

系统合成生物学小组作业展示

(第一组) 文章主题: 新时代的高通量测序技术

小组成员及任务:

- 秦家冉: 翻译文章, ppt制作, 演讲
- 雷正楠: 翻译文章
- 彭崇雄: 翻译文章

目录:

- 简介
- 常用测序辅助技术
- SBL测序方法——SOLiD cPAL
- SBS测序方法——CRT (Illumina CRT系统, GeneReader系统)
SNA (454焦磷酸测序装置, Ion Torrent平台)
- 短读平台的比较
- 长读平台简单概述
- 连接阅读测序与合成长读

一.Short-read NGS

简介；此测序方法有两个类别：SBL(通过连接测序) SBS(通过合成测序)

SBL:与荧光团结合的探针序列与DNA片段杂交，并与相邻的寡核苷酸连接成像，荧光团的发射光谱指示与探针内的特定位置互补的碱基的同一性

SBS: 通过单分子列阵实现在小型芯片上进行PCR反应，应用可逆阻断技术实现每次只合成一个碱基，并标记荧光基团，激发后捕获激发光，读取对应碱基信息

常见辅助测序技术

A. DNA微阵列

即**基因芯片**，在固相表面合成成千上万个寡核苷酸探针再与放射性同位素或荧光标记的**DNA或cDNA杂交**，通过放射性自显影或激光共聚焦显微镜扫描后，对杂交结果进行计算机软件处理，获得杂交信号的强度及模式分布图

B. NANOSTRING

设计一对一对的探针，每对探针均分为**Capture探针（捕获探针）**、和**Report探针（报告探针）**，两种探针均有50个碱基，CAPTURE探针的作用是把目标MRNA给抓住（对应目标MRNA的5'段），并固定到玻璃板上。REPORT探针，与目标MRNA相结合的同时带了一串6个彩色的小珠子，通过小珠子的颜色和排列，光学识别软件可以在后期识别过程当中，把不同的REPORT探针给识别出来，并加以区分、和统计。（每次实验，最多可以测800种探针。也就是说一次最多可以测800种基因的MRNA）

常见辅助测序技术

C. qpcr

实时定量PCR技术，在常规PCR体系中加入荧光基团，**利用荧光信号累计实时监测pce扩增反应中每一个循环产物量的变化**，最后通过标准曲线对未知模板进行定量分析<http://wenku.baidu.com/view/6c0a011ffad6195f302ba605.html>

D, Optical mapping

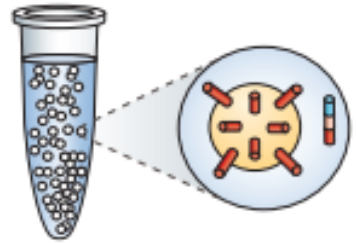
基因组DNA以单个DNA分子的形式固定在长的平行的阵列中，固定的DNA依然保持其原有的静电属性。**DNA被限制性内切酶处理**，切得的限制性酶片段按DNA原来的顺序保持不变。将基因组DNA进行染色、扫描、计算、组装一系列工作，作出全基因组限制性内切酶酶切图谱。

消化的DNA使用荧光染料染色，再置于全自动的荧光显微观察仪采集数据。测量每个DNA分子限制性内切酶片段的大小和顺序，光信号转化成数字信号生成单个DNA分子的限制性内切酶酶切位点图谱。通过DNA分子限制性内切酶酶切位点图谱相互重叠的部分拼接得到全基因组限制性内切酶酶切位点图谱——光学图谱

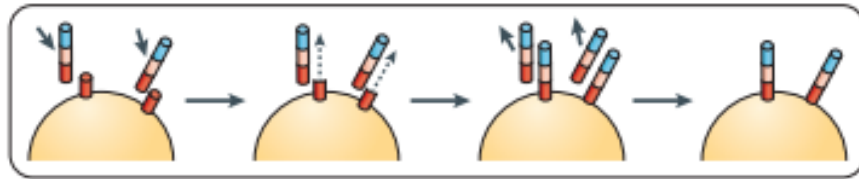
模板群体DNA的克隆策略

- 基于珠的（乳液PCR）：DNA模板产生是样品DNA的断裂，随后连接到普通衔接子用于克隆扩增和测序。对于珠基试剂，**一个衔接子与固定在珠子上的寡核苷酸片段互补**（图1a）。使用乳液PCR 8，DNA模板被扩增，使得多达100万克隆DNA片段被固定在单个9珠上。

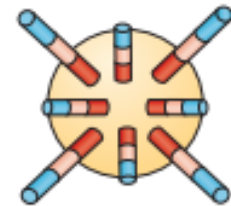
a Emulsion PCR
(454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), Ion Torrent (Thermo Fisher))



Emulsion
Micelle droplets are loaded with primer, template, dNTPs and polymerase



On-bead amplification
Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates



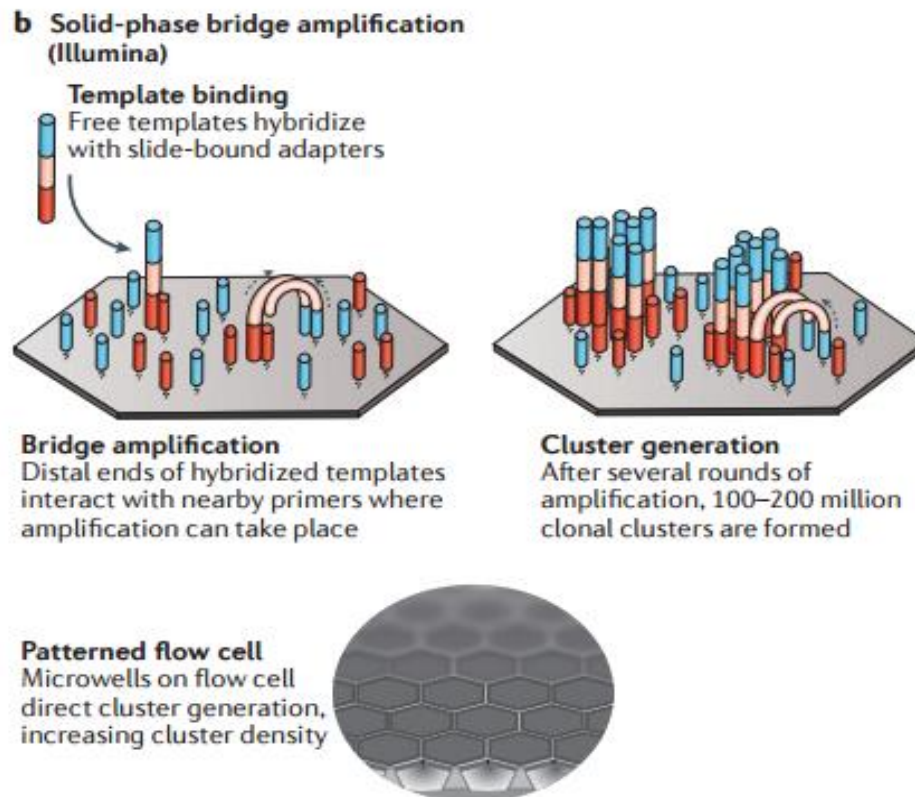
Final product
100-200 million beads with thousands of bound template

