



新一代测序技术简介

www.illumina.com/technology/next-generation-sequencing.html

I. 欢迎使用新一代测序

a. 基因组学科的发展历史

自 20 世纪 70 年代二维色谱法出现以来，DNA 测序已取得了很大的进展。随着 1977 年 Sanger 链终止法的出现，¹ 科学家们获得了以一种可靠、可重复的方式对 DNA 进行测序的能力。十多年后，Applied Biosystems 推出了最早的自动化毛细管电泳（CE）测序仪器，ABI 370（1987 年）和 ABI 3730xl（1998 年）这两款仪器成为了美国国立卫生研究院（NIH）和 Celera 公司牵头的人类基因组计划的主要分析设备。² 尽管在当时这些“第一代”仪器被认为具有较高通量，但随着 2005 年 Genome Analyzer 的出现，测序运行从每次运行 84 千碱基对（kb）提升到了每次运行 1 兆碱基对（Gb）。³ 短 read、大规模平行测序技术是一种完全不同的方法，它彻底改变了测序能力并发起了基因组学科的“新一代”技术。自彼时起，新一代测序（NGS）的数据产出能力每年超过摩尔定律的两倍多（图 1）。

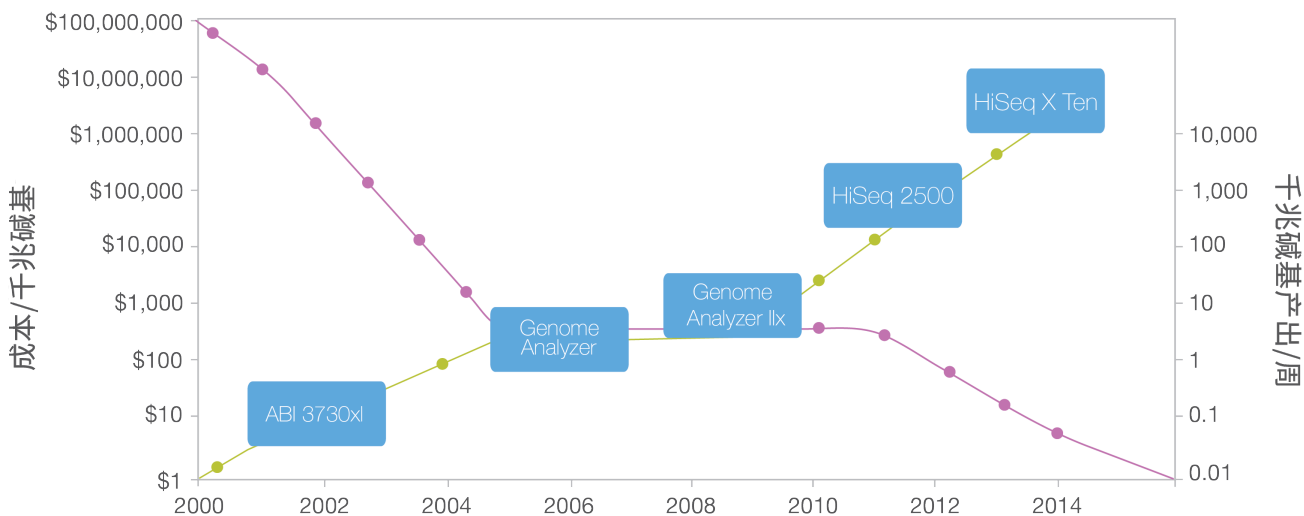


图 1：自 2000 年起的测序成本和数据产出 - 自 2000 年起，数据产出急剧上升而测序成本同时下降。图表两边的 Y 轴为对数。

2005 年，Genome Analyzer 单独运行一次便大约能得到一兆碱基对的数据。截至 2014 年，该比率攀升为单一测序运行得到 1.8 兆兆碱基对（Tb）的数据，令人震惊地增加了 1000 倍。这很明显地反映在了第一张人类基因组图谱的实际情况中，众所周知，该图谱于 2001 年联合发表在《科学》与《自然》杂志上，它需要 15 年的时间来进行测序，而成本接近三十亿美元。与之相反，2014 年发布的 HiSeq X[®] Ten 系统在一天内便可对 45 个人类基因组进行测序，每个花费约 \$1000（图 2）。⁴

除了数据产出急剧增加外，新一代测序技术的引入也转变了科学家们对遗传信息的思考方式。1000 美元基因组可实现群体规模的测序，并将个性化基因组医学确立为标准医学护理的基础部分。如今，研究人员可以在一年内分析成千上万份样本。正如麻省理工学院和哈佛大学下属的 Broad Institute 创始院长兼人类基因组计划主要领导人 Eric Lander 所说：

