

三、新一代DNA测序技术

DNA测序技术已广泛应用于生物学研究的各个领域，很多生物学问题都可以借助高通量DNA测序技术予以解决。过去三年，大规模平行测序平台（massively parallel DNA sequencing platform）已经发展为主流的测序技术，这项测序技术的出现不仅令DNA测序费用降到了以前的百分之一，还让基因组测序这项以前专属于大型测序中心的“特权”能够被众多研究人员分享。目前，新的测序技术及手段还在不断涌现，比如最新的进展就包括建立序列数据库、建立序列数据分析新方法以及设计测序试验等等。新一代DNA测序技术有助于人们以更低廉的价格，更全面、更深入地分析基因组、转录组及蛋白质之间交互作用组的各项数据。今后，各种测序将成为一项广泛使用的常规实验手段，这有望给生物学和生物医学研究领域带来革命性的变革。

DNA测序技术经历了漫长而曲折的发展历程。迄今为止，我们获得的绝大部分DNA序列都是基于Sanger测序法获得的。在过去5年间，人们至少从以下四个方面刺激了DNA测序技术的发展（表2）。

表2 刺激DNA测序技术发展的四个方面

人类基因组计划的出现，使人们面临了巨大的经费问题，因为传统的Sanger测序法无论如何优化，也无法大幅度降低测序费用；

人类以及其它主要模式生物参考序列数据库的建立使得短片段作图成为可能，这极大地促进了短片段测序技术的发展；

不断涌现的新型分子生物学技术极大地促进了分子生物学的发展，随之而来的越来越多的比如基因变异、RNA表达、蛋白与DNA相互作用以及染色体构象等等生物学现象需要人们用高通量DNA测序手段去解释，这也极大地促进了新型测序技术的发展；

其它学科、各项相关技术的发展，例如显微镜技术、表面化学技术、核苷生化技术、聚合酶工程技术、计算机技术、数据储存及分析技术等，为DNA测序技术提供了极大的支持。

1. 具有代表性的新一代DNA测序仪

最近市面上出现了很多新一代测序仪产品，例如美国Roche Applied Science公司的454基因组测序仪、美国Illumina公司和英国Solexa technology公司合作开发的Illumina测序仪、美国Applied Biosystems公司的SOLiD测序仪、Dover/Harvard公司的Polonator测序仪以及美国Helicos公司的HeliScope单分子测序仪。所有这些新型测序仪都使用了一种新的测序策略——循环芯片测序法（cyclic-array sequencing），也可将其称为“新一代测序技术或者第二代测序技术”。

所谓循环芯片测序法，简言之就是对布满DNA样品的芯片重复进行基于DNA的聚合酶反应（模板变性、引物退火杂交及延伸）以及荧光序列读取反应。2005年，有两篇论文曾对这种方法做出过详细介绍。与传统测序法相比，循环芯片测序法具有操作更简易、费用更低廉的优势，于是很快就获得了广泛的应用。

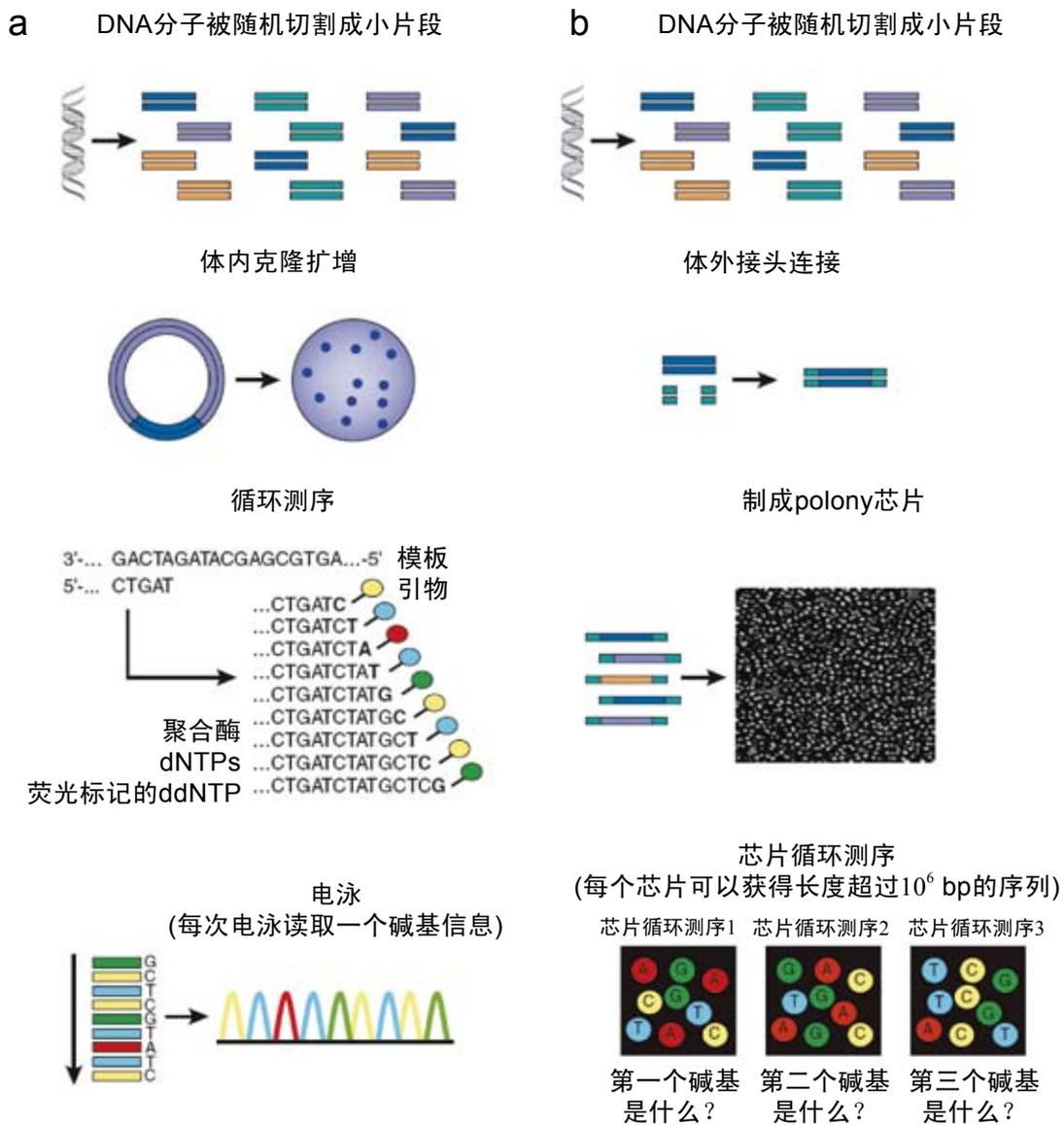


图1 传统的Sanger测序法及新一代DNA测序技术工作流程图。

(a) 高通量鸟枪Sanger测序法。首先基因组DNA被随机切割成小片段分子，接着众多小片段DNA被克隆入质粒载体，随后转化到大肠杆菌中。最后培养大肠杆菌提取质粒，进行测序。每一个测序反应都在只有几微升的反应体系中完成。测序后获得一系列长短不一的末端标记有荧光的片段，最后通过对每一个延伸反应产物末端荧光颜色进行识别来读取DNA序列。(b) 鸟枪循环芯片测序法。首先将基因组DNA随机分割成小片段DNA分子，然后在这些小片段DNA分子的末端连接上普通的接头，最后用这些小片段DNA分子制成polony芯片。每一个polony中都含有一个小片段DNA分子的许多个拷贝。许多这样的polony集合在一起就形成了polony芯片。这样一次测序反应就可以同时对众多的polony进行测序。然后与sanger法中一样，通过对每一个延伸反应产物末端荧光颜色进行识别来读取出DNA序列。重复上述步骤就能获得完整的序列。

虽然这些新一代测序仪以及芯片的实际制作过程似乎都和传统的测序方法有很大的不同，而且各有特点（表3），但实际上它们背后的原理和技术都是非常相似甚至是相同的（图1b）。新一代测序法首先也是将基因组DNA随机切割成小片段DNA分子，然后在体外给这些小片段分子的末端连接上接头制成文库，也可以使用配对标签（mate-paired tag）制成跨步文库（jumping libraries）。随后可以通过原位polony（in situ polony，小词典1）、微乳液PCR（emulsion PCR）或桥式PCR（bridge PCR）（图5）等方法获得测序模板。

上述方法有一个共同点，那就是任何一个小片段DNA分子的PCR扩增产物都是在空间上聚集的：原位

polony法和桥式PCR法中所有的产物都集中在平板的某处，在微乳液PCR法（emulsion PCR）中所有的产物都集中在微珠的表面。真正的测序反应本身和传统测序法一样，是由重复的聚合酶促反应和最后的荧光读取分析反应组成（图6）。本文讨论的所有测序仪都是使用合成测序法（sequencing by synthesis），即通过聚合酶或连接酶不断地延伸引物获得模板序列，最后对每一轮反应的结果进行荧光图像采集、分析，获得序列结果。

表3 新一代测序技术的特点

体外构建DNA文库，随后体外克隆扩增待测DNA片段，这就解决了传统Sanger测序技术中好几个限制平行测序规模的瓶颈问题，比如转化大肠杆菌以及阳性克隆挑选等问题；
基于芯片的测序方式能极大地提高平行测序的能力。由于每一个待测DNA片段克隆的直径都不到1 μm，因此可以在芯片上同时对上亿个待测DNA片段进行测序。
因为每一个待测DNA片段克隆都固定在芯片上的固定位置，因此可以一次对芯片上的所有片段进行测序，而只需要使用几微升的酶等反应试剂，这就相当于每一个待测片段只使用了几皮升（picoliter）或几飞升（femtoliter）的反应试剂，从而极大降低了测序反应的费用；
测序片段长度短：几乎所有的新一代测序技术的测序长度都要明显短于传统测序技术的测序长度；
准确率不高：平均来说，新一代测序技术的测序准确率要比传统测序技术低10倍。

注：虽然目前测序片段长度短和准确率不高这两个缺点限制了新一代测序技术的应用，不过应该坚信，我们最终一定会克服这些问题。就好像经过了30年的努力，传统的测序技术也今非昔比，到达了今天的水平一样。