模板群体DNA的克隆策略

- 固相桥放大

固态扩增避开使用emPCR有利于直接在载玻片上扩增（图1b, c）。在这种方法中，**正向和反向引物随机地或在图案化的载玻片上共价结合到载玻片表面。**

这些引物提供了单链DNA (ssDNA) 模板可以结合的互补末端



模板群体DNA的克隆策略

- 固相模板扩增

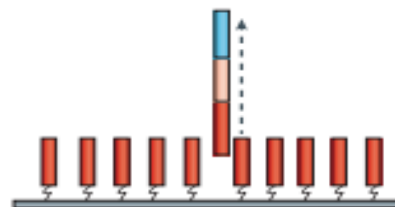
游离DNA模板与结合的引物杂交

第二链被扩增

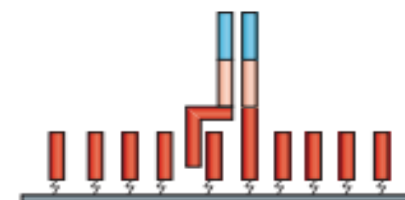
dsDNA部分变性，允许自由端与附近的引物杂交

扩增结合的模板以再生游离DNA模板序列

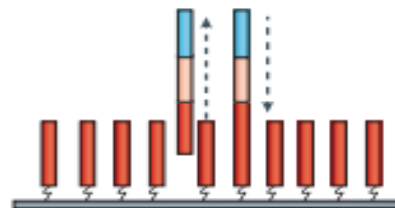
c Solid-phase template walking (SOLiD Wildfire (Thermo Fisher))



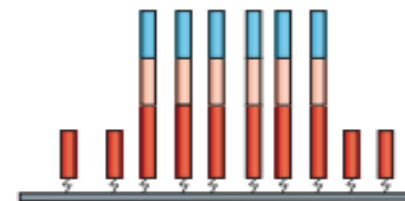
Template binding
Free DNA templates hybridize to bound primers and the second strand is amplified



Primer walking
dsDNA is partially denatured, allowing the free end to hybridize to a nearby primer



Template regeneration
Bound template is amplified to regenerate free DNA templates



Cluster generation
After several cycles of amplification, clusters on a patterned flow cell are generated

模板群体DNA的克隆策略

- 溶液中DNA纳米球产生

衔接子连接将一组衔接子连接到DNA模板的任一末端，随后为模板环化

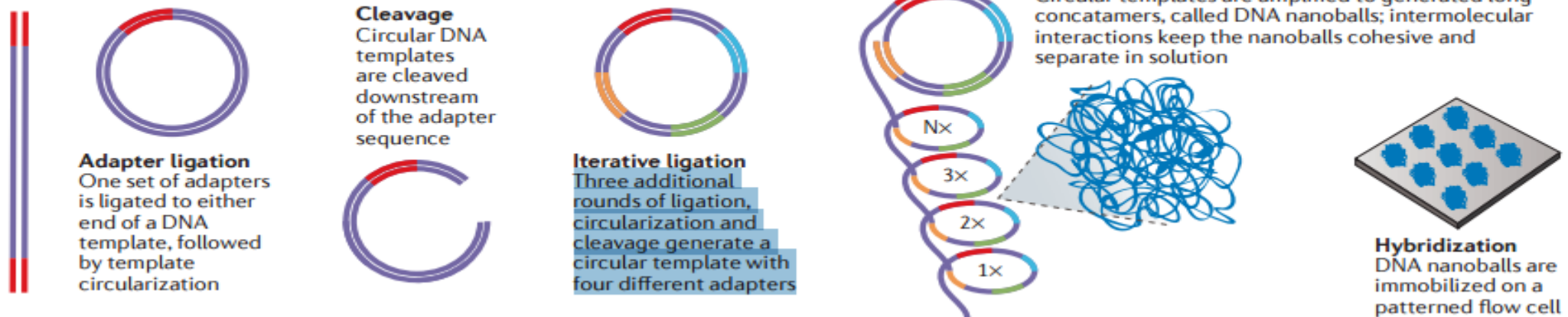
在衔接子序列的下游切割环形DNA模板

另外三轮结扎，环化和切割产生具有四个不同衔接子的环形模板

将环形模板扩增产生长的连接物，称为DNA纳米球；分子间相互作用保持纳米球在溶液中粘性和分离

DNA纳米球固定在图案化的流动池上

d In-solution DNA nanoball generation (Complete Genomics (BGI))



SBL测序方法

- SBL方法包括标记的探针和锚定序列与DNA链的杂交和连接。

探针由一个或两个已知碱基和一系列简并或通用碱基编码，驱动探针和模板之间的互补结合，而锚定片段编码与衔接子序列互补并提供起始连接的位点的已知序列

连接后，对模板成像，并鉴定探针中的一个或多个已知碱基。在完全去除锚-探针复合物或通过裂解以除去荧光团并再生连接位点之后开始新的循环

SOLiD平台使用两个碱基编码的探针，其中每个荧光信号代表二核苷酸。因此，原始输出不与已知核苷酸的掺入直接相关。

因为16种可能的二核苷酸组合不能与光谱可分辨的荧光团单独相关，使用四个荧光信号，每个代表四个二核苷酸组合的子集。因此，每个连接信号代表几种可能的二核苷酸之一，在数据分析期间必须去卷积