“进步的速度令人震惊。随着成本继续下降，我们正迈入一个获得疾病基因完整目录的时代。这样我们可以查看数千个人的基因组，并了解他们之间的差异，从而发现那些导致癌症、自闭症、心脏病，或精神分裂症的关键基因。”⁵

每年进行测序的人类基因组

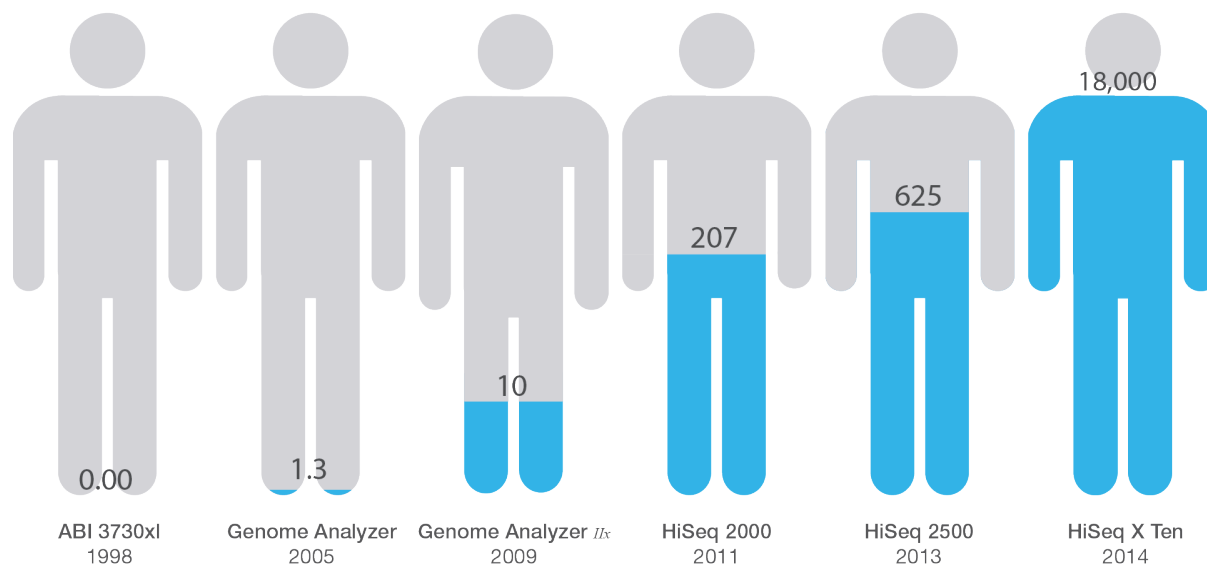


图 2：几十年来的基因组测序 - 自二十世纪 90 年代起，对人类基因组所有 32 亿个碱基（按 30 倍覆盖度）进行测序的能力呈指数级增加。2005 年，随着 Illumina Genome Analyzer 系统的引入，每年便可对 1.3 个人类基因组进行测序。差不多 10 年以后，随着 Illumina HiSeq X Ten 等测序系统的出现，该数字攀升为每年对 18,000 个人类基因组进行测序。

b. NGS 测序化学的基础

大体上来讲，新一代测序技术背后的概念与 CE 测序类似。在 DNA 合成的连续循环中，DNA 聚合酶催化促进荧光标记的三磷酸脱氧核苷酸（dNTP）合并到 DNA 单链模板中。在每次循环期间合并的时候，核苷酸都会被荧光基团激发所识别。关键的区别在于，与单一 DNA 片段测序不同，新一代测序将此过程以一种大规模并行的方式扩展到数以百万的片段。世界上超过 90% 的测序数据均由 Illumina 边合成边测序（SBS）化学产生。* 该技术带来了较高的准确度、较高产出的无错误序列，以及较高比例的 > Q30 碱基检出。⁶⁻⁸