1.1 454测序仪

454测序仪的出现极大促进了测序业务的开展，科研人员已经将测序技术作为解决科研工作中许多常见问题的利器。这是因为454测序仪在以下几个方面取得了质的突破：首先是解决了高通量测序问题；其次它简化了样品准备步骤，将以往转化大肠杆菌扩增质粒的繁琐过程全部用简单的体外PCR扩增法替代了；最后，它缩小了测序反应体积，节省了试剂。这样，454测序仪做到了以极其低廉的价格进行大规模平行测序反应。它的测序规模之大、测序费用之低是以往的测序仪无法匹敌的。454测序仪与其它的新一代测序仪一起，降低了测序检测的费用，推动了测序技术平民化进程，使得小实验室也能开展测序检测项目，打破了以往只有少数几个大型测序中心才能进行测序研究的“垄断地位”。在过去的18个月里，由于有了454测序仪的帮助，人们对人类基因组的结构有了更深入的了解，同时第一次使用非Sanger测序法对个人进行了测序，还建立了一种发现小RNA的新方法。不过，要能让更多的人使用上新一代的测序产品，它们还需要变得更便宜，并且更加容易操作。在一段时间之内，454测序仪必定会进一步降低测序费用，帮助人们迎接个人基因组时代的到来。

自从诺贝尔奖得主Frederick Sanger和Walter Gilbert（图2）分别发明了链终止法DNA测序技术（sequencing by chain termination technique）和链断裂法DNA测序技术（sequencing by chain fragmentation technique）之后，人们就一直希望能够扩大DNA测序技术的处理规模。到了今天，我们对测序技术的需求和对计算机技术的需求一起出现了迅猛的增长，因为测序技术的发展速度已经远远跟不上实验要求的增长速度。于是出现了好几种替代Sanger测序法的新型测序方法，比如杂交测序法、借助原子力显微镜（atomic force microscopy）直接DNA成像测序法（direct imaging of DNA sequence）、质谱分析法、

合成测序法以及微液流测序法等等。在我们进行人类基因组计划时还出现了三项技术改进方法，即使用荧光标记物取代了放射性标记物来标记终止碱基（双脱氧碱基）；使用毛细管电泳（**capillary electrophoresis**）取代了传统的平板凝胶电泳；建立了末端配对测序法（**paired-end sequencing**）来对质粒、**fosmid**、人工细菌染色体（**BAC**）等短片段序列进行测序，解决了测序长度带来的限制问题。同时，开展研究的自动化液体分装技术（**liquid-handling robotics**）帮助我们摆脱了人工试管操作，可以用自动化的方式在微量滴定板（**microtiter plate**）上装载待测序样品（质粒等），极大地降低了测序的费用和劳动强度。

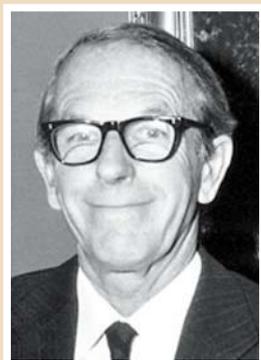


图2 Frederick Sanger（左）与Walter Gilbert（右）。图片来源：Nobelprize.org

随着美国**454 Life Sciences**公司（该公司现已被美国罗氏公司收购）的第一台新一代测序仪——454测序仪的面世，我们获得了一种完全不同的测序方式。454测序仪引领的新一代测序技术在一直困扰传统测序技术的三个瓶颈问题上取得了突破。这三个问题分别是文库制备、模板制备和测序。而且，在随后出现的其它新一代测序仪产品身上，我们或多或少都会发现在454测序仪上使用到的技术，这也足以说明454测序仪的技术创新的确取得了巨大的成功。

454测序仪的先行者地位使它对整个测序业的影响远远超过了其它新一代测序仪竞争对手。这一点从Leamon、Rothberg等人撰写的一篇介绍2005年技术进展的论文被引用了570多次的事实，以及有100多篇经过同行审议的关于人类遗传学、代谢组学、生态学、进化学以及古生物学的论文（**peer-reviewed publications**）都是使用454测序仪开展的研究多个事实中都能够得到证明。454测序仪技术是继Sanger测序技术之后出现的第一个用于对细菌基因组进行从头测序的新技术，也是第一个被用来对人类基因组进行测序的非Sanger测序技术。其它使用454测序仪开展的重要研究项目包括探究蜜蜂消失原因的项目、研究人类基因组重排复杂性的项目、建立用于研究传染性疾新方法的项目以及对尼安德特人（**Neanderthal**）基因组的测序项目等。

1.1.1 摩尔定律对454测序仪的影响

454测序仪的迅猛发展不是因为我们想要Sanger测序仪小型化，而是因为新型奔腾芯片的出现以及摩尔定律法则给我们带来的希望。很明显，常规的人类基因测序项目会对我们处理测序技术的能力提出更高要求，这与我们对计算机处理能力的要求是一样的。不过，只有将计算机的电子管换成晶体管，才为后来集成电路技术的发展提供了可能，这正是计算机产业发展的关键所在。而希望对传统的毛细管电泳技术进行改良，提高它的速度和处理规模，正如只用电子管直接制作集成电路一样不可能。因此，如果将各种测序技术比作一个个晶体管，将一系列测序步骤整合起来比作集成电路，那么也就可以用摩尔定律来预测DNA测序技术的发展速度了。

合成测序法概念虽然在提出的时候还不算成功，但它的出现为测序仪小型化奠定了基础。基于合成测

序法出现了两种策略：一种是循环可切除终止测序法（cyclic reversible termination technology），即依次逐个添加荧光标记的碱基，继而检测荧光信号，切除荧光基团，如此往复；另一种策略是焦磷酸测序法（sequenced by detecting pyrophosphate release）。454测序仪采用的正是焦磷酸测序法，因为它似乎比第一种方法的效率更高。结果证明，454公司的选择是正确的。454测序仪采用的是小型化焦磷酸测序反应，测序模板准备和焦磷酸测序反应步骤都是在固态芯片上完成的。

实际上，早在上世纪90年代中期，焦磷酸测序技术就已经被科研界用来进行基因分型工作了，但那时的焦磷酸测序技术还不能够满足标准的测序实验要求，因为它的测序长度太短，因此只能用于旨在发现SNP的基因分型研究当中。当时进行基因分型操作时，是在微量滴定板（microtiter plate）上进行的，可以连续进行最多96次基因分型实验，平均每个样品花费20美分。那时焦磷酸测序还不能用于从头测序工作，因为从头测序需要对每一个尤其是第一个碱基都能准确地区分清楚，而焦磷酸测序只能简单地对已知位点的碱基进行检测，而且从头测序要求的测序长度也是焦磷酸测序法无法达到的。

不过，由于焦磷酸测序的原理是通过检测碱基掺入时发出的光来进行测序的（图3），所以它并不需要类似于电泳之类的物理分离过程来对碱基进行区分。这也就是说焦磷酸测序仪可以“缩小（减）”到只需要检测光线就够了，而不需要像传统的测序仪还需要电泳设备，而这正是限制传统电泳仪小型化的关键所在。发光检测方法还能够进行多路平行操作，但是直到454测序仪出现之前，还没有人这样做过，以前都是依次进行检测的。和晶体管早期的遭遇一样（当时人们也怀疑晶体管替代不了电子管），人们同时对高密度的，用于并行焦磷酸测序的反应也充满了疑问。不过，当我们不再在溶液中进行测序反应，而是将测序模板、所有的试剂（酶）都固定在平板上制成芯片之后，就获得了小型化的，能进行多路并行处理的测序仪，这就与晶体管被小型化并整合成集成电路的过程一样。此外，借助微量滴定板上一个个的小孔所达到的将不同测序反应进行分隔这一目的，也能通过在单个固相支持物上进行严密包裹（隔离）的反应来实现。在这些各自隔绝的反应体系中，链聚合反应速度和发光速度都能通过对反应试剂和产物弥散状况进行严密的控制来进行精密的调整。

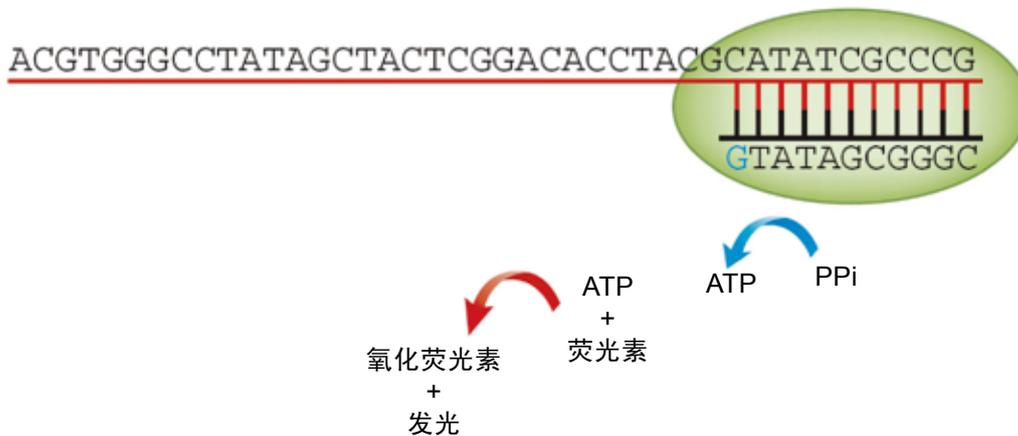


图3 焦磷酸测序反应原理。

图中红色单链表示模板链，引物以黑色表示，DNA聚合酶以绿色的椭圆形表示。每当掺入一个碱基（如图中蓝色的G）时，就会释放出焦磷酸（PPI），然后被磷酸化酶（sulfurylase，图中蓝色箭头所示）转化成ATP。然后荧光素酶（图中红色箭头所示）就可以在ATP参与下将荧光素转变成氧化荧光素，同时发光也会被测序仪检测到。

1.1.2 新的并行试验方法

在开发新型高通量、高并行运行方法时碰到的一个关键问题是，如何将反应试剂同时加入数量如此之多的各个反应体系中？在焦磷酸测序的过程当中需要反复加入不同的碱基以供测序反应使用，而当时的自动化加样设备无法有效地做到对这么多的反应体系同时循环加样。于是，开发一种全新的高密度并行处理方法这