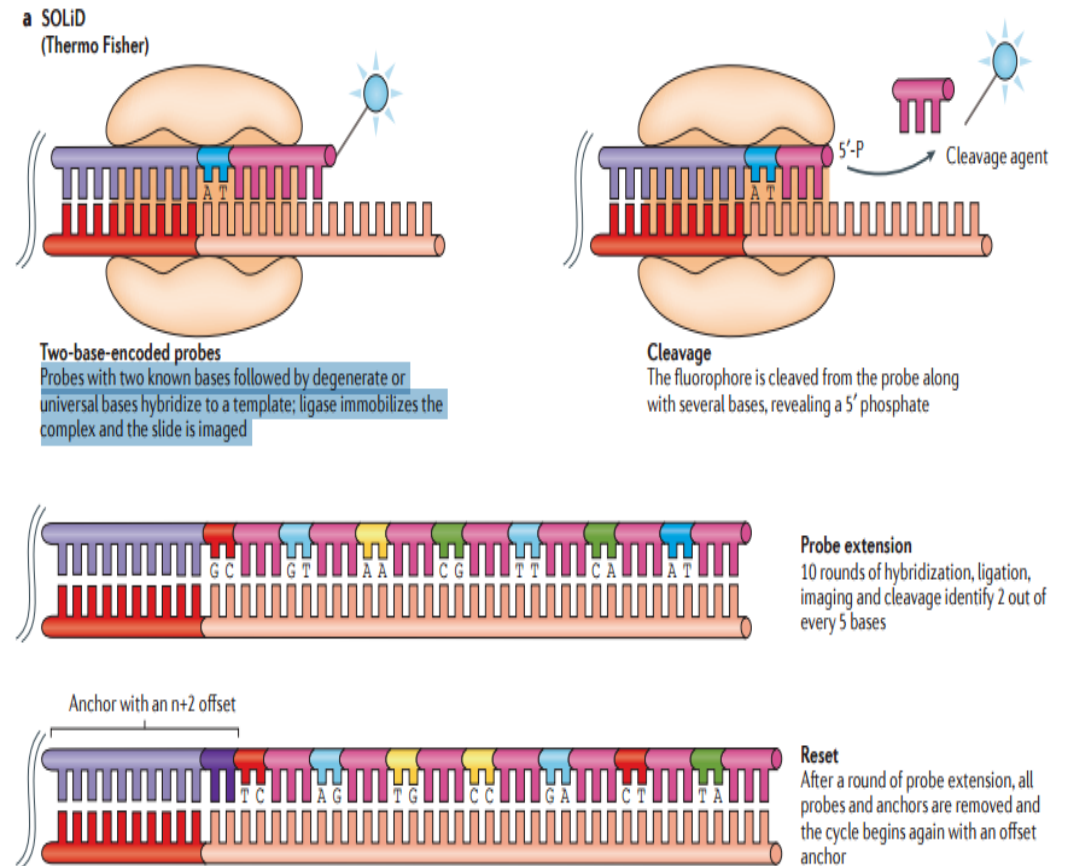
1. SOLiD 平台

探针具有两个已知的碱基，随后简并或通用碱基与模板杂交;**连接酶固定复合物**并对载玻片成像

荧光团与探针一起从几个碱基切割，显示磷酸5'端

10轮杂交，连接，成像和清洗鉴定每5个碱基中的2个

重启阶段经过一轮探针扩展后，所有的探针和锚点被移除，并且循环以偏移锚点再次开始，之后在将得到的数据进行重叠组装



2. cPAL 平台

完整基因组学使用组合探针 - 锚连接 (cPAL) 或 **组合探针 - 锚定子合成进行DNA测序**。在cPAL中 (图2b)，锚定序列 (与四个衔接子序列之一互补) 和探针在几个位置与DNA纳米球杂交。在每个循环中，杂交探针是单碱基编码探针池的成员，其中每个探针含有在恒定位置的已知碱基和相应的荧光团成像后，去除整个探针 - 锚复合物，并且新的探针 - 锚定子组合被杂交。每个随后的循环利用具有在 $n + 1$ 位置中的已知碱基的探针组。该方法中的进一步循环也使用可变长度和化学性质的衔接头，允许在衔接子序列的上游和下游进行测序。

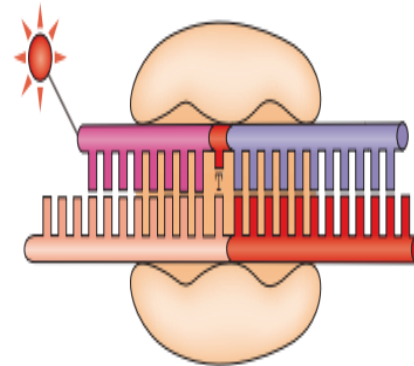
(1) 具有单个已知碱基和简并碱基的探针与模板杂交并成像

(2) 在每个成像步骤之后，去除探针和锚

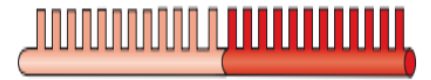
(3) 配对末端测序

(4) 随后的杂交和连接循环使用偏移锚定序列更远的碱基

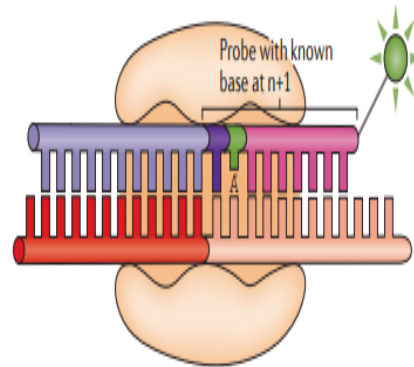
b Complete Genomics (BCI)



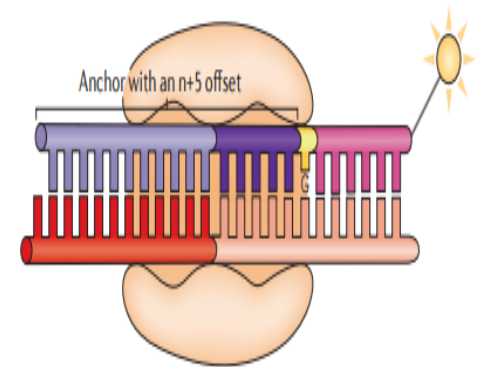
Single-base-encoded probes
A probe with a single known base and degenerate bases hybridizes to a template and is imaged



Reset
After each imaging step, both the probe and anchor are removed



Paired-end sequencing
Sequencing is performed for both the left and right sides of the adapter



Offset anchors
Subsequent rounds of hybridization and ligation use offset anchors to sequence more-distant bases

SBS测序方法

1. 循环可逆终止 (CRT)

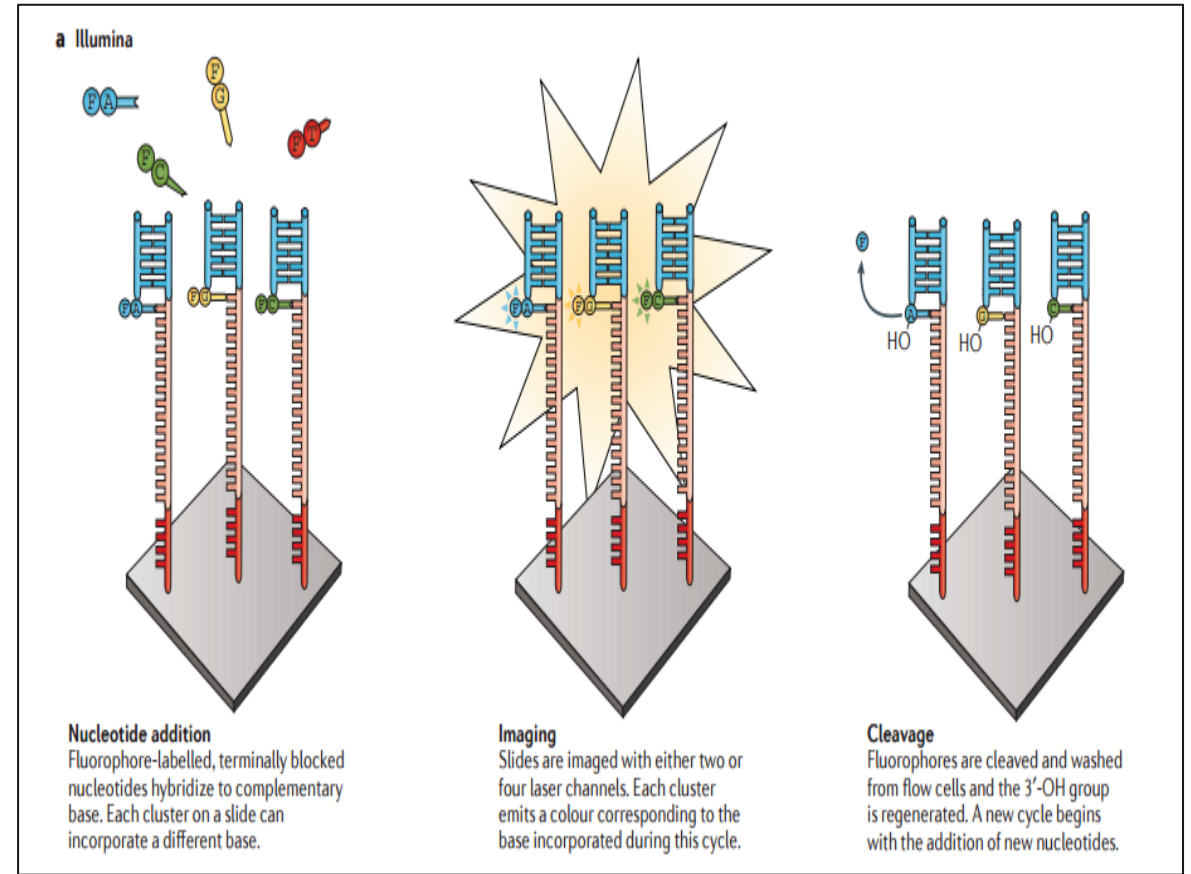
CRT方法通过使用与在Sanger测序中使用的终止子分子类似的终止子分子来定义，其中核糖3'-OH基团被封闭，从而防止伸长。为了开始该过程，DNA模板由与衔接子区域互补的序列引发，其将启动聚合酶结合到该双链DNA (dsDNA) 区