Illumina 新一代测序工作流程包括四个基本步骤：

- 文库制备**—通过 DNA 或 cDNA 样本的随机片段化制备测序文库，然后是 5' 和 3' 接头连接（图 3A）。或者，“标签片段化”将片段化和连接反应合并到单独的一个步骤，极大地提高了文库制备过程的效率。⁹ 然后，对接头连接的片段进行聚合酶链式反应扩增和凝胶纯化。
- 簇生成**—在此过程中，文库被上样到流动槽，在这里文库片段被一层与文库接头互补的结合在表面的寡核苷酸所捕获。通过桥式扩增，将每个片段扩增成不同的克隆簇（图 3B）。当簇生成完成后，模板可立即用于测序。
- 测序**—Illumina SBS 技术利用一种专利的基于可逆终止子的方法，当单个碱基掺入 DNA 模板链时检测它们（图 3C）。由于每个测序循环中存在全部四种可逆终止子结合的 dNTP，天然竞争最大限度减少了掺入偏向，与其他技术相比大大降低了原始错误率。^{6,7} 最后结果是高度准确的逐个碱基测序，即使在重复序列区域和均聚物中，也几乎消除了序列背景特异的错误。
- 数据分析**—数据分析和比对过程中，新识别的序列读段与参考基因组进行比对（图 3D）。比对之后，便可能有很多的分析变化形式，例如单核苷酸多态性（SNP）或插入—缺失（插入缺失）识别、RNA 方法的读段计数、系统发育或元基因组分析等等。