一重要课题又再一次摆在了科研人员的面前。这一次，我们找到了一个非常简单但是又很巧妙地方法。在高密度的反应芯片表面使用层流（laminar flow）加样方式，反应试剂会通过扩散作用很好地进入每一个反应体系，而且也可以用层流的方式洗去多余的反应试剂。现在，所有的新一代测序仪都采用了这种层流加样方法。

为了将每个单独的测序反应都分隔开来，我们一开始使用平板（芯片），不过在平板上平均每一平方厘米的面积上最多只能同时进行数百至数千个反应。但我们希望达到的是在每平方厘米的面积上同时进行100万个测序反应，这样才能令测序仪小型化，同时节省试剂并进行快速成像和测序。为了实现更高密度的测序反应，我们在平板上制作了很多小孔，将每个反应体系都安置在这些小孔中，这些小孔都足够深，足以分隔每个反应体系。虽然这种方法极大提高了测序反应的密度，缩小了平板的面积，但是要达到我们的要求还是需要60mm×60mm大小的芯片才行。

针对图像采集问题使用了商业化的天文学照相（astrological grade camera）器材，在电荷偶合装置（CCD）的表面连接上光纤束（fiber-optic bundle）。这些光纤是锥形排列的，这样可以将大范围的光信号都传输到CCD表面上很小的一个范围。采取下面两个步骤，我们就可以制成含有高密度小孔的芯片：先将光纤束连接到类似于载玻片一样的一次性芯片上，然后用酸蚀刻（acid etching procedure）技术在玻片的另一面打上小孔。这种酸蚀刻技术是根据制作生物传感器的技术改进而来的。

454公司制作的每张芯片上可以达到数百万个小孔，每一个小孔都是一个独立的“反应站”，互不干扰，测序反应发出的光被连接在芯片上的光纤传送到CCD记录下来（图4）。这种芯片就好像集成电路一样一次可以同时处理数百万个测序反应。这种芯片同样也能被其它通过发光检测技术的产品所使用。454测序仪也没有像以前的96孔板焦磷酸测序仪那样使用液态的试剂，而是将试剂和模板统统都吸附在一个个微珠上，然后把把这些微珠一个个地放到芯片上的小孔中，每孔一个微珠。这种固定步骤不仅保证了每孔测序反应的独立性，也极大地节省了试剂消耗费用。

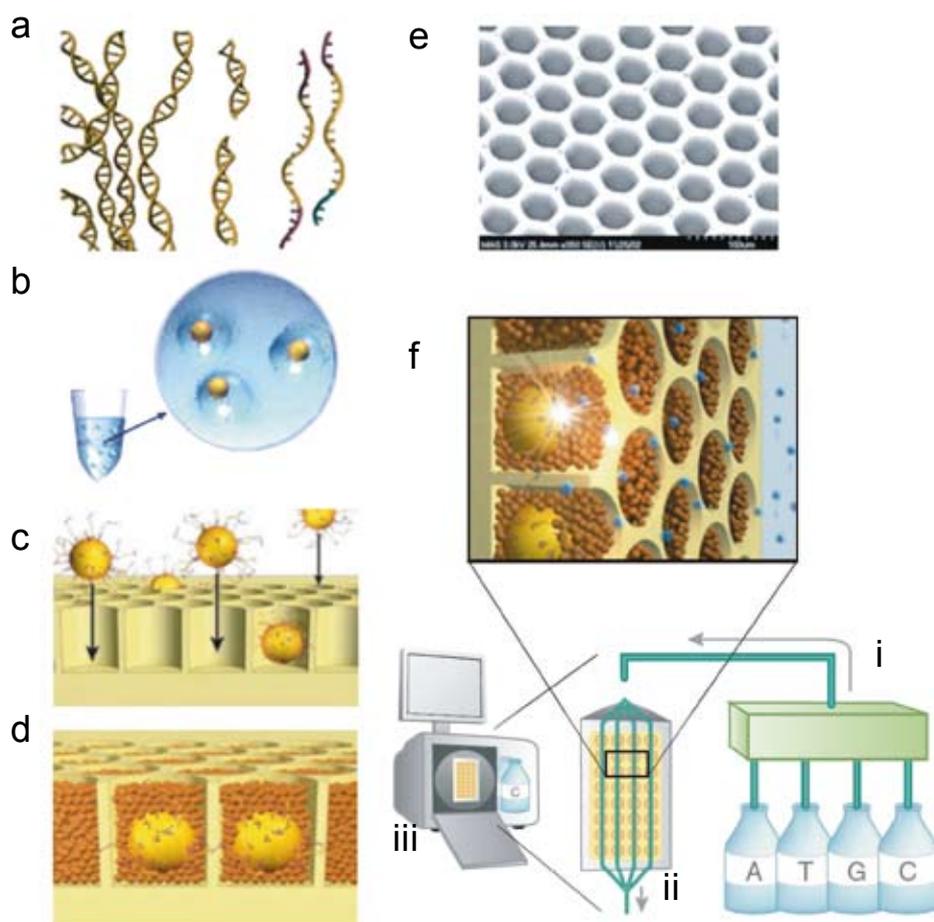


图4 454测序仪技术应用简介。

(a) 分离基因组DNA，随即切割成小片段，每个片段两端连接上接头序列，并变性形成单链；(b) 将(a)中制成的单链分子与微珠连接，每个微珠连接上一条单链分子，然后将这些微珠在乳液中包裹成一个个油包水的小液滴，每个液滴中包含一个微珠，然后进行乳液PCR扩增，最后每个微珠上都会携带有上千万条待测模板分子；(c) 打破液滴，收集微珠，然后将微珠放置到芯片上的小孔中，每个小孔中一个微珠；(d) 在每个小孔中置入吸附有焦磷酸测序反应所需酶的小微珠；(e) 微珠置入前的芯片图像；(f) 454测序仪主要包括以下几个部分：(i) 液体试剂供应装置；(ii) 反应池；(iii) 光线探测成像系统和计算机控制系统。

要想实现高通量基因组测序，只对测序步骤进行优化还是远远不够的。人类基因组计划花费的30亿美元经费中有很大大一部分都用在了测序样品制备阶段。当时即使是采用最简单的制备样品方法也需要将目标片段克隆到细菌中，挑克隆，再转到96孔板，然后进行克隆扩增，提取质粒，制备测序模板。这种工作流程既耗时也耗钱。

如果采用新型的文库制备方法就可以极大地节省这部分开支，这种新型的方法是先分离基因组DNA，随机切割成小片段分子，然后通过有限稀释（limiting dilution）和聚合酶扩增反应，即体外克隆方式（clones without bacterial）制备模板片段。这样，从模板制备到最后的测序反应整个过程都能够在体外完成。

1.1.3 从发明到创新

从概念的提出到最后技术上的实现，454测序仪主要关注两个方面，首先是开发蚀刻光纤玻片；其次，改进焦磷酸测序方法使其能在固相支持物上进行，即将其改造成固态焦磷酸测序法，同时也对模板及文库构建方法进行了改进，让454测序仪能进行长片段测序工作和从头测序工作。

1.1.3.1 在蚀刻板上的小孔中进行固态、长片段焦磷酸测序反应

蚀刻技术经过改良之后能在75mm×75mm的玻片上刻出深55 μm、宽44 μm的小孔。而开发固态测序方法和改良测序长度则是两个紧密相关的问题，因为在固定的小孔中反应实际上就能改进测序质量和测序长度。由于反应试剂能迅速渗透到小孔中，因此反应速度也会加快。而且这里也没有使用三磷酸腺苷双磷酸酶（apyrase）提取未参与反应的碱基，而是将芯片置入反应池中通过层流液体的快速渗透作用将多余的未参与反应的碱基和反应副产品洗掉，由此得到100bp~500bp的测序长度。在能有效去除多余碱基的同时，每轮反应中聚合酶的效率也得到了极大提高。这样高效率的聚合反应使得454测序仪具有较长测序长度的同时也保证了高准确性，测序长度在200bp时的准确率高达99.5%。这是因为通过降低小孔中残存的未参与反应的碱基浓度，可以降低这些碱基对聚合酶活性的抑制作用，或者降低这些碱基导致的延后错误（carry-forward error，即由于未参与反应的碱基导致的测序反应不同步现象）的发生率。454测序仪在测序长度和准确率方面具有优势还因为其在应用流体学、表面化学和酶学（包括选择更好的聚合酶、在更高的温度进行测序反应以及更换及平衡各个酶组分）等方面都有创新（表4）。

还有一些能提高测序精度和测序长度的技术，不过暂时还没有商业化产品。这些技术包括使用可切除的终止碱基（reversible terminator）提高对同聚物（homopolymers）的检测精度；双末端测序法（double-ended sequencing），即同一模板的两条链均不测序；以及选择性酶固定法（alternative enzyme-immobilization method）等。这些技术改进还都没有用到测序仪产品中，有一部分原因是因为现在还没有必要使用。

表4 通过与454合作进行的研究应用项目

应用领域	合作项目
细菌测序及比较基因组学	结核分枝杆菌
古生物学及古代DNA	对尼安德特人的DNA进行测序
小RNA	拟南芥DICER功能分析
环境基因组学	有关导致蜂群崩溃症的微生物研究
基因组结构研究	人类基因组中的变异
人类全基因组测序	对James Watson的基因组进行测序

注：蜜蜂群崩溃症（honeybee colony collapse），指的是来自养蜂业的蜂箱或自然界存在的欧洲蜜蜂群的工蜂突然消失的现象，又称作Colony collapse disorder（CCD）。

完全的体外大规模模板制备工作是达成高通量、低价格测序技术的前提。已广泛使用的乳液PCR扩增技术就是一种很好的方法。不过，由于很难在热循环测序反应中保证乳液微滴的稳定性，因此最开始实验的模板扩增方法是恒温扩增法（isothermal）。