在每个循环期间，加入所有四种单独标记的和3'-封闭的脱氧核苷酸 (dNTP) 的混合物，在将单个dNTP掺入每个伸长的互补链后，除去未结合的dNTP，并将表面成像以鉴定在每个簇处掺入了哪个dNTP。然后可以除去荧光团和封闭基团并开始新的循环

(1). Illumina CRT系统

与其他平台相比，Illumina CRT系统（图3a）占了测序仪器的最大市场份额21。Illumina的短读测序仪器包括小，低通量台式装置到专用于群体水平全基因组测序（WGS）的大型超高通量仪器。

dNTP鉴定通过使用两个或四个激光通道的全内反射荧光（TIRF）显微镜实现。在大多数Illumina平台中，每个dNTP结合到特定于该碱基类型的单个荧光团，并且需要四个不同的成像通道，而NextSeq和Mini-Seq系统使用双荧光团系统

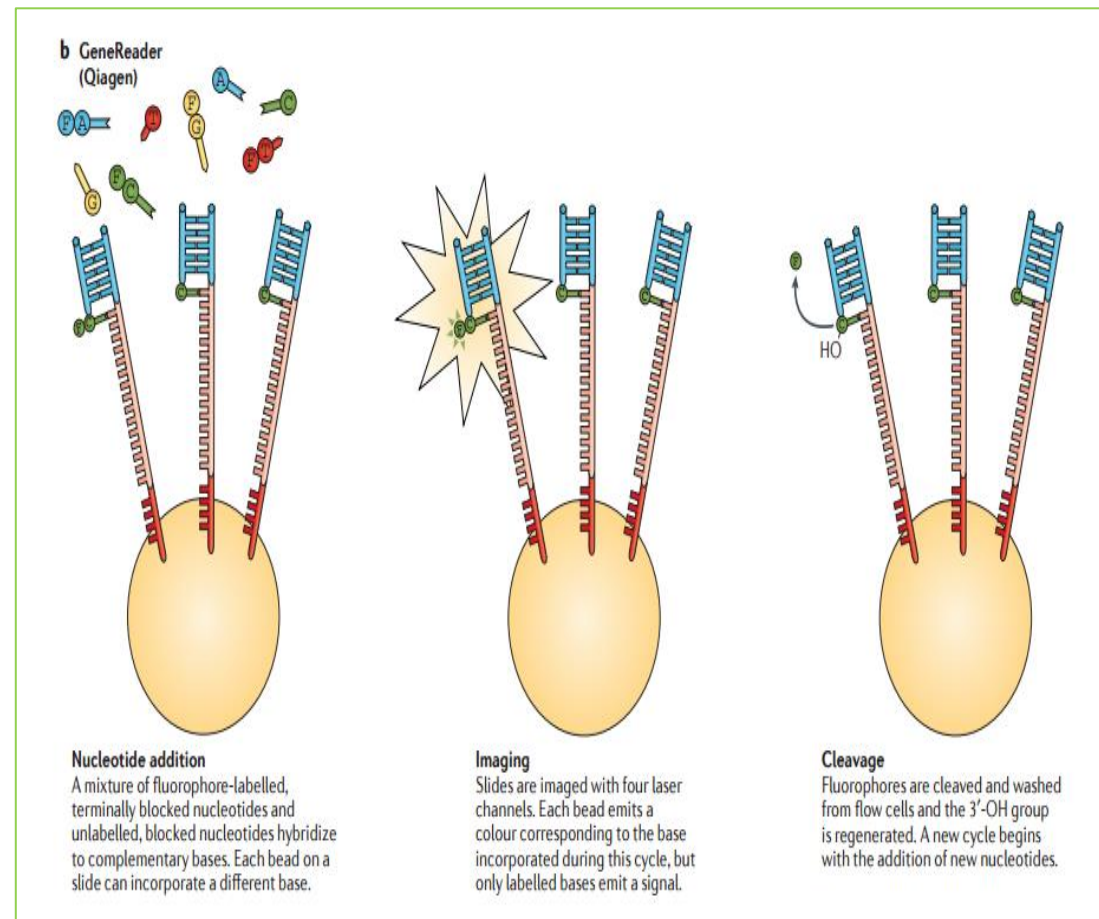


(2) GeneReader 系统

Qiagen收购了Intelligent BioSystems CRT平台，该平台已经商业化，并于2015年作为GeneReader 22重新推出（图3b）。

与其他系统不同，此平台旨在成为一个一体化的NGS平台，从样品制备到分析。为此，GeneReader系统与QIAcube样品制备系统和Qiagen Clinical Insight平台捆绑以进行变体分析。

GeneReader使用几乎与Illumina使用的方法；然而，其目的不是确保每个模板结合荧光标记的dNTP23。相反，GeneReader旨在确保合并足够标记的dNTP以实现鉴定。



2. 单核苷酸添加 (SNA)

概述:

与CRT不同，SNA方法依赖于单个信号来标记dNTP掺入延伸链中。因此，**四个核苷酸中的每一个必须迭代地添加到测序反应中，以确保只有一个dNTP负责信号。**此外，这不需要dNTP被阻断