如需观看 SBS 测序化学的详细动画，请访问：
www.illumina.com/SBSvideo。

* 根据 Illumina 公司 2015 年存档数据计算。

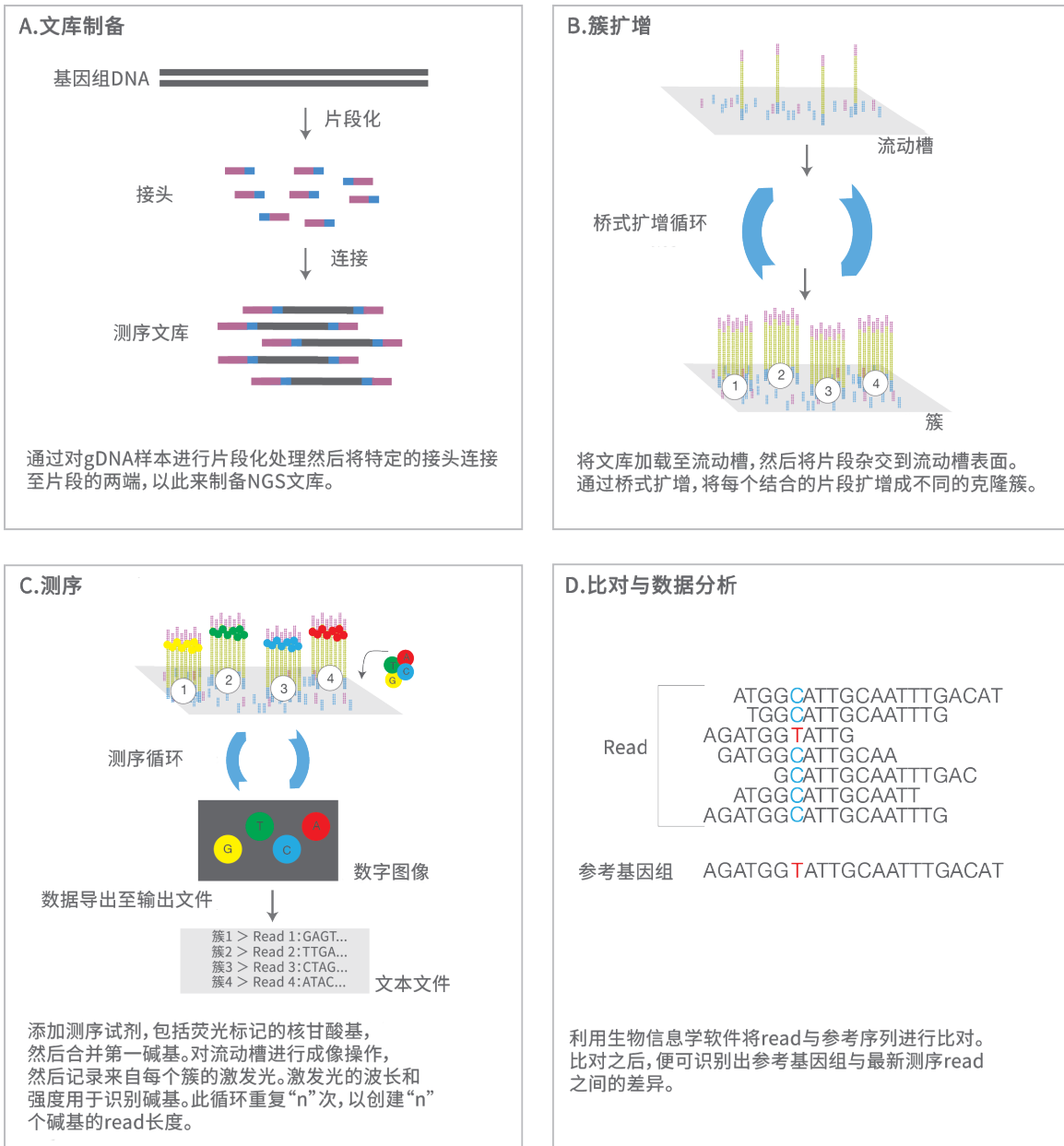


图3：新一代测序化学概述 - Illumina 新一代测序包括四个步骤：(A) 文库制备、(B) 簇生成、(C) 测序以及 (D) 比对和数据分析。

c. 测序技术的进展

双端测序

随着双端 (PE) 测序的发展，新一代测序技术获得了重大进展 (图4)。PE 测序指对一个文库内的 DNA 片段两端分别进行测序，然后将正向和反向 read 作为 read 对进行比对。除了在相同的时间内、开展相同文库制备工作的情况下产生两倍数量的 read 以外，作为 read 对进行比对的序列可以实现更准确的 read 比对，而且还能检测出单端测序数据无法检出的插入缺失。⁸ 差别 read 对位点间隔的分析还可以移除聚合酶链式反应复制品，这是文库制备过程中聚合酶链式反应扩增所带来的常见假象。此外，PE 测序能在 read 对比对后得到更高数量的单核苷酸多态性检出。^{8,9} 尽管单端测序能对诸如小 RNA 测序之类的某些应用起到很好的作用，但是目前大多数研究人员采用双端方法。

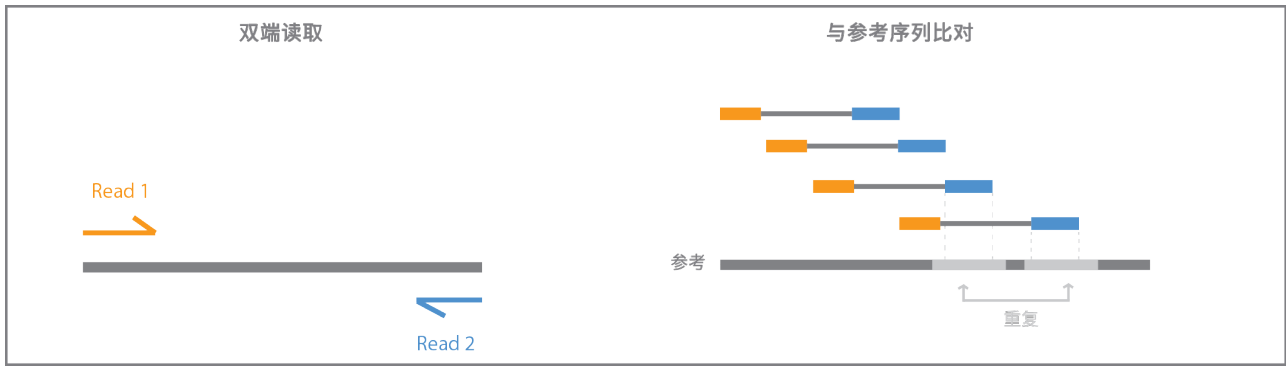


图 4 :双端测序和比对 - 双端测序技术能够对 DNA 片段序列的两端进行测序。因为每对 read 之间的距离是已知的，比对算法可以利用此信息更精确地在重复区域上来标示 read。这能实现更好的 read 比对，特别是在基因组中难以测序的、重复的区域上。

可调覆盖度与无限动态范围

新一代测序的数字化特性，可让诸如基因表达分析之类的 read 计数方法实现几乎无限的动态范围。芯片测量连续的信号强度，检测范围在低端受噪声限制、在高端受信号饱和限制，而新一代测序可以对分散的数字化测序 read 计数进行定量分析。通过增加或减少测序 read 的数量，研究人员可以调节实验的灵敏度，以满足不同研究目标的需求。由于新一代测序的动态范围可调整且几乎无限制，研究人员能够采用比传统芯片方法高得多的灵敏度对微妙的基因表达变化进行定量分析。测序运行可以调整，以较高的分辨率对特定的基因组区域进行放大，或者以较低的分辨率提供更全面的视图。