乳液PCR不需要借助细菌的帮助就能扩增模板，虽然这一点非常诱人，但最开始时并没有合适的表面活性剂能帮助乳液在热循环过程中保持稳定。于是出现了恒温扩增法，即滚环扩增反应（RCA）。虽然滚环扩增反应的产量非常高，但这些产物中大部分都不能用来作为测序模板。因此，还需要找到一种不需要细菌扩增，能用于有限稀释的模板扩增新方法。于是，人们又把目光转回了PCR法。在RCA法中，首先将模板克隆有限稀释之后置入光纤玻片上的小孔中，然后用橡胶衬垫把光纤玻片封闭起来，将玻片放入传统的平顶PCR仪进行扩增。这种方法取得了成功，但是效率不高，因为在玻片中的热质量（thermal mass）和它的钳效应（clamping mechanism）需要更长的PCR循环时间，而且模板的有限稀释度不能低于10%。孔与孔之间的相互污染现象也是一个不容忽视的问题。不过无论如何，该方法还是第一个首先从全基因组文库中扩增模板然后使用非Sanger、非Gilbert测序法对基因组进行从头测序的方法，也是第一个使用体外模板扩增技术进行全基因组（腺病毒基因组）测序的方法。

乳液滴的热稳定性问题最终通过加入用于制造炸药的表面活性剂得到了解决，于是乳液PCR技术马上在众多新一代测序仪中得到了广泛的应用。因为乳液PCR技术具有高效性、可扩展性，既能从30Kb的腺病毒基因组中扩增模板，也能从好几Mb的肺炎链球菌（*Streptococcus pneumoniae*）基因组中扩增模板。

随着测序精度、测序长度、乳液滴稳定性等各方面技术的不断发展，454测序仪已经不仅仅用于对细菌级别的基因组进行测序了，还能对更高级、更复杂的生物基因组进行测序，例如现代人类基因组、尼安德特人基因组以及环境基因组等。

1.1.3.3 文库制备

文库制备包括以下几个步骤，首先随机切割样品基因组，获得大量DNA片段，然后接上接头进行扩增反应。454测序仪的样品制备程序和Craig Venter等人的鸟枪法样品制备程序有着本质的差别。454公司采用的是如图4中所示的有限稀释、乳液PCR扩增法，而没有鸟枪法中的细菌克隆繁殖步骤。去掉了细菌繁殖步骤极大地提高了整个测序工作的速度和效率，同时避免了由于细菌繁殖导致的序列丢失的可能性。这种方法同样对古老DNA和代谢基因组学的研究也非常适用。末端配对文库制备方法的建立同样帮助454测序仪获得了对复杂基因组从头测序、对重复片段测序以及对基因组结构（复制、重排）展开系统研究三种能力。这种末端配对文库的制备方法是受到了Bender科研小组对果蝇（*Drosophila*）制备跨步文库（jumping library）方法的启发而发展得来的。

1.1.4 应用范围

随着越来越多重要的研究领域受到测序技术的影响，454公司开始和其它商业和学术机构开展合作，进行样品测序和分析工作。这些合作项目又进一步验证了454测序仪使用的技术能够在众多领域中发挥作用，例如末端配对文库技术对于研究基因组结构的作用和乳液PCR技术捕获目的DNA片段的作用等。

1.1.4.1 细菌基因组测序和比较基因组研究

为了测试454测序仪在全基因组测序方面的能力，454公司一开始就参与了一项合作项目，该研究项目会对4株结核分支杆菌基因组进行测序，这四株结核分支杆菌分别是一株对R207910具有耐药性的结核分支杆菌（*Mycobacterium tuberculosis*）菌株，基因组大小约4Mb；两株对R207910具有耐药性的耻垢分支杆

菌 (*Mycobacterium smegmatis*)，基因组大小约6Mb；以及一株正常的耻垢分枝杆菌 (*Mycobacterium smegmatis*)，基因组大小约6Mb。他们希望能发现结核分枝杆菌 (*Mycobacterium tuberculosis*) 对R207910产生抗药性的机制。该项研究清晰的展现了454测序仪在测序速度和测序精度方面的优势。使用传统的Sanger测序法对一个4Mb的基因组和3个6Mb的基因组进行测序需要好几个月的时间，而用454测序仪，在只有一位实验人员参与实验的情况下，包括样品制备等步骤在内所用的时间仅需要一周。而且使用454测序仪还避免了传统测序方法中细菌克隆阶段可能出现的错误，获得了高质量的测序结果，发现了导致结核分枝杆菌对R207910产生抗药性的两个点突变位点。这项研究成果让我们在最近的40年内第一次找到了特异性治疗结核病的药物，同时也对454测序仪在细菌基因组测序方面的应用价值有了深刻的体会。随后，454测序仪又参与了比较基因组学研究项目、对高致病性细菌空肠弯曲菌 (*Campylobacter jejun*) 基因组的从头测序项目、对幽门螺杆菌 (*Helicobacter pylori*) 在慢性胃炎致病过程中的进化研究项目、从南极海冰细菌 (Antarctic sea ice bacterium) 中新发现冰结合蛋白 (ice-binding protein) 并对其测序的研究项目，以及在引起肺炎、脑膜炎和泌尿道感染的细菌中发现致病因素的研究项目等。

由于454测序仪不会因为细菌克隆产生测序误差，所以在对结核分枝杆菌抗药性的研究中表现出了非常强的发现突变位点的能力，这一点也被后来的其它研究项目所证实。此外，最近在用454测序仪进行的人类基因组测序项目中发现了长达29Mb的片段与人类基因组参考序列build-36不相符，这些片段被认为是参考序列中不存在的序列，属于基因组中的常染色质部分。不过，还需要注意的是，有些报道称由于重复片段的存在会出现序列组装错误，而且小模板片段雾化 (nebulization) 处理这种方式也会造成测序错误出现。

1.1.4.2 小RNA测序

对于包括miRNA在内的小RNA的研究兴趣从2005年开始就持续不断升温，而2005年恰好也是454测序仪上市的那一年。454测序仪以其不需要进行传统的细菌克隆步骤和足以覆盖只有21bp长的miRNA的测序长度等优势，很快就在miRNA的作用研究之中占据了一席之地。454测序仪最早参与进行的miRNA研究是对拟南芥 (*Arabidopsis thaliana*) miRNA开展的研究。随后马上又参与了另一项研究项目，在这个项目中我们在小鼠体内发现了一种新型的小RNA——piRNA。这些研究项目为我们在人类、黑猩猩、斑马鱼和肿瘤细胞系中开展小RNA研究铺平了道路。454测序仪具有的这种对小RNA进行研究的能力使它在众多有关RNA的研究领域都能有所作为，例如转录体研究领域、EST研究领域、5'-RATE研究领域和基于转录体的SNP研究领域等。

1.1.4.3 在古生物学和古DNA研究领域的作用

要用传统的测序方法对尼安德特人的基因组进行测序研究非常困难，因为这些古老DNA量非常少，而且都早已裂解成了片段。一开始，454公司使用比较容易得到的不太重要的古代DNA样品检验了454测序仪对它们的测序能力，结果非常好，尽管当时454测序仪的测序长度只有100bp。不过，尼安德特人的基因组片段长度基本上都介于40bp~90bp之间，而且最近开发的乳液PCR方法也能够对微量（单分子）样本进行很好的扩增。于是，454测序仪参与了对38,000年前古老的尼安德特人的基因组进行测序的工作，研究结果分别发表在了好几篇论文当中，引起了广泛的关注，并促进了古生物学基因组的研究。随后有人对长毛象 (woolly mammoth) 和更新世狼 (Pleistocene wolves) 的基因组开展了测序研究。

1.1.4.4 环境基因组学和感染性疾病研究领域

美国在2001年爆发了炭疽恐怖袭击危机之后，454公司便对如何使用454测序仪对复杂的、未知的、未人工培养的环境微生物基因组进行测序展开了研究。前后两个合作研究项目均表明454测序仪能够用于从DNA混合样品中发现未知微生物并对其进行分类。在第一个研究项目中，有三名患者都接受了同一名澳大利亚器

官捐赠者的器官，之后均因不明原因而死亡。从这三名死者身上提取了非人类DNA样品进行测序，结果获得了144,000条序列。分析后发现，这些序列分别属于一种沙粒病毒科（*Arenaviridae*）家族病毒的14个不同基因。随后进行的第二项研究在对健康蜂群和患病蜂群进行环境基因组学比较研究之后发现，以色列急性麻痹病毒（*Israeli acute paralysis virus*）是导致蜜蜂蜂群崩溃症的元凶。上述这些研究都突出了454测序仪的一个特点，即在样品准备前不需要进行克隆或预扩增步骤，因此非常适用于对未知的未能人工培养的物种进行测序。这些特点也在其它对地下矿藏、深海、土壤和高盐等环境下进行的环境微生物构成方面的研究所证实。

1.1.4.5 基因组结构研究领域

454测序仪技术的进步使它适用于更多的科研领域。最新开发的末端配对测序法（*paired-end sequencing*）就非常适用于发现人类基因组当中的结构变异。末端配对作图过程（*paired-end mapping*），简单来说就是对一个非洲人和一个欧洲人的基因组进行测序后发现结构变异并对其作图，最终将1,000多个3Kb或更长的结构变异片段定位到人类基因组参考序列中。研究发现，在人类基因组当中存在的结构变异远远超过了人们的预计，其中有很多变异都会造成非常重要的表型改变。这项对诺贝尔奖得主James Watson基因组进行测序的项目和其它相关研究，一起使得“人类基因多样性（*human genetic variation*）”这一科学命题成为了《科学》（*Science*）杂志的年度重大科技突破。

1.2 Illumina测序仪

Illumina测序仪通常也被称作Solexa测序仪（Illumina测序仪的特点见表5）。它适用于采用各种方法制备的DNA文库，文库中DNA片段可以长达数百bp，并可通过桥式PCR来扩增模板片段（图5b）。在桥式PCR反应中，正向引物和反向引物都被通过一个柔性接头（*flexible linker*）固定在固相载体（*solid substrate*）上。经过PCR反应，所有的模板扩增产物就都被固定到了芯片上固定的位置。

值得注意的是，Illumina测序仪使用的桥式PCR与传统的桥式PCR有所不同，它会交替使用*Bst*聚合酶进行延伸反应以及使用甲酰胺（*formamide*）进行变性反应。这样，经过桥式PCR扩增之后，也会在固相载体上形成一个模板“克隆”。一块芯片的8条独立“泳道”上每一条泳道都可以容纳数百万的模板“克隆”，这样一次就可以同时对8个不同的文库进行测序。