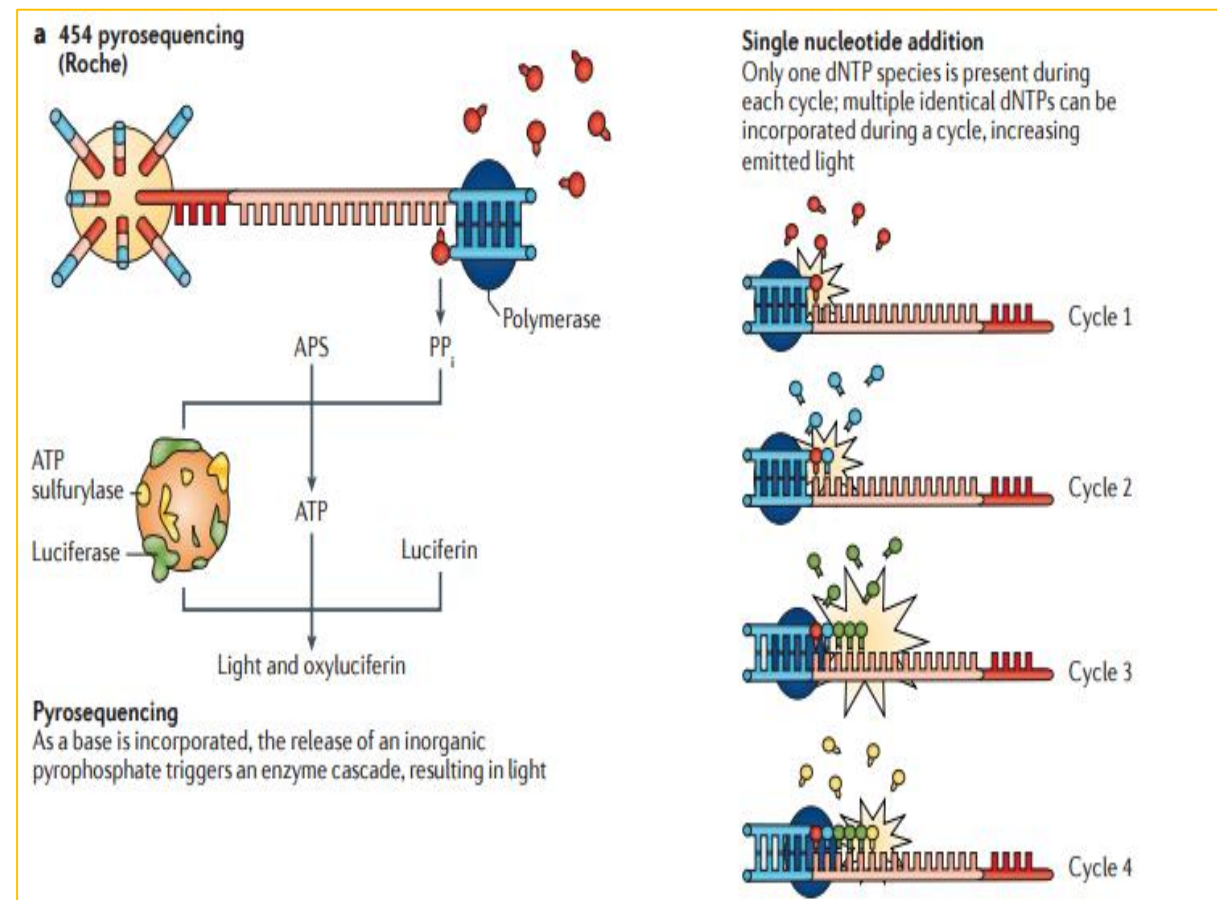
因为测序反应中不存在下一个核苷酸阻止了延伸，但其中添加了相同dNTP的多聚物区域是个例外，其中序列鉴定依赖于并入多个dNTP时信号的比例。

(1). 454焦磷酸测序装置

454焦磷酸测序装置是开发的第一个NGS仪器。

这个SNA系统将模板结合的珠子分布到PicoTiterPlate和含有酶混合物的珠。

当dNTP掺入链中时，发生酶级联，产生生物发光信号。由于在特定珠粒处掺入一种或多种相同的dNTP（图4a），因此通过电荷耦合器件（CCD）照相机检测的每个光脉冲

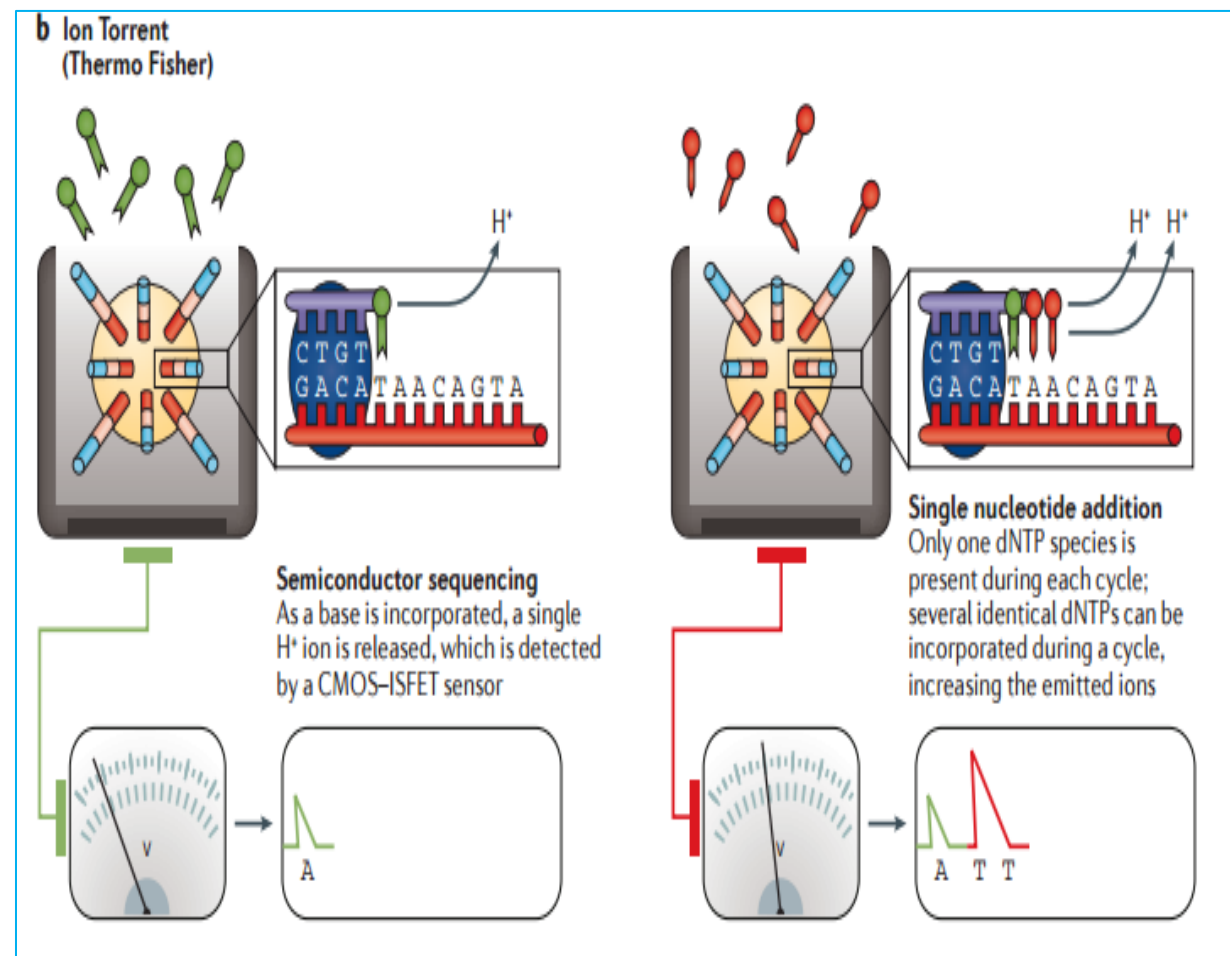


(2) Ion Torrent 平台

Ion Torrent是第一个没有光学传感的NGS平台。

不使用酶级联产生信号，Ion Torrent平台检测在每个dNTP结合时释放的H⁺离子。所得的pH变化由集成互补金属氧化物半导体（CMOS）和离子敏感场效应晶体管（ISFET）检测，（图4b）。

