技术, 那时目前用来预测的模型将过时. 每个人都在猜测, 这要多久才能实现? 但有一点已明确, 即答案将在 1~2 年内揭晓. 基于现在经验曲线的轨迹, 我们预计千美元基因组在 1~2 年内, 可能依靠基于循环阵列合成测序的第 2 代测序技术实现. 但是对于百美元基因组, 现有的技术可能难以达到这一目标, 即便是正在开发中的单分子测序技术. 若是这样, 将不得不等到更新一代的革命性技术出现, 如第 3 代测序技术.

2.5 客户的测序成本

真能这么快实现目标吗? 真的即将看到 1000 美元客户成本的人类基因组测序吗? 没有那么快. 目前所有用于测序成本的分析的数据基本上都是耗材与试剂的成本. 即使是对同一设备, 人们估算的成本范围之大也令人惊讶^[52]. 经常被低估的是下游信息分析的成本. 分析测序仪产生的数据, 精确注释人类基因组是耗时和人力密集型的. 很多时候, 在成本估计或声明时, 人们都没有将这部分成本考虑在内. 如果将所有的因素都考虑在内, 例如耗材、设备摊销、维护、人工以及计算分析, 产生高质量人类基因组数据的实际成本要比那些数字高出许多倍.

除了测序的实际成本, 还必须考虑市场的力量, 包括企业的经营成本. 也就是说即使测序成本降到 1000 美元左右, 在市场服务的初期阶段, 其价格也会远远高于此——这是简单的市场供求关系所决定的. 综合考虑这些因素, 意味着可能很难在不久的将来看到真正意义的 1000 美元基因组.

目前, 真实的成本结构中试剂成本高是人为造成的. 因为构建文库与测序反应试剂的价格完全由仪器公司控制. 只有具有实力的竞争者进入市场, 打破对试剂的垄断, 才可能发生有利于客户的改变. 因此, 目前用户所要做的是将所要测序的基因组排定优先级别, 从最需要完成的开始, 而不是, 马上开始所有的测序.

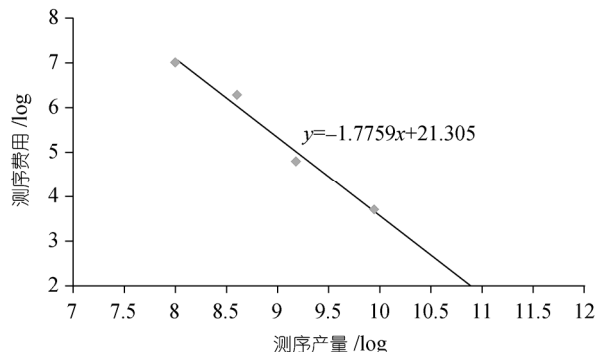


图 4 第 2 代测序技术费用和产量发展经验曲线

2.6 测序运作模式

测序成本方面, 另一个需要考虑的因素是规模经济效应. 从某种意义上说, 这一因素已经包括在经验曲线之内, 做的越多越便宜. 这引起另一个有趣的问题: 测序业务将朝着哪个方向发展? 是集成的生产型中心还是分散的进行测序? 预计市场需求将使运作模式多样化.

存在两种测序运营模式: 作为常规实验室技术或高通量的操作. 对于毛细管电泳测序仪 ABI-3730XL, 为测定有限数目的样本或特殊任务所必需, 将按照常规实验室技术的方式运作. 可以预见, 目前的第二代测序仪的运作将集中于缝隙市场(niche market), 例如市场上的 Roche/454 以及 Illumina GA 测序仪, 他们将用于小规模的操作, 满足单独实验室一次或多次运行来分别测定一些细菌或是人类基因组的需求. 高通量的机器将测序能力提升至 20 倍覆盖率人类单倍体基因组数据的水平, 相当于每次运行产生 100~150 G 的数据.