经过上述PCR扩增步骤之后，所有的模板都被线性化处理（*linearization*）而形成单链模板，接着与测序引物退火、杂交。随后使用修饰的DNA聚合酶和四种核苷酸混合试剂进行单碱基延伸测序反应（图6b）。这些核苷酸试剂都经过两种方式处理过，它们都是可逆的终止子（*reversible terminator*）。这些核苷酸的3'羟基端都有一个可被化学法切除的基团，这样每一次反应都只会掺入一个核苷酸，同时每种核苷酸都标记上了可被化学法切除的不同颜色的荧光基团，以标识每种碱基。经过一轮单碱基掺入反应采集到信号之后，就可以通过化学方法切除上述被掺入核苷酸上标记的两个基团，然后就能够继续掺入下一个核苷酸，重复测序反应了。这种测序方法对36bp长度的序列测序准确率是非常高的，不过如果处理更长的序列，准确率就会有所降低了。

表5 Illumina测序仪的特点

对于检测同聚物来说，Illumina测序仪要比454测序仪好很多；

Illumina测序仪的平均出错率在1%~1.5%，不过可以通过质量优化将错误率降低到不到0.1%；

与其它测序仪一样，经过改进之后，Illumina测序仪也能进行配对测序（*mate-paired read*），比如每一次测序都能同时对两条不同的模板进行测序，各获得其中长约36bp的序列信息；

Illumina测序仪在测序长度方面有所缺陷，这是因为它存在下列诸如被掺入核苷酸上标记的荧光基团或者终止基团切除不完全等原因导致测序信号发生了衰减（*decay*）和相移（*dephasing*）；

Illumina测序仪在检测碱基替换突变时最易出错，其次是检测插入或缺失突变时。

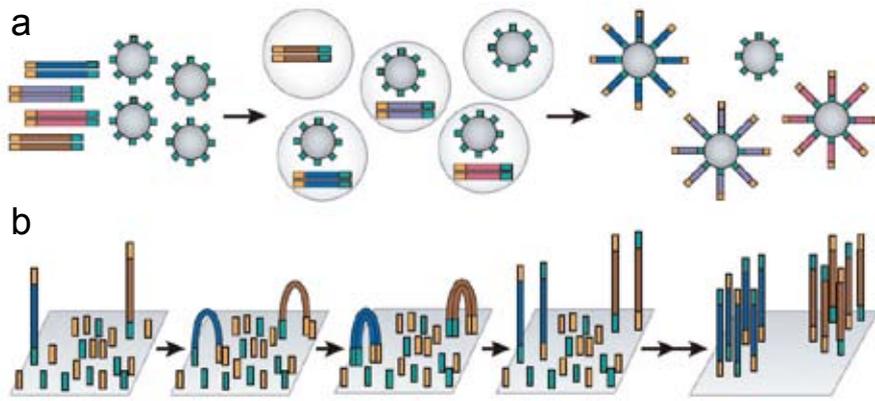


图5 克隆扩增模板片段。

(a) 在454测序仪、Polonator测序仪以及SOLiD测序仪中广泛用来扩增模板片段的微乳液PCR技术示意图。首先，体外构建好的两端连接有接头（如图中金色和绿色所示）的DNA文库在一个油包水（water-in-oil）的乳液环境中经PCR扩增，由于此处只使用了一对针对接头的引物（有一条PCR引物的5'端被固定在微珠表面），因此进行的是多模板PCR（multi-template PCR）而不是多重PCR（multiplex PCR）。另外，由于在模板浓度还很低的情况下，大部分包裹着微珠的微乳液滴中最多只能有一条引物，因此经过PCR扩增之后，每一个微珠表面上都只会连接上一种模板分子的扩增产物。打破微乳液滴，将这些携带有大量模板分子的微珠收集起来制成芯片。（b）Illumina测序仪中使用的桥式PCR技术示意图。首先，体外构建好的两端连接有接头（如图中金色和绿色所示）的DNA文库经由密集固定在载体上的引物（引物的5'端借助一个柔性接头固定在载体上）进行PCR扩增。这样扩增产物也会被固定在载体上，PCR反应结束之后，每一个模板克隆都包含有1,000条模板产物。仔细检测模板浓度既可以保证在芯片上最大限度地携带模板，同时也能避免模板过于拥挤的现象发生。

1.3 AB SOLiD测序仪

AB SOLiD测序仪可以对由任何方法制成的DNA文库进行测序。AB SOLiD测序仪有一个极大的特点就是能够将富集模板片段的微珠在芯片上进行高度可控的任意排列。AB SOLiD测序仪也是使用如图5a中所示的微乳液PCR方法扩增模板片段的，不过，它这里使用的是直径只有 $1\mu\text{m}$ 的小磁珠。PCR扩增反应结束之后，微乳液滴被打破，小磁珠被富集起来固定到固态平板上，制成高密度测序芯片。后面的合成测序法由DNA连接酶而非DNA聚合酶完成。

首先，通用引物与模板片段两端的接头序列互补结合，然后连接酶将一个被荧光标记的8bp长的核酸探针片段（fluorescently labeled octamers）连接到引物末端（图6c）。这段8bp长的核酸探针片段是经过设计的，比如其中第五位碱基上就标记了荧光。连接反应完成之后，就可以采集荧光图像，然后在第五位碱基和第六位碱基之间切断，去掉荧光标签。如此反复，就可以获得每间隔四个碱基的第五号碱基的确切信息，比如第5号碱基、第10号碱基、第15号碱基以及第20号碱基等等。经过几轮这样的循环之后，已经获得延伸的引物会变性脱落，再重新结合上新的引物从头开始新一轮测序，不过这一次可能获得的是第4号碱基、第9号碱基、第14号碱基以及第19号碱基的信息。我们可以使用不同长度的引物（+1或者-1）或者使用在不同位点（比如第2号碱基）标记荧光的8bp核酸探针片段达到这个目的。如此反复，最终就能获得整条模板片段的完整序列信息。

AB SOLiD测序仪还有一个特点就是使用了双碱基编码技术（two-base encoding），该技术具有误差校正功能，因为它是通过两个碱基来对应一个荧光信号而不是传统的一个碱基对应一个荧光信号，这样每一个位点都会被检测两次，因此出错率明显降低。

Polonator测序仪是一个和AB SOLiD测序仪比较相似的产品，因为它也运用了J.S等人和哈佛大学Church研究小组开发的部分系统。Polonator测序仪同样也使用微乳液PCR法扩增模板片段，使用连接酶法测序。不过，Polonator测序仪的价格要比其它第二代测序仪低得多。而且更重要的是，Polonator测序仪是一个开源的设备，用户可以通过自己编程“设计”出最适合自己的测序仪。不过，Polonator测序仪目前可测序的长度还非常有限。

值得注意的是，454测序仪、SOLiD测序仪以及Polonator测序仪还都存在一个共同的不足，那就是微乳液PCR技术实在是太过麻烦并且对实验操作的技术要求较高。不过从另一方面来说，使用仅仅只有1 μ m大小的磁珠构成的高密度测序芯片进行测序（不论是使用聚合酶法、连接酶法，还是其它的生化方法）是最有可能实现的高通量测序方法。因为1 μ m是衍射技术（diffraction）所能分辨的极限大小了。另一方面，最近报道的使用1 μ m磁珠进行高分辨率芯片点样技术的突破，使我们有望实现每个测序模板一个像素（one pixel per sequencing feature）的愿望。

1.4 HeliScope测序仪

HeliScope测序仪是由Quake团队设计开发的，它实际上也是一种循环芯片测序设备。不过，HeliScope测序仪最大的特点是无需对测序模板进行扩增，它使用了一种高灵敏度的荧光探测仪直接对单链DNA模板进行合成法测序。首先，将基因组DNA切割成随机的小片段DNA分子，并且在每个片段末端加上poly-A尾。然后通过poly-A尾和固定在芯片上的poly-T杂交，将待测模板固定到芯片上，制成测序芯片。最后借助聚合酶将荧光标记的单核苷酸掺入到引物上（图6d）。采集荧光信号，切除荧光标记基团，进行下一轮测序反应，如此反复，最终获得完整的序列信息。根据最近的报道，经过数百轮这种单碱基延伸可以获得25bp或更长的测序长度。HeliScope测序仪的其它特点见表6。