由传感器检测的pH变化与检测到的核苷酸数目不完全成比例，测量多聚物长度的精度有限。



短读平台的比较

个体短片测序平台在吞吐量，成本，错误轮廓和读取结构方面有所不同（表1）。

尽管存在几个NGS技术提供商，NGS研究正越来越多地在Illumina仪器中进行。虽然这意味着他们的数据的高可信度。

但它也引起了对使用单一测序方法26-28导致的系统性偏倚的关注。

如今，研究人员越来越多地选择将多种测序方法与互补的优势整合。

Sequencing by synthesis: CRT

Platform	Throughput	Read Length	Flow Cell	Run Time	Error Rate	Cost (per run)	Cost (per Gb)
Illumina MiniSeq Mid output	150 (SE)*	2.1–2.4 Gb*	14–16 M*	17 h*	<1%, substitution [‡]	\$50,000 (REF. 118)	\$200–300 (REF. 118)
	75 (SE)	1.6–1.8 Gb	22–25 M (SE)*	7 h	<1%, substitution [‡]	\$50,000 (REF. 118)	\$200–300 (REF. 118)
	75 (PE)	3.3–3.7 Gb	44–50 M (PE)*	13 h			
	150 (PE)*	6.6–7.5 Gb*		24 h*			
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M (SE)	4 h	0.1%, substitution [‡]	\$99,000 [†]	–\$1,000
	25 (PE)	750–850 Mb	24–30 M (PE)*	5.5 h			\$996
	150 (PE)	4.5–5.1 Gb		24 h			\$212
	250 (PE)*	7.5–8.5 Gb*		39 h*			\$142 [‡]
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb	44–50 M (PE)*	21–56 h*	0.1%, substitution [‡]	\$99,000 [†]	\$250
	300 (PE)*	13.2–15 Gb*					\$110 [‡]
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb	Up to 260 M (PE)*	15 h	<1%, substitution [‡]	\$250 [†]	\$42
	150 (PE)*	32–40 Gb*		26 h*			\$40 [‡]
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb	400 M (SE)*	11 h	<1%, substitution [‡]	\$250 [†]	\$43
	75 (PE)	50–60 Gb	800 M (PE)*	18 h			\$41
	150 (PE)*	100–120 Gb*		29 h*			\$33 [‡]
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11 Gb	300 M (SE)*	7 h	0.1%, substitution [‡]	\$690 [†]	\$230
	50 (PE)	25–30 Gb	600 M (PE)*	16 h			\$90
	100 (PE)	50–60 Gb		27 h			\$52
	150 (PE)	75–90 Gb		40 h			\$45
	250 (PE)*	125–150 Gb*		60 h*			\$40 [‡]
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution [‡]	\$690 [†]	\$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			\$78
	100 (PE)*	270–300 Gb		11 d*			\$45 [‡]
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution [‡]	\$690 [†]	\$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			\$58
	100 (PE)	360–400 Gb		5 d			\$45
	125 (PE)*	450–500 Gb*		6 d*			\$30 [‡]
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution [‡]	\$740/\$900 (REF. 156)	\$50
	75 (PE)	325–375 Gb					\$31

短读平台的比较

(1) SOLiD和Complete Genomics系统cPAL的Revolocity系统

- SOLiD和Complete Genomics系统使用的SBL技术为这些技术提供了非常高的精度（~99.99%），因为每个基底被探测多次。虽然准确，两个平台也显示敏感性和特异性之间的折衷的证据，使得出现假阳性的结果，使得假变异体被误判。SOLiD平台显示一些替代错误和一些富含GC的低表达。也许这些技术受限于广泛采用的特征是它们**非常短的读取长度**。
- 尽管两个平台都可以产生单端和配对末端测序读取，但SOLiD的最大读取长度仅为75bp，Complete Genomics33的最大读取长度为28-100bp，限制了其用于基因组装配和结构变体检测应用。不幸的是，由于这些限制，以及大约几天的运行时间，SOLiD系统已被降级到行业内的一个小领域。
- 此外，虽然基于cPAL的Revolocity系统旨在与Illumina HiSeq在成本和整个方面进行竞争，但其发布在2016年暂停，并且现在仅可用作人类WGS的服务平台，而基于cPAS的BGISEQ -500平台仅限于中国大陆

短读平台的比较

(2) Illumina系统

- Illumina主导短读排序行业，部分原因在于其作为**技术的成熟度**，高水平的跨平台兼容性及其广泛的平台。可用的仪器套件包括从低通量MiniSeq到超高通量HiSeq X，其能够将~1,800个人类基因组测序至每年 $30\times$ 覆盖。
- **选项的多样化——运行时间，读取结构和读取长度（高达300 bp）。**
- 由于Illumina平台依赖于CRT方法，它对SNA平台中观察到的**同聚物误差更不敏感**。虽然它的总体准确率大于99.5%，该平台在富含AT和富含GC的区域中一些显示比较模糊，有替代错误的倾向

(2) 综述

主体	功能	优点	缺点
双色标签系统	通过减少扫描到两个颜色通道和减少荧光基团使用，提高了速度并降低了成本。	提高了速度并降低了成本。	双通道系统导致对于低多样性样品的稍高的错误谱和性能不佳。
最高通量仪器 HiSeq X			仅限于仅几个应用。
Qiagen GeneReader 全自动二代测序系统	履行和 illumina MiSeq46 类似的作用。		每个 gigabase 测序的成本可能更低。
454 和 Ion Torrent 系统		侧重于重复或复杂 DNA 的应用。	插入和缺失 (indel) 错误占主导地位，

长读平台

- **概述：** 已经显而易见的是，基因组是高度复杂的，具有与进化，适应和疾病相关的许多长的重复元件，拷贝数改变和结构变化。然而，这些复杂元素中的许多都长到短读配对端技术不足以解决它们。
- 长读取测序提供超过几千碱基的读数，允许分辨这些大的结构特征。这样的长读数可以跨越具有单个连续读数的复杂或重复区域，从而消除基因组元件的位置或大小的不确定性。长读也可以用于转录组研究，因为它们能够跨越整个mRNA转录物，允许研究人员确定外显子和脱氧合酶同种型的精确连接。

长读平台

- **类型:** 单分子实时测序方法和合成方法，其依赖于现有的短读技术来构建长读。它们不依赖于短DNA片段，因为它们不需要对信号进行扩增。它们也不需要产生实际的长读；相反，它们使用计算方法来组装库。单分子长读序列（PacBio和ONT）。目前，最广泛使用的长读平台是Pacific Biosciences (PacBio) (图5a) 使用的单分子实时 (SMRT) 测序方法——仪器使用专门的流动池。

A Real-time long-read sequencing

Aa Pacific Biosciences

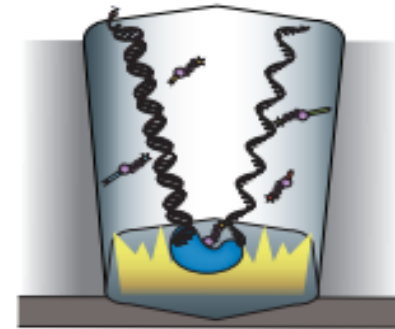
SMRTbell template
Two hairpin adapters allow continuous circular sequencing