相信未来大规模测序业务将转移到大的商业化服务中心或服务商, 而不是在单独的实验室或小的研究机构中进行. 如果现在的预期都能变成现实, 当测序成本下降到 1000 美元每基因组水平之下, 那将很难说花费 50 万美元购买设备是正确的(假设设备的价格将与现有设备持平——基于经验的合理假设).

要迅速收回仪器成本, 并达到盈亏平衡点, 这需要承接大量的测序工作. 运作良好的服务导向型测序服务商拥有大量的客户资源, 其更具有财务优势. 简单的说, 这就是规模经济. 这一点在某种程度上类似与 IT 行业相似, 大型 Web 主机托管服务商承担了众多大小机构的网站托管业务. 基因组测序服务商已经萌芽, 如 Agencourt Bioscience, Cofactor Genomics, Complete Genomics, Knome, SeqWright 等, 以上只是几个例子. 这一趋势将持续下去, 但随着市场的成熟, 这些公司最终将合并成几个较大的服务商.

2.7 技术共存

不同世代测序技术并不互相排斥, 新一代技术出现并不意味着原有的平台完全过时. 由于不同世代技术之间功能上的互补性, 它们将长期共存(表 3). 典型的例子是第 2 代测序平台的测序通量远高于基于 Sanger 法的第 1 代测序平台, 但依然没有完全取代第 1 代测序技术. 具有读长优势以及原始数据准确的 Sanger 法将用于小规模测序的缝隙市场, 而第 2 代测序仪将用于大规模的项目.

3 中国应该怎样做?

发达国家, 如美国和欧盟, 在新的测序技术开发方面进行着激烈的竞争, 这也标志着测序技术的重

表 3 不同世代测序技术的应用领域^{a)}

技术		特征	功能性					
			PCR 产物 测序	新测序		重测序	GT	1000/100
				WGS	CBC			
第 1 代	1.1	平板胶	+	+	+	+	+	NA
	1.2	毛细管-四色荧光	+++	++	++	++	++	NA
第 2 代	2.1	LR	NA	+++	NA	++	NA	NA
		SR	NA	++	NA	+++	NA	NA
	2.2	高通量/无反应仓	NA	+++	NA	+++	NA	NA
	2.3	单分子	NA	+++	NA	+++	+	1000
第 3 代	3.1	化学/纳米技术	NA	+++	NA	+++	+	1000
	3.2	纳米技术	NA	+++	NA	+++	+	100
	3.3	纳米技术	NA	+++	NA	+++	+	100

a) WGS: 全基因组鸟枪测序(whole genome shotgun); CBC: 克隆测序(clone by clone); GT: 基因型分型(genotyping); 1000/100: 千/百美元人基因组测序目标; LR: 长读长(long reads); SR: 短读长(short reads)

要性. 对于中国, 发展这项复杂的技术仍然是巨大的挑战, 但同时也是一个巨大的机遇.

传统上, 中国缺少财务支持以及专有技术(Know-how)来开发例如新一代测序仪之类的复杂分析仪器. 迄今为止, 中国开发的分析仪器局限于实验室基本设备如离心机、摇床以及电泳设备等. 高端设备, 如高通量测序仪等, 都依赖进口. 考虑到过去中国的经济发展与技术状态, 这些都是可以理解的.

要开发这类高级分析技术平台, 不仅需要技术、工程基础、合作以及跨学科的整合, 例如生物化学、半导体、电子工程、机械工程、计算机等, 而且还需要巨大的公共或私人投资机构提供强大的金融支持, 但潜在的利益是巨大的.

首先, 通过开发这类分析仪器, 中国的科学家与工程师不仅可以在解密生命密码方面做出贡献, 而且可以学习并掌握这方面的技术与组织技能; 其次, 除了推进技术、工程以及科学 Know-how 的发展, 还有巨大的经济价值. 先进的基因组测序技术打开了个体化医疗的大门, 这样做不仅可以带来巨大的经济利益, 而且有利于医疗条件的改善与提升. 假定只

有 10% 的中国人选择用 1000 美元来测定他们的基因组, 这将创造 1300 多亿美元的价值. 因此认为, 尽管这一前期开发的投入较大, 但对中国来说, 如果不做, 潜在的损失是无法估量的.