轻松调节覆盖度水平的功能，实现了多项实验设计优势。例如，体细胞突变可能仅存在于给定组织样本的一小部分细胞内。如果采用肿瘤—正常细胞混合的样本，就必须按极高的覆盖度（通常在 1000x 以上）对隐藏有突变的 DNA 区域进行测序，才能在混合的细胞群中检测出这些低频率的突变。而在覆盖度范围的另一方面，诸如发现基因组变异的应用通常又需要更低的覆盖度水平。在这种情况下，研究设计便会涉及到按较低的分辨率对多份样本（成百上千）进行测序，以便在给定的群体内实现更高的统计功效。

文库制备的进展

采用 Illumina 新一代测序技术，文库制备的流程已经有了迅速改善。最早的新一代测序文库制备实验方案涉及 DNA 或 RNA 样本的随机片段化、基于凝胶的大小选择、特定于平台的寡核苷酸连接、PCR 扩增以及若干纯化步骤。尽管生成这些早期 NGS 文库所需的 1-2 天时间相较于传统克隆技术已是极大的进步，但当前的 NGS 实验方案，例如 Nextera[®] XT DNA Library Preparation，可将文库制备时间缩短到 90 分钟以内。¹⁰ 无需聚合酶链式反应、无需凝胶的试剂盒还可用于灵敏的测序方法。无需聚合酶链式反应的文库制备试剂盒，能在传统意义上具有挑战性的区域，例如高 AT/GC 富集区域、启动子和同聚物区域，实现优异的覆盖度。¹¹



如需了解 Illumina 文库制备试剂盒完整列表，请访问：www.illumina.com/products/by-type/sequencing-kits/library-prep-kits.html。

多重分析

除了每次运行的数据产出增加以外，新一代测序的每次运行样本通量也会随着时间的推移而增加。多重分析允许在单一测序运行中同时对大量文库进行混合与测序（图 5）。采用多重文库时，唯一索引序列在文库制备过程中被添加到每个 DNA 片段两头，以便在最终数据分析前识别每一个 read 并对其分类。采用双端测序和多重分析时，新一代测序已极大缩短了多样本研究得出数据的时间，让研究人员能够快速且简便地从实验得到数据。

由于在最终数据分析之前来自混合文库的测序 read 需要在被称为多重分离的流程中进行计算机化识别和分类，多重分析虽然获得了通量收益，却也增加了额外的复杂性（图 5）。自多重分析研发之时起，多重分析文库之间的索引错配现象便是一个已知的问题。¹² 索引跳跃是索引错配的一个特定原因，它会导致来自预期索引的文库错误分配到池内的另一个索引上，进而导致错配和不准确的测序结果。

如需了解有关索引跳跃的相关信息，包括其发生的机制、Illumina 如何衡量索引跳跃以及缓解索引跳跃对测序数据质量所产生影响，请参阅《索引错配对多重分析和下游分析的影响白皮书》。