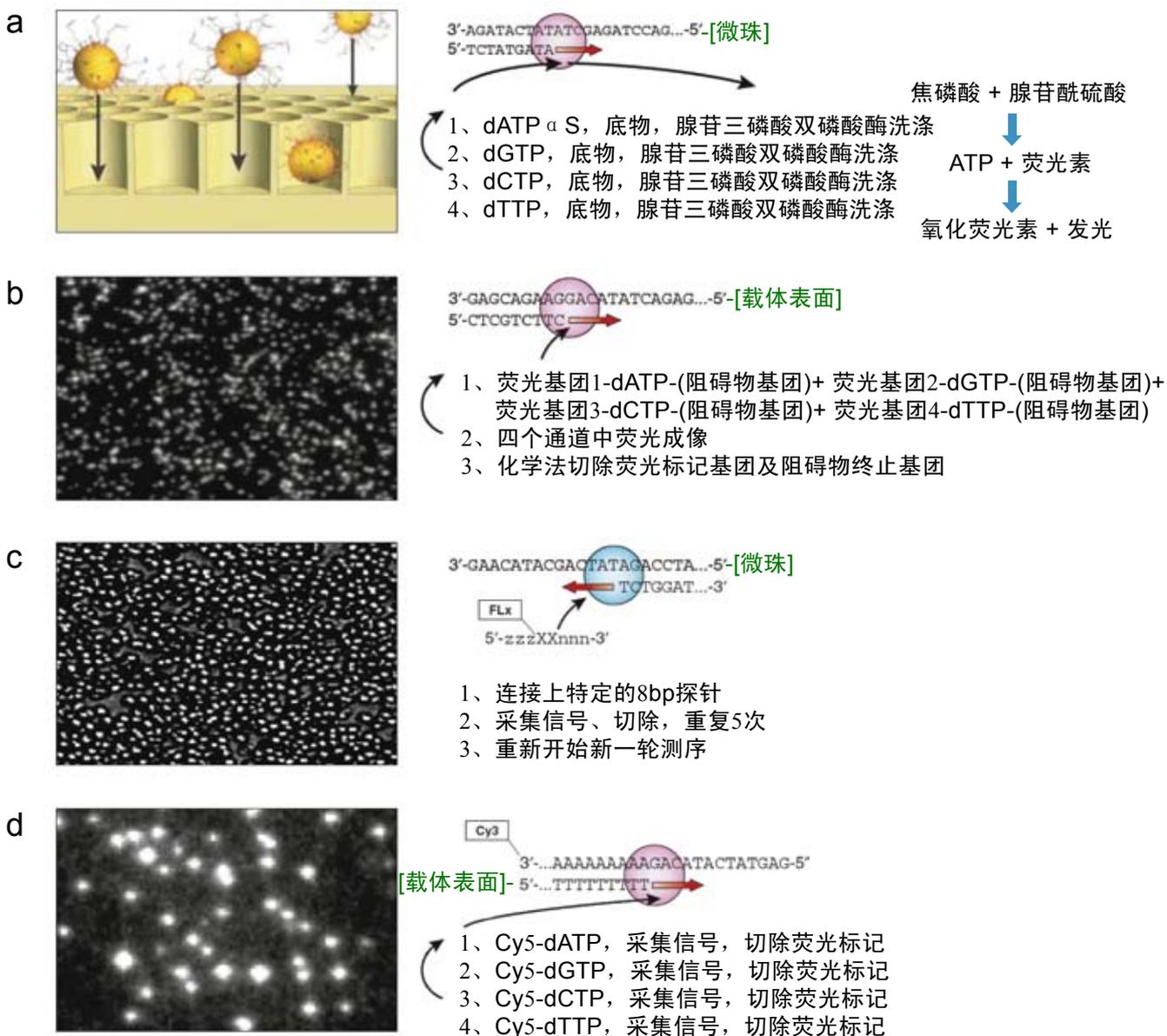


图6 循环芯片测序技术示意图。

(a) 454测序仪使用的方法，经微乳液PCR法扩增后，携带有大量模板分子的微珠被放置到芯片上的微孔中。随后使用焦磷酸法测序，每一轮测序反应都会掺入一个核苷酸，随后加入反应试剂荧光素和5'腺苷酰硫酸，这样在每一个小孔中每当有聚合酶将核苷酸掺入时都会发光。最后用腺苷三磷酸双磷酸酶洗涤去掉多余的核苷酸。(b) Solexa测序仪使用桥式PCR直接在芯片上进行模板片段扩增，然后同时加入四种经过修饰的脱氧核苷酸，每一个核苷酸都携带一种荧光基团和一个可被去除的终止基团。经过修饰的DNA聚合酶催化引物延伸测序反应。采集图像，然后切除荧光标记基团和终止基团，重复上述反应，完成测序。(c) SOLiD测序仪和Polonator测序仪使用微乳液PCR法扩增模板片段，然后吸附有大量扩增片段的直径1 μm的磁珠被制成高密度测序芯片。接着使用连接酶而不是聚合酶测序法完成测序。在SOLiD测序仪中，每一次反应都会在引物末端加上一个荧光标记的8bp的探针，在探针中央的两个碱基上（而不是1个碱基，即双碱基编码技术）标记有荧光基团，探针被连接上之后发出荧光，随后荧光基团部分（zzz）被切除，重新进行下一轮反应。经过几轮这样的测序反应之后可得到一段不连续的碱基序列。然后变性去掉已被延伸的引物，重新结合上新的引物进行新一轮测序反应。(d) HeliScope测序仪，单核苷酸分子不经扩增直接测序。Poly-A尾被添加到DNA文库片段末端，通过与固定在芯片上的Poly-T互补杂交将模板链固定到芯片上，制成测序芯片。模板上标记有Cy3以标识出它们在芯片上的位置。DNA聚合酶将荧光标记的单核苷酸掺入到引物上，采集荧光信号获得序列信息，切除荧光基团，重复上述步骤，完成测序。

表6 HeliScope测序仪的其它几项特点

它和454测序仪一样，测序是不同步的，即每条模板测序的程度不同，有快有慢，测序速度要依据每条模板本身的序列而定，比如在有些循环中有的模板序列就没有相应的合适碱基掺入。不过，由于使用的是单分子测序，因此也不存在相移问题，所以也不会因为相移问题出现错误；

在荧光标记的核苷酸上没有终止基团。比如在454测序仪中，同聚物问题是一个非常重要的问题。而在HeliScope测序仪中则可以很容易地通过控制每一轮测序反应中碱基加入反应的速度来解决这个问题。而且，Harris等人还提到在产生同聚物时连续掺入的标记有荧光的碱基之间会发生淬灭（quenching）现象，这样就无法正确分辨同聚物的长度具体是多长，是A、AA还是AAA；

可以通过“两步法”，即测序两次来提高测序的准确性。这是由于在测序过程中新合成的链同样通过引物固定在芯片上，所以可以变性使新合成的链与模板链分离，重新以反方向再次测序，这样在两次测序的phred样质量值（小词典3）接近30；

由于大量的未标记碱基、不发荧光碱基或污染碱基掺入，所以使用HeliScope测序仪检测缺失突变的出错率最高，一次检测误差率为2%~7%，两次检测误差率为0.2%~1%。不过，检测碱基替换突变的误差率非常低，一次检测误差率为0.01%~1%，两次检测误差率为0.001%，这是目前所有测序仪当中最高的准确率。

原文检索：Jonathan M Rothberg & John H Leamon. (2008) The development and impact of 454 sequencing. *Nature Biotechnology*, 26(10): 1117-1124.

 筱玥/编译

2. 用于处理新一代测序技术数据的软件 and 标准

各种新一代测序仪的飞速发展面临着一个极其重要的问题，那就是生物信息学问题，这些问题包括序列质量评分（sequence quality scoring）问题、序列比对问题、序列组装问题、数据发布问题等。下面将逐个进行讨论。

2.1 序列质量问题

目前，序列质量评分问题是受到广泛关注的一个问题。造成这种现象的原因主要是因为所有新一代测序仪的测序质量都不高，而且不同的序列情况都有各自的误差率。随着新一代测序仪产品的不断成熟，在临床及科研工作中的应用范围越来越广，它们的测序质量也就变得重要起来，而且我们也需要对各个测序仪的测序质量有一个清晰的、可靠的评价标准。由于这个问题还只是刚刚出现，所以我们有可能会设立一个全球统一的、标准化的评价体系对目前现有的以及将来即将出现的测序仪进行评价。我们希望避免再次发生类似过去几个芯片厂家之间进行数据比较的尴尬局面。对于测序仪的应用范围进行标准化的质量评价也是有好处的。比如评价从头测序的质量、评价测序结果与参考序列的相似度、评价测序仪发现突变以及多态性的能力以及对测序仪在进行大规模测序项目研究时的质量可靠性进行评价等。表7列出了几项应该被重点评价的项目。

表7 应该被重点评价的项目

测序仪技术的可重复性；
测序仪的准确性，例如phred样质量值；
测序仪系统误差模式；
在标签计数应用项目中测序仪真实比值的偏差。

这些质量数据都应该以一种简单、标准化的方式包含在测序结果中。现在所有的测序仪器生产商也都在他们的测序报告中加入了测序质量信息，消费者可以借此对数据进行交叉比较，甚至还有可能各取所长，将不同测序仪的测序结果整合起来，获得最佳的测序结果。目前，旨在从短片段测序结果中发现多态性以及突变位点的重测序项目经常会依靠“主要投票机制（majority voting scheme）”。该方法易于操作，但是容易出错，假阴性率较高。诸如Brockman小组和Quinlan小组开发的，更多更好的用于发现单核苷酸多态性的方法是将误差率与单个碱基信号联系起来，即误差率与测序质量和序列内容相关，这样就能获得更准确的结果。我们估计，像phred样质量值之类的评价体系以及“第三方”算法的不断发展，最终一定会诞生一个很好的发现多态性的工具。

2.2 用于分析数据的软件以及生物信息学工具

虽然这些软件和工具都还没有完全商业化，但我们还是能获得一些软件用来分析新一代测序仪的测序结果（表8）。这些软件和工具的功能主要集中在4个方面（表9）。

序列比对和组装是一个颇受关注的问题。虽然BLAST和BLAT等序列比对工具对于长片段的序列比对来说非常有用，但是它们不适合新一代测序仪获得的短片段序列。因此，有大量的针对短测序片段的比对工具出现，这些工具中允许错配或者空隙出现。这些工具中有些借鉴了非常成熟的序列比对算法，例如Smith-Waterman算法等，不过还是有很多新的专门针对短序列片段的算法出现。比如SOAP就是这样一种新型算法，这个软件包能非常有效地进行有空隙或者无空隙的序列比对。该软件包使用了一种内存密集种子（memory-intensive seed）算法和查表格（look-up table）算法来加快比对速度，同时能够对最容易出错的3'端序列进行反复调整。

其它用来加速比对过程的方法还包括“字节编码（bit encoding）”技术。该技术能将序列信息压缩成计算机信息，这样更利于计算机管理和分析。越来越多的比对软件在进行比对时开始考虑数据质量，比如能处理Solexa测序仪测序结果和SOLiD测序仪测序结果的MAQ软件，以及能用最新的“字符空隙转化成彩色空

隙（color-space to letter-space）”的Smith-Waterman算法对SOLiD测序仪的双碱基编码测序结果进行分析的SHRiMP工具（<http://compbio.cs.toronto.edu/shrimp/>）。

表8 可用于分析短片段序列数据的生物信息学工具

软件名称	软件功能	下载地址
Cross_match	比对	http://www.phrap.org/phredphrapconsed.html
ELAND	比对	http://www.illumina.com/
Exonerate	比对	http://www.ebi.ac.uk/~guy/exonerate
MAQ	比对及变异检测	http://maq.sourceforge.net
Mosaik	比对	http://bioinformatics.bc.edu/marthlab/Mosaik
RMAP	比对	http://rulai.cshl.edu/rmap
SHRiMP	比对	http://compbio.cs.toronto.edu/shrimp
SOAP	比对	http://soap.genomics.org.cn
SSAHA2	比对	http://www.sanger.ac.uk/Software/analysis/SSAHA2
SXOligoSearch	比对	http://synasite.mgrc.com.my:8080/sxog/NewSXOligoSearch.php
ALLPATHS	序列组装	
Edena	序列组装	http://www.genomic.ch/edena
Euler-SR	序列组装	
SHARCGS	序列组装	http://sharcgs.molgen.mpg.de
SHRAP	序列组装	
SSAKE	序列组装	http://www.bcgsc.ca/platform/bioinfo/software/ssake
VCAKE	序列组装	http://sourceforge.net/projects/vcake
Velvet	序列组装	http://www.ebi.ac.uk/%7Ezerbino/velvet
PyroBayes	碱基识别	http://bioinformatics.bc.edu/marthlab/PyroBayes
PbShort	突变检测	http://bioinformatics.bc.edu/marthlab/PbShort
ssahaSNP	突变检测	http://www.sanger.ac.uk/Software/analysis/ssahaSNP