4 结论

30 年的创新和发展开创了基因组测序的新时代. 测序技术从手工的一次一个样品发展成为基于大规模阵列的高度自动化技术. 测序通量呈指数提高, 而成本急剧降低. 由于当前的以及即将出现的更新一代测序技术, 1000 美元基因组的目标将变得更加现实. 快速、廉价的测序能力将引领我们开辟比较基因组学分析、疾病诊断以及个性化(个体化)医疗等新领域. 由于其在科研以及经济方面的巨大利益, 中国应该在 21 世纪这一关键技术的研究开发方面发挥更大的作用, 并积极促进其在生命科学研究以及医药领域的广泛应用. 当全部现存生物的基因组及其有意义的变异都被发现并掌握时, 做出这些贡献的人应该感到无比的自豪!

参考文献

- 1 Gilbert W. DNA sequencing and gene structure. In: Forsén S. Nobel Lectures in Chemistry 1971-1980. 1980. Singapore: World Scientific Publishing Co, 1993. 408—426
- 2 Sanger F, Coulson A R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol, 1975, 94: 441—448
- 3 Shendure J, Mitra R, Varma C, et al. Advanced sequencing technologies: methods and goals. Nat Rev Genet, 2004, 5: 335—344
- 4 Moore G E. Cramming more components onto integrated circuits. Electronics, 1965, 38: 114—117
- 5 Sanger F. Sequences, sequences, and sequences. Annu Rev Biochem, 1988, 57: 1—28
- 6 Smith L M, Fung S, Hunkapiller M W, et al. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res, 1985, 13: 2399—2412
- 7 Applied Biosystems Timeline. <http://www.appliedBiosystems.com>
- 8 Mitra R D, Shendure J, Olejnik J, et al. Fluorescent in situ sequencing on polymerase colonies. Anal Biochem, 2003, 320: 55—65
- 9 Shendure J, Porreca G J, Reppas N B, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science, 2005, 309: 1728—1732
- 10 Adessi C, Matton G, Ayala G, et al. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res, 2000, 28: e87
- 11 Dressman D, Yan H, Traverso G, et al. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci USA, 2003, 100: 8817—8822
- 12 Mitra R D, Church G M. *In situ* localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res,