ZMW wells
Sites where sequencing takes place



Labelled nucleotides
All four dNTPs are labelled and available for incorporation



Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light

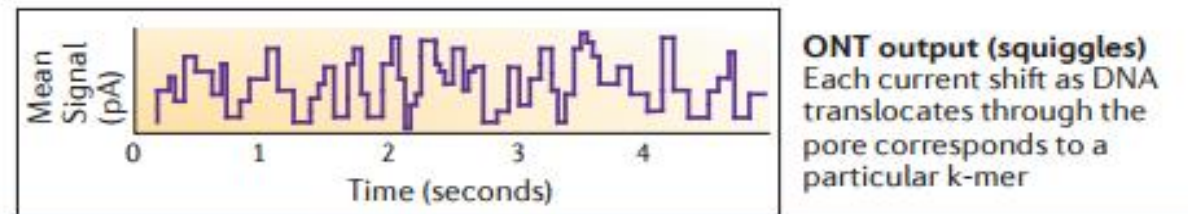
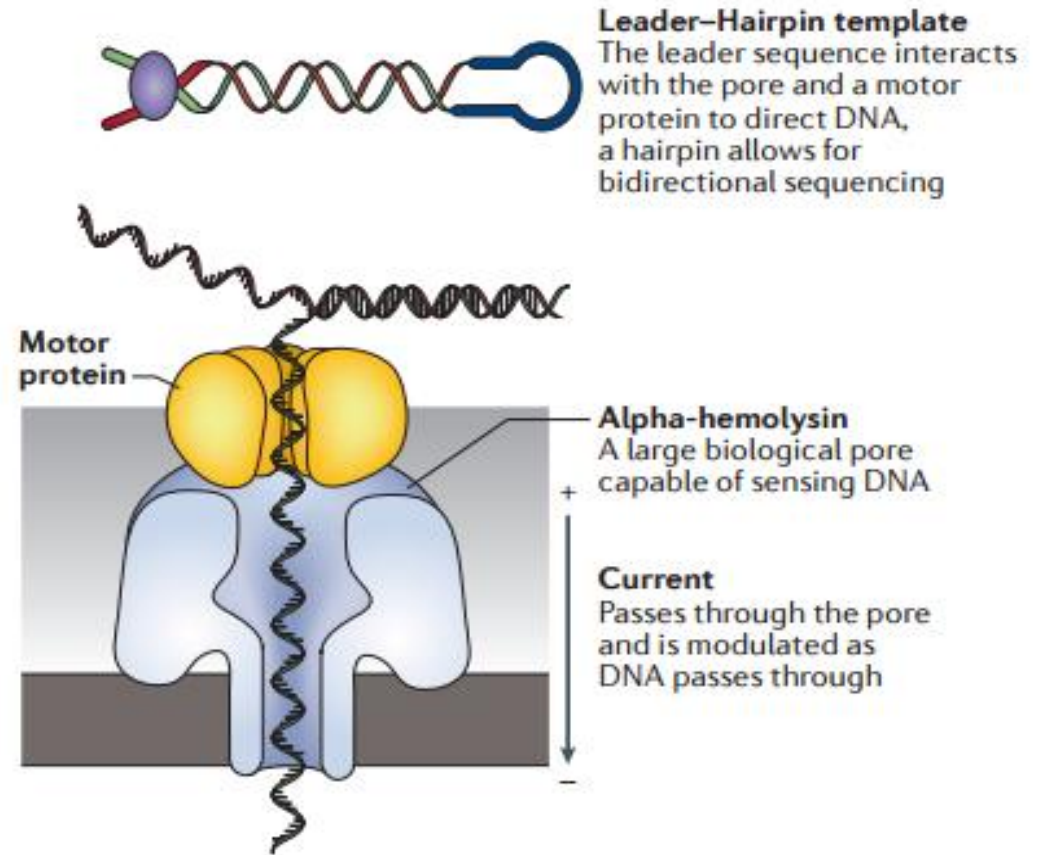
连接阅读

短读SBS技术结合DNA并允许聚合酶沿着DNA模板移动，PacBio将聚合酶固定在底部并允许DNA链通过ZMW。由于固定酶具有恒定的掺入位置，该系统可以集中在单个分子上。因为标记的核苷酸在掺入ZMW的底部期间暂时停止，因此可以通过一个记录发射光的颜色和持续时间的激光和照相机系统连续显现每孔中每个单分子模板上的dNTP掺入。聚合酶在掺入过程中切割dNTP结合的荧光团，允许其在掺入下一个标记的dNTP之前扩散远离传感器区域。

SMRT平台还使用独特的环形模板使得每个模板在聚合酶重复穿过环状分子时被测序多次。尽管对长于~3kb的DNA模板进行多次测序是困难的，但是可以对较短的DNA模板可以作为模板长度的函数而进行多次测序。这些多遍被用于产生插入片段的一致性读取，称为环状共有序列（CCS）。

- 在2014年，纳米孔测序仪的第一个消费者原型 - 牛津纳米技术（ONT）的Minion实现可用。**纳米孔测序仪不监测由模板DNA链引导的核苷酸的掺入或杂交。而其他平台使用次级信号，光，颜色或pH，纳米孔测序仪直接检测天然ssDNA分子的DNA组成。**
- 为了进行测序，当电流通过孔60时，DNA通过蛋白质孔（图5b）。当DNA通过次级运动蛋白的作用移位时，发生电压阻挡，其调制通过孔的电流。这些电荷的时间跟踪称为波形空间，电压的变化是孔中特定DNA序列的特征，其然后可以被解释为k聚体。不仅仅只是具有1-4个可能的信号，该仪器可以产生1000多个——每个可能的k聚体。
- 对于**2~4Tb的数据**在满负荷（火力全开）的设备上只需要**两天左右**的时间。

Ab Oxford Nanopore Technologies



合成长读

- 与真正的测序平台不同，合成长期阅读技术依赖一种用于关联在现有短读序列器上排序片段的条形码系统
- 这些方法将大DNA片段分配到微量滴定孔中或乳液，使得每个分区中存在非常少的分子。在每个分区内，模板片段被剪切和条形码。该方法允许在现有的短读取仪表测序之后，数据被条形码分割，并且与共享条形码的片段（源自相同的原始大片段）进行信息重新组合。类似于早期的技术
- BAC-by-BAC测序，合成条形码读取提供从较大的片段衍生的小片段之间的关联。通过分离片段，可以分离重复或复杂的区域，允许每个都在本地组装。这防止组件中的不可分辨的分支点，其导致断裂（间隙）和更短的组装的连续序列。

(1) Illumina合成长读序列测序平台

- Illumina系统（以前的Moleculo）不需要专门的仪器便可将DNA分成微量滴定板。然而，10X Genomics仪器（GemCode和Chromium）使用乳液分配DNA，并需要使用微流体仪器进行测序反应。只有1ng的起始材料，10X Genomics仪器可将任意大的DNA片段（高达~100kb）分成称为“GEM”的胶束，其通常包含 $\leq 0.3 \times$ 拷贝的基因组和一个独特的条码。在每个GEM内，凝胶珠溶解并且较小的DNA片段从原始大片段扩增，每个具有识别源GEM的条形码。测序后，将读取对齐并链接在一起以在原始片段的跨度上形成一系列锚定片段。

B Synthetic long-read sequencing

Ba Illumina

DNA fragment

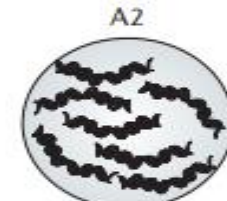
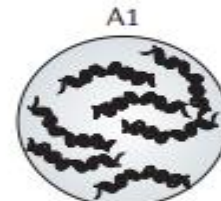
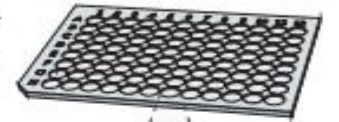
DNA is fragmented and selected to ~10 kb



-3,000 molecules per well

Enzymatic cleavage

DNA is barcoded and fragmented to ~350 bp



Barcodes

DNA from the same well shares the same barcode

Pooling
DNA from each well is pooled and undergoes a standard library preparation