表9 表8中的软件和工具的主要功能

将测序结果与参考序列比对；
进行碱基识别以及发现多态性；
将配对测序结果或非配对测序结果从头组装；
基因组浏览和注释。

和序列比对算法的进展情况不同，如何将相对低质量的短片段序列从头组装起来还是一大难题。不过，已经有好几种组装工具开始投入使用了。对于这种短片段序列的从头组装工作来说，配对序列对于整个工作的成功与否影响最大，因此已经有人开始利用配对序列设计新的组装工具了。

2.3 序列数据出版、发行、组织存档原则

目前对于如何组织、存档以及发布这些新一代测序仪产生的短片段序列结果正处于热烈的讨论之中，人们希望制定一个类似芯片试验（microarray experiments, <http://uhts.lbl.gov/>）时制定的MIAME（Minimum

Information About a Microarray Experiment) 规则。这些早期的工作经验在如何处理包括生物学注释信息、临床原始数据、关键试验细节（比如样品特征、样品处理方法）在内的元数据，以及如何处理、出版发行这些数据等方面给了我们良好的建议。如何对这些新一代测序仪的测序结果数据进行公共管理也是一个需要探讨的问题。NCBI最近专门为短片段序列建立了数据库Short Read Archive (SRA)，并同步制定数据提交格式。SRA数据库不仅会收集包括实验注释信息、试验参数等信息的数据，而且还会被整合到Entrez查询系统当中。目前的工作主要包括开发线上搜索工具、数据图形化工具，这些工作有望在近期完成。

3. 新一代测序技术的前景

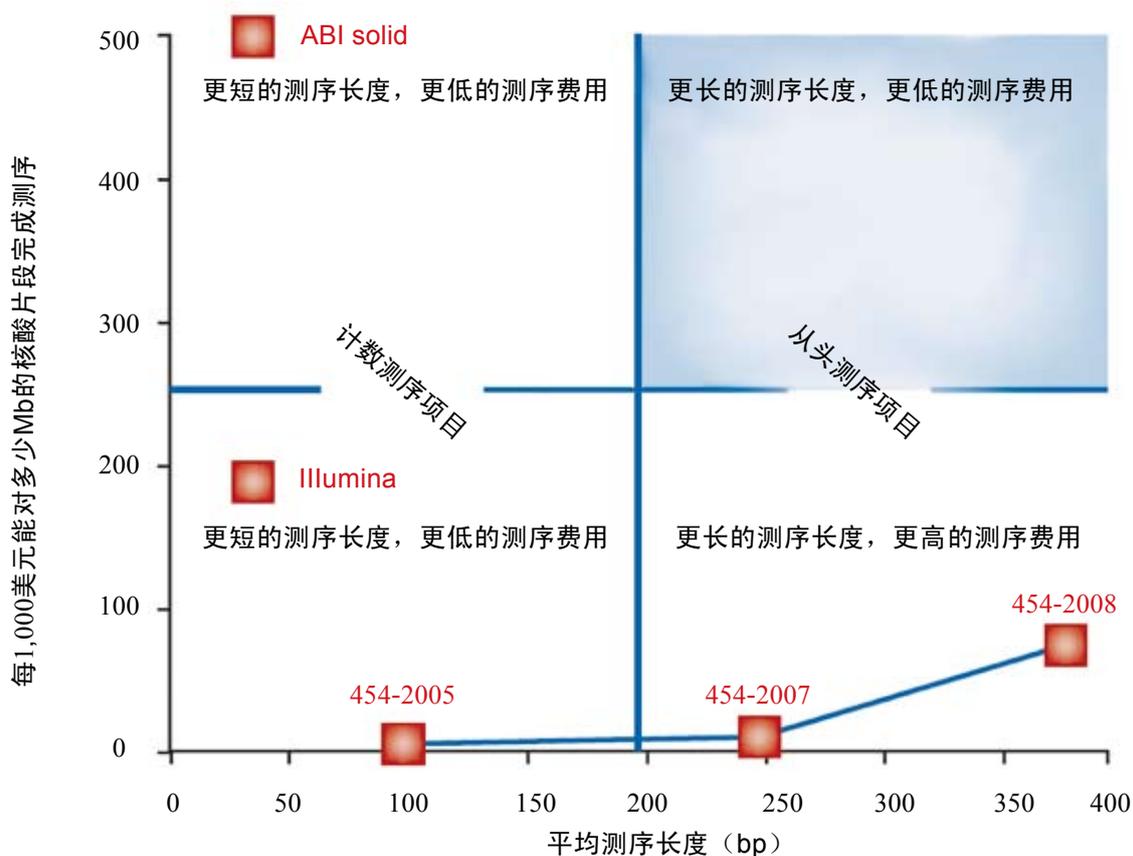


图7 新一代测序仪的比较情况。

对两种不同测序仪从测序长度和测序费用两方面进行比较。Illumina公司的Solexa测序仪和ABI公司的SOLiD测序仪的测序费用数据和测序长度数据来自许多渠道，包括各自生产厂家的网页和最新的技术资料。454公司从2005年至2008年上半年的数据反应了它们在测序长度和测序费用方面的进步。

在2007年6月，James Watson的基因组序列登录到了GenBank数据库当中，这是第一次使用非Sanger测序法获得了人类个体基因组序列，并且第一次将个人基因组序列公之于众。整个测序过程在两个月之内就完成了，花费不到100万美元，这占耗时10年之久的人类基因组计划使用经费的千分之一，同时还是2007年5月在网上公布结果的Venter基因组计划费用的百分之一。我们比较了454测序仪最初的技术参数（每次可以获得两千万碱基序列，测序长度100bp，准确率96%）和用于对James Watson进行测序时的技术参数（每次可以获得一亿碱基序列，测序长度250bp，准确率超过99%），结果发现摩尔定律真的适用于基因组测序领域。

454测序仪和其它的新一代测序仪（图7）一起，展示出了小型化技术和并行处理技术的威力，它们提高了处理通量，降低了测序费用。除了引领新一代测序技术的发展之外，454公司的研发团队还开发了体外DNA文库构建、模板扩增等技术，而且这些技术现在都已经被市场上其它新一代测序仪所广泛使用。很快，

随着计算机技术的飞速发展，个体基因组测序的费用将会由100,000美元降低到10,000美元，继而降低到1,000美元甚至更低。

个人基因组时代马上就要到来了！

4. 新一代测序技术与传统测序技术的优缺点分析

从费用角度、适用范围和限制性来说，传统测序仪和新一代测序仪之间具有明显的差距。因此，对于每一个具体的项目来说，都需要仔细考虑，选择出最合适的测序仪。传统的Sanger测序法适用于对kb~mb长度的DNA片段进行的小规模的测序项目。Sanger测序法相比新一代测序法而言具有极大的“间隔尺寸（granularity）”，既能用于大型项目也能用于小型项目。虽然与传统测序仪相比，新一代测序仪在某些方面很明显地处于劣势，比如在测序长度和准确率方面，但即便如此，在处理大规模的测序项目时大家还是倾向于选择新一代测序仪。

看看新一代测序仪对以往使用传统测序仪进行的生殖细胞突变和体细胞突变研究的帮助就可以认识到它们的作用有多么强大。在这项研究里，使用Sanger测序法除了试剂这一项费用之外，其它的费用也远远超过了使用其它新一代测序仪。这些其它费用包括在96孔板或384孔板中处理样品的费用、电泳费用、大量的生物信息学处理费用以及设备维护人工费用等。研究人员最近对100份样品中的100个基因使用传统测序方法究竟需要花费多少费用进行了一次非正式的调查，假设每个基因平均由10个外显子组成，结果发现整体费用在30万美元至100万美元不等，价格依据测序单位是非盈利的基因组测序中心还是商业化的测序服务机构而不同。很显然，这么高昂的费用对于任何一个实验室来说都是难以承受的。新一代测序仪除了能将测序费用降低好几个数量级之外，它们还具有所需仪器设备少的优点，不过新一代测序仪在后续数据处理方面会碰到问题。

各款新一代测序仪之间也有非常明显的差异（表10），它们都有各自“拿手”的绝活（表11）。有一些测序项目，比如重测序（resequencing）对于测序仪的测序长度要求就没有从头测序的要求高。对于需要依靠标签计数（tag counting）的测序项目，例如在定量分析蛋白质与DNA之间的相互作用时，我们就会更加需要能将待测片段分割成尽量多、尽量小片段的测序方法。测序的准确度和各自相对拿手的项目，比如是善于发现插入、缺失突变还是善于发现碱基替换突变也是需要着重考虑的问题。另外，在进行从头测序或发现结构性变异的研究时使用的配对测序法已经广泛应用于各种新一代测序仪当中。这时，这些配对的模板片段在芯片上的分布情况，比如相互之间的距离远近等就是需要重点考虑的问题了。

表10 新一代DNA测序平台

新一代测序仪名称	模板扩增方法	合成测序方法中使用到的酶和试剂	测序费用（美元）/Mb	设备价格（美元/台）	是否末端配对	发生1°错误的突变类型	测序长度
454	微乳液PCR	聚合酶（焦磷酸测序）	~60	500,000	是	插入或缺失突变	250bp
Solexa	桥式PCR	聚合酶（可被切除的终止基团）	~2	430,000	是	替换突变	36bp
SOLID	微乳液PCR	连接酶（由双碱基编码技术设计的8bp长度探针）	~2	591,000	是	替换突变	35bp
Polonator	微乳液PCR	连接酶（九聚物）	~1	155,000	是	替换突变	13bp
HeliScope	单个分子检测	聚合酶（非同步延长）	~1	1,350,000	是	缺失突变	30bp

注：DNA测序领域的快速发展使得对各类测序方法的价格及读长的评估在很短时间内便失去意义。Roche Applied Science、Illumina及Applied Biosystems公司目前都在不断推出新的产品。表中列出的测序费用只是对使用的反应试剂费用的一个估算。测序长度指的是单链长度。