- 1999, 27: e34
- 13 Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*, 2006, 34: e22
 - 14 Margulies M, Egholm M, Altman W E, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437: 376—380
 - 15 Ronaghi M, Karamohamed S, Pettersson B, et al. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 1996, 242: 84—89
 - 16 Macevitz S C. DNA sequencing by parallel oligonucleotide extensions. US patent 5750341, 1998
 - 17 Complete Genomics Technology Paper. <http://www.completegenomics.com>
 - 18 Dahl F, Drmanac R, Sparks A. Methods and oligonucleotide designs for insertion of multiple adaptors into library constructs. US patent application 20090176652, 2009
 - 19 Holt R A, Jones S J M. The new paradigm of flow cell sequencing. *Genome Res*, 2008, 18: 839—846
 - 20 Drmanac R, Sparks A B, Callow M J, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, 2009, 327: 78—81
 - 21 Braslavsky I, Hebert B, Kartalov E, et al. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci USA*, 2003, 100: 3960—3964
 - 22 Harris T D, Buzby P R, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320: 106—109
 - 23 Harris T D, Buzby P R, Jarosz M, et al. Optical train and method for TIRF single molecule detection and analysis. US patent application 20070070349, 2007
 - 24 Hardin S, Gao X, Briggs J, et al. Methods for real-time single molecule sequence determination. US patent 7329492, 2008
 - 25 Eid J, Fehr A, Gray J, et al. Real-Time DNA sequencing from single polymerase molecules. *Science*, 2009, 323: 133—138
 - 26 Levene M J, Korfach J, Turner S W, et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 2003, 299: 682—686
 - 27 Korfach J, Marks P J, Cicero R L, et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc Natl Acad Sci USA*, 2008, 105: 1176—1181
 - 28 Array based sequencing-by-synthesis. <http://www.mobious.com>
 - 29 Densham D H. Nucleic acid sequence analysis. EU Patent Application EP1229133, 2002
 - 30 Driscoll R J, Youngquist M G, Baldeschwieler J D. Atomic-scale imaging of DNA using scanning tunnelling microscopy. *Nature*, 1990, 346: 294—296
 - 31 Ikai A. TM and AFM of biolorganic molecules and structures. *Sur Sci Rep*, 1990, 26: 263—332
 - 32 Tanaka H, Kawai T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nat Nanotechnol*, 2009, 4: 518—522
 - 33 Bension R. Rapid sequencing of polymers. US patent application 20040214177, 2004
 - 34 Glover III, Roy W. Systems and methods of analyzing nucleic acid polymers and related components. US patent 7291467, 2007
 - 35 Branton D, Deamer D W, Marziali A, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 2008, 26: 1146—1153
 - 36 Meller A, Nivon L, Branton D. Voltage-driven DNA translocations through a nanopore. *Phys Rev Lett*, 2001, 86: 3435—3438
 - 37 Fologea D, Gershow M, Ledden B, et al. Detecting single stranded DNA with a solid state nanopore. *Nano Lett*, 2005, 5: 1905—1909
 - 38 Ling X S, Bready B, Pertsinidis A. Hybridization-assisted nanopore sequencing of nucleic acids. US patent application 20070190542, 2007
 - 39 Lagerqvist J, Zwolak M, Di Ventra M. Fast DNA sequencing via transverse electronic transport. *Nano Lett*, 2006, 6: 779—782
 - 40 He J, Lin L, Zhang P, et al. Identification of DNA basepairing via tunnel-current decay. *Nano Lett*, 2007, 7: 3854—3858
 - 41 Butlera T Z, Pavlenokb M, Derringtona I M, et al. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc Natl Acad Sci USA*, 2008, 105: 20647—20652
 - 42 Wu H C, Astier Y, Maglia G, et al. Protein nanopores with covalently attached molecular adapters. *J Am Chem Soc*, 2007, 129: 16142—16148

-
- 43 Clarke J, Wu H C, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*, 2009, 4: 265—270
- 44 Cheikh C, Koper G. Influence of the stick-slip transition on the electrokinetic behavior of nanoporous material. *Physica A*, 2007, 373: 21—28
- 45 Benner S, Chen R J, Wilson N A, et al. Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat Nanotechnol*, 2007, 2: 718—724
- 46 IBM Press Release. Advancing the Science of DNA Sequencing. 2009, <http://www.ibm.com>
- 47 Postma H W. Rapid sequencing of individual DNA molecules in graphene nanogaps. *Nano Lett*, 2010, DOI: 10.1021/nl9029237
- 48 Albertorio F, Hughes M E, Golovchenko J A, et al. Base dependent DNA-carbon nanotube interactions: activation enthalpies and assembly-disassembly control. *Nanotechnology*, 2009, 20: 395101
- 49 Gigliott B, Sakizzie B, Bethune D S, et al. Sequence-independent helical wrapping of single-walled carbon nanotubes by long genomic DNA. *Nano Lett*, 2006, 6: 159—164
- 50 Meng S, Maragakis P, Papaloukas C, et al. DNA nucleoside interaction and identification with carbon nanotubes. *Nano Lett*, 2007, 7: 45—50
- 51 Wright T P. Factors Affecting the Cost of Airplanes. *J Aeronaut Sci*, 1936, 3: 122—128
- 52 Karow J. The Cost of Sequencing a Human Genome? Answers Differ, Even for the Same Platform. In *Sequence*, 2009