Sequencing
DNA is sequenced on a standard short-read sequencer

(2) 基于10X Genomics乳液的系统

- 与Illumina系统不同，该方法不尝试对单个DNA片段的无间隙，端对端覆盖。相反，它依赖于链接的读，其中衍生自单个长分子的分散的小片段共享公用条形码。虽然这些片段在没有任何覆盖的情况下裂解原始大分子的片段，通过确保在初始制备中存在许多来自相同基因组区域的长片段，从而产生读云，其中来自每个长片段的连接读数可以被堆叠，将它们各自的覆盖组合到总图中（图5d）

Bb 10X Genomics

Emulsion PCR
Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs



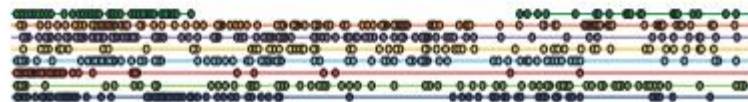
GEMs
Each micelle has 1 barcode out of 750,000



Amplification
Long fragments are amplified such that the product is a barcoded fragment ~350 bp



Pooling
The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation



单分子和合成长扫描序列的比较

(1) PacBio RS II仪器（长读序列中使用最广泛的仪器）

- 其平均读长度为10-15kb，用于揭示复杂的长标记基因组结构64和用于全长转录物测序是理想的
- 有几个显著的限制。长读取的单通错误率高达15%，这些误差随着读取而随机分布,需足够高的覆盖来克服

该仪器的运行时间和吞吐量可以通过控制传感器监测ZMW的时间长度来调节;较长的模板需要较长的时间，且成本较高

(2) ONT Minion

- ONT Minion是一个小型（~3厘米×10厘米的MK1）基于USB的设备，运行在个人计算机，使其成为任何当前测序平台的最小尺寸对于文库制备（例如，热循环器）仍然需要大量的辅助设备，但是文库制备和设备优化的改进可以想到将完全功能的测序系统所需的空间减小到单行李行李的尺寸
- MinION对要排序的片段的大小有很少的约束。理论上，任何大小的DNA分子可以在装置上测序，但是在实践中，当处理超长片段时存在一些**限制70ONT MinION具有大的误码率 - 一维读取可达30%** - 并且由indel误差控制。有效的同聚物测序仍然是ONT MinION的挑战。当同聚物超过k聚体长度时，可能难以鉴定当一个k聚体离开孔和另一个k聚体进入。修饰的碱基也对该装置构成挑战，因为修饰的碱基将改变给定k-mer的典型电压漂移。幸好，在最近**改进的化学和碱基调用算法中提高了精确度**

(3) Illumina合成长期阅读方法

- 吞吐量和误差分布与当前Illumina设备的吞吐量和误差分布相同。然而，由于DNA分裂的原因，该系统需要比典型的短期项目所需的更多的覆盖，因此相对于其他Illumina应用，增加了与该技术相关的成本

(4) 基于10X Genomics乳液的平台

与Illumina合成长读平台相似，依赖于微流体技术，例如，使用乳液。与传统的短读平台相比，乳液平台允许使用少量的DNA，例如，1ng的DNA。乳液平台允许使用少量的DNA，例如，1ng的DNA。

读取之间的位置关系共享相同的条形码会导致模糊，使分析更困难。为了纠正这一点，10X Genomics计划在2016年中期发布Chromium系统。

尽管化学基本上保持不变，但是可能的胶束分隔物的数量将从100,000增加到100万，并且条形码的数量将从750,000增加到~400万。

应用

- **WGS正在成为NGS中最广泛使用的应用之一。** 通过这项技术，研究人员可以获得基因组信息和相关生物学意义的最全面的观点。

例子：2010年，1000个Genomes项目从179个人的WGS发布其初步结果，并定向测序697个个体76。截至2015年，来自26个不同人口的2,504人的基因组已被重建77,78，提供了对人群水平的人类变化的无与伦比的洞察结果，项目序列更大的人正在接近完成或正在进行79-81。人群水平测序被证明是理解人类疾病的重要工具，具有令人兴奋的结果。在一个例子中，Sidore等人82，在2,120个撒丁岛人群上进行WGS，发现脂质水平和炎症标志物的新位点，从而更好地了解驱动血液胆固醇水平的机制。全外显子组和靶向测序也证明了测序研究的价值。

应用

通过限制所用基因组材料的大小，可以在测序运行中对更多的单个样品进行测序。这可以增加基因组研究的宽度和深度。 使用外显子测序，Iossifov等人⁸⁴对超过2,500个单亲家庭进行测序，每个家庭有一个被诊断患有孤独症谱系障碍（ASD）的儿童，他们发现了从头错义突变，从头基因破坏突变和拷贝数变异 ~30%的病例。这和其他有关突变的基因类型为ASD的可能原因的工作提供了一个框架。

NGS还提供了对基因组的调节机制的洞察。 蛋白质-DNA相互作用可以通过富集蛋白质相互作用的DNA片段来探测，通常通过免疫沉淀，如ChIP-seq⁴¹的情况。相反，ATAC-seq使用多活性转座酶从未受蛋白质或核小体保护的区域产生短读NGS相容性DNA片段。

修饰碱基的分析也是可能的。 甲基化包括甲基化DNA⁸⁸的捕获和富集，甲基化或未甲基化区域的选择性消化和/或甲基化碱基的修饰，使得其将SNP引入DNA序列。纳米孔平台还显示出用于直接检测修饰碱基的前景，因为可以通过碱基修饰调节通过孔的电压的特征偏移，允许在不需要化学操作的情况下进行鉴别。

应用

- **NGS最近的范式转变是对非常长的DNA序列进行测序的能力。**历史上重复和复杂区域难以使用短读测序方法94-96组装和解析。最近，使用长读测序技术，Chaisson等人。
- **能够通过间隙闭合和延伸，向人类GRCh37参考基因组添加超过1Mb的新序列，**并且它们鉴定出 $> 26,000$ 个长度 ≥ 50 bp的indels，提供了可用的最全面的参考基因组之一。除了简单改善参考基因组，长的阅读被证明比短的阅读法更有在效地识别临床相关的结构变化。合成方法增强了研究阶段基因组的能力99。在2014年，Kuleshov等人100显示合成方法如何可以减少基因组定相99所需的覆盖率10倍，同时也逐步增加到可以检测到99%SNP。这允许研究人员跟踪跨代突变的**历史——家庭研究的一个重要工具。**
- **转录组研究也从更好的NGS获得中受益。**今天，研究人员利用NGS的力量深入序列到单转录的灵敏度。在2014年，Treutlein等人101证明了单细胞RNA-seq的**功能，长期用于表示不同的细胞群在发育中，国家的组织和发现细胞亚群的新标记。**虽然，长读法转新的剪接同种型51。例如，最近的人类转录组的长读谱显示 $> 10\%$ 的读取表示新的剪接同种型102。

结束语

我们正处在NGS技术的新革命的尖端。超高通量测序的出现推动了仅在几年前不可能进行的研究，并且在临床部门中变得更加广泛。这场革命带来了一系列新的挑战。

由于NGS的目标是在临床环境中无处不在，时间仍然是一个挑战。尽管临床可能受到快速数据缺乏的挑战，但是NGS的其他方面受到大量数据的困扰。NGS在临床上的增殖也引起了与基因组信息相关的效用和伦理的关注。

然而，尽管有这些挫折，新的应用和仪器正以惊人的速度发展。在接下来的几年中，更多的开发者寻求通过新的测序解决方案进一步泛化这个领域。这些现有的和即将出现的NGS工具可能引发生命科学革命性的突破，包括RNA或蛋白质的直接测序，实时基因组病原体监测或基于个人基因组测序的精确医学。