最后，需要考虑的当然是价格因素，各个新一代测序仪的费用都不相同，作为消费者，当然希望各个测序仪生产厂家之间的竞争更加激烈一点。单纯比较每个碱基的测序费用是一个不错的选择方法，不过有时这也会误导我们，比如准确率更高的方法当然费用会高一些。

表11 新一代测序技术的应用

应用范围	应用举例
全基因组测序 (Complete genome resequencing)	人类个体基因组多态性及突变的全面检测
约化表示测序法 (Reduced representation sequencing)	大规模多态性检测
靶向再测序 (Targeted genomic resequencing)	靶向多态性及突变检测
末端配对测序 (Paired end sequencing)	遗传及获得性结构变异检测
环境基因组测序 (Metagenomic sequencing)	传染性及其共生菌群检测
转录组测序 (Metagenomic sequencing)	定量基因表达及选择性剪切；转录注释；转录SNPs或体细胞突变检测
小RNA测序 (Small RNA sequencing)	microRNA表达谱
酸性亚硫酸盐标记DNA测序 (Sequencing of bisulfite-treated DNA)	基因组DNA中胞嘧啶甲基化模式的测定
染色质免疫沉淀测序 (ChIP-Seq)	全基因组蛋白质与DNA相互作用图谱
核酶片段及测序 (Nuclease fragmentation and sequencing)	核小体定位
分子条码 (Molecular barcoding)	多个体来源样品的多通路测序

5. 总结

过去几年间，新一代测序技术获得了突飞猛进的进展，同时有好几款使用大规模平行循环芯片测序技术的测序仪得到了广泛的应用。这几款测序仪虽然使用的技术有所差异，但是在测序数据的质量和数量方面都有着同样的特征，因此也都面临着同样的试验设计、数据分析和注释的问题。不过，这些新一代测序仪将以往的测序费用降低了好几个数量级。鉴于此，以前只有大型测序中心才能够开展的项目，现在在小型实验室里也能顺利进行了。由于新一代测序仪的出现，测序研究领域也开始升温，有些研究团队正在努力开发新的测序技术希望能够取代现有的新一代测序仪。按照目前的发展速度，我们很难估计几年之后的情况。不过，能够预计的是，下、下一代或者说是第三代测序仪一定会像十年前的芯片技术一样，迅速地普及开来，从而成为常规的技术。希望人们不仅关注测序技术本身的发展，更加关注如何利用测序技术来揭开生物学和医学上的众多谜团。

原文检索：

Jay Shendure & Hanlee Ji. (2008) Next-generation DNA sequencing. *Nature Biotechnology*, 26(10):1135-1145.

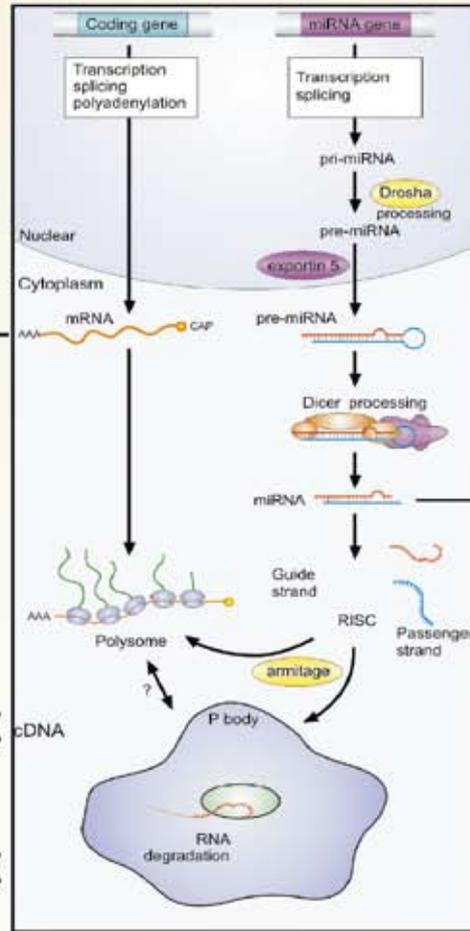
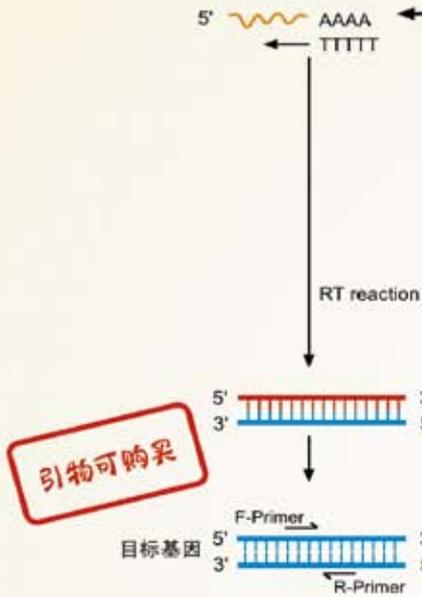
Jonathan M Rothberg & John H Leamon. (2008) The development and impact of 454 sequencing. *Nature Biotechnology*, 26(10): 1117-1124.

Real Time PCR检测服务

mRNA水平基因的检测

优质可靠的反应试剂

独特的引物设计技术

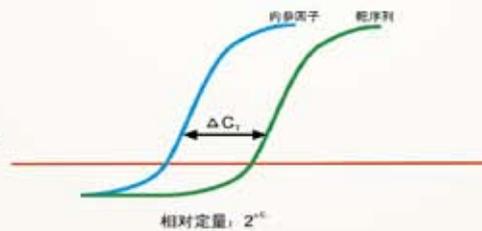
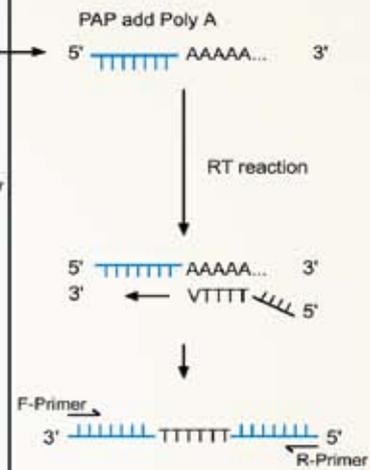


mature-miRNA的检测

严格的服务流程

快速、可靠的检测过程

经济、高通量的检测



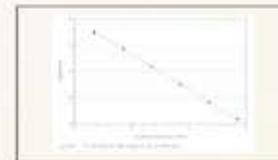
RNA提取质量好



可在宽广范围内
精确定量



扩增产物单一



扩增效率~100%

- 相关产品: 1、经验证的OmicsAmp™ qPCR Primer
2、经验证的OmicsAmp™ miRNA Primer
3、All-in-One™ qPCR Mix

- 4、miRNA qPCR Detection Kit



GeneCopeia, Inc.

19520 Amaranth Drive
Germantown, Maryland 20874
USA
Tel: 301-515-6982
Fax: 301-515-6983
Web: <http://www.genecopeia.com>



广州复能基因有限公司

地址: 广州科学城国际企业孵化器D座8楼 510663
电话: (020)32052376、32052410、32290874
传真: (020)32052877
网址: <http://www.fulengen.com>
服务热线: 020-32068595 020-32052376-666
电子信箱: sales@fulengen.com

1. Polony芯片

该方法是由瑞士日内瓦雪兰诺制药研究中心的Eric Kawashima、俄罗斯科学院的Alexander Chetverin和当时在美国哈佛大学的密特拉共同开发的。他们制作出了可排列在载玻片或凝胶分子层上的一个个独立的聚合酶群落，成为polony，每一个聚合酶群落里都含有一个DNA模板分子，通过PCR反应就可以获得大量的模板。这些扩增后的聚合酶群落就像菌落一样一个个散布在载玻片或凝胶分子层上，成为polony芯片。每一个polony直径约为1 μm，一个芯片上可以承载数十亿个聚合酶群落。

2. 元基因组 (metagenomics)

元基因组又称环境基因组学，是指同时被研究的整个微生物群落的DNA。

3. Phred样质量值 (Phred-like quality scores)

Phred样质量值最初是被定义用来衡量自动DNA测序质量的。Phred样质量值被广泛用来衡量DNA测序仪的测序质量好，并且被用来比较不同测序方法之间的效率高低。

Phred样质量值	在多少个碱基中掺入1个错误碱基数	测序准确率
10	10	90 %
20	100	99 %
30	1,000	99.9 %
40	10,000	99.99 %
50	100,000	99.999 %

四、新型纳米孔测序技术

新型纳米孔测序法 (nanopore sequencing) 是采用电泳技术，借助电泳驱动单个分子逐一通过纳米孔来实现测序的。由于纳米孔的直径非常细小，仅允许单个核酸聚合物通过，因而可以在此基础上使用多种方法来进行高通量检测。此外，纳米级别的孔径保证了检测具有良好的持续性，所以测序的准确度非常高。对于长达1,000个碱基的单链DNA分子、RNA分子或者更短的核酸分子而言，根本无需进行扩增或标记就可以使用纳米孔测序法进行检测，这使得便宜、快速地进行DNA测序成为可能。如果对现有纳米孔测序法进行进一步发展和改进，那么它将有成为第三代测序技术（也可称为下、下一代测序技术），从而帮助人们实现24小时内只花费1,000美元完成二倍体哺乳动物基因组测序这一目标。

一个盛满电解质溶液的容器被一纳米孔膜隔成两半，如果施以比较小的电压，如约100mV电压，就能使用标准的电生理检测手段测量通过纳米孔的电流大小。很多生物电通道的开关都是靠小肽段分子是否堵塞通道来实现的。基于这个事实，加州大学圣克鲁兹分校 (University of California Santa Cruz, UCSC) 的Deamer和哈佛大学 (Harvard University) 的George Church都不约而同地提出一个构想：如果DNA分子或者RNA分子也能堵塞某个通道，那么应该可以运用上述方法来检测电流。接下来，Deamer和Branton等人证明了单链DNA和RNA分子能通过蛋白质组成的孔道，并且能检测到它们通过这种纳米级孔道时所造成的电流改变 (图8a)。他们使用的孔道蛋白是金黄色葡萄球菌α溶血素 (Staphylococcus aureus toxin, α-hemolysin)。这种蛋白以前曾被Bayley小组用作生物传感器。Bayley小组发现，α溶血素蛋白非常稳