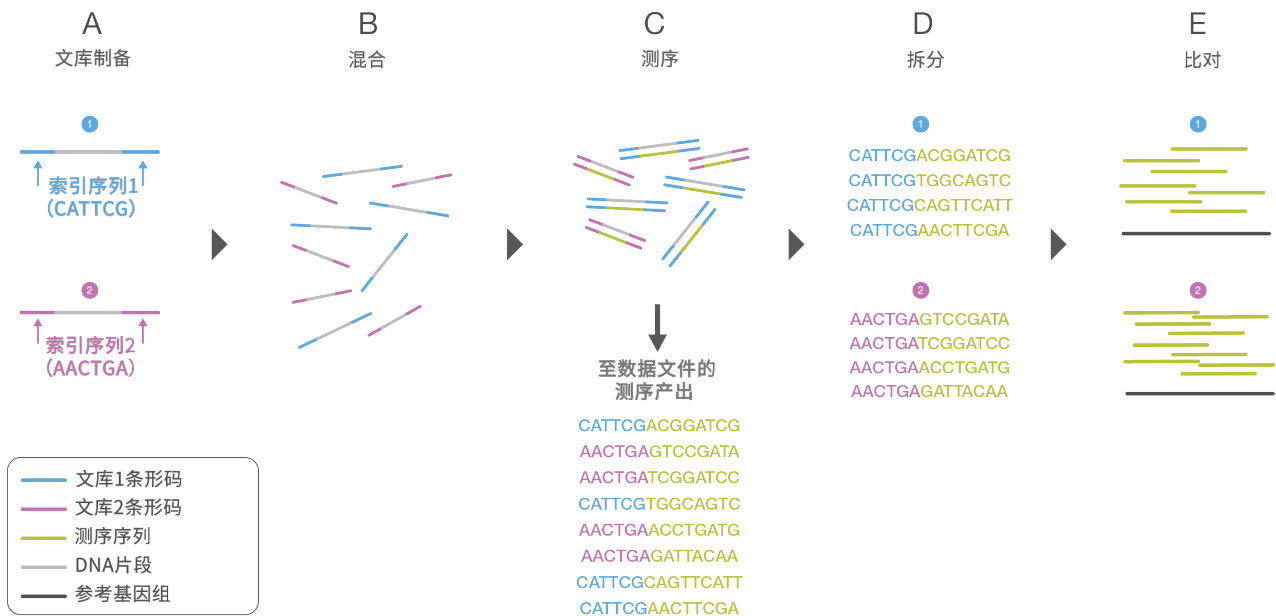


图 5：文库多重分析概述 - (A) 文库制备过程中唯一索引序列添加到两个不同的文库。(B) 文库混合到一起，然后上样到同一流动槽通道。(C) 在单一仪器运行过程中对文库进行同时测序。所有序列均导出至单一的输出文件。(D) 多重分离算法依据 read 的索引将其分类到不同的文件。(E) 每组 read 与相应的参考序列进行比对。

灵活可扩展的仪器

最新的 NGS 平台会产出大量数据，同时具有高度灵活性和可扩展性。测序系统可用于所有研究方法和研究规模，从小型实验室到大型基因组中心均可（图 6）。Illumina 新一代测序仪器囊括从台式 MiniSeq™ 系统到 NovaSeq™ 6000 系统的各款产品，前者的产出范围为 1.8–7.5 Gb，适用于靶向测序研究，后者能够在约两天的时间内生成引人注目的 6 Tb 和 20 B read[†]，适用于群体规模的研究。

灵活的运行配置也是 Illumina 新一代测序仪的设计思路之一。例如，HiSeq® 2500 系统提供两种运行模式以及单个或两个流动槽测序，而 NextSeq® 系列测序系统提供两种流动槽类型，以适应不同的通量要求。HiSeq 3000/4000 系列采用与 HiSeq X 仪器相同的图案化流动槽技术，以实现物美价廉、大规模测序。新款 NovaSeq 系列系统结合最新高性能成像与新一代 Illumina 图案化流动槽技术，以实现通量的大幅上升。此灵活性可让研究人员采用其选择的仪器配置适用于其特定研究要求的运行。



如需深度比较 Illumina 平台，请访问：
www.illumina.com/systems/sequencing.html，
如需探究测序平台比较工具，请访问：
www.illumina.com/systems/sequencing-platforms/comparison-tool.html。



图 6 :适用于几乎所有规模的测序系统 - Illumina 提供创新型新一代测序平台，能够在从小型台式测序仪到大规模测序系统不等的较广泛规模上提供优异的数据质量和准确性。

II. 新一代测序方法

新一代测序平台可采用各类不同的方法，让研究人员能够探究与任何生物的基因组、转录组或表观基因组相关的几乎所有问题。测序方法的不同主要是取决于如何获得 DNA 或 RNA 样本（例如，生物、组织类型、正常的还是受影响的、实验条件等）以及使用什么数据分析选项。不论采用什么方法，测序文库制备之后，实际的测序阶段从根本上来讲仍旧是相同的。有各类不同的标准文库制备试剂盒可供选择，为全基因组测序（WGS）、RNA 测序（RNA-Seq）、靶向测序（例如外显子组测序或 16S 测序）、自定义选择区域、蛋白质结合区域等提供实验方案。尽管新一代测序方法的数量不断增加，但最常见方法的简要概述陈述如下。

a. 基因组学

全基因组测序

基于芯片的全基因组关联分析（GWAS）已成为识别全基因组范围内疾病相关性的常见方法。虽然 GWAS 芯片能够对每份样本检测超过四百万个标记，但是，检测人类基因组 32 亿对碱基最全面的方法是全基因组测序。大幅降低的测序成本和快速生成大量数据的能力，使得全基因组测序成为基因组研究的强大工具。尽管全基因组测序通常与人类基因组测序相关，但该方法可扩展、灵活的特性使其对诸如重要农业家畜之类的任何物种、植物基因组或疾病相关微生物基因组开展测序都同样有用。这一广泛的实用性已在 2011 年欧洲的大肠杆菌爆发事件中得以证实，它促使了一次迅速的科学应对措施。使用最新的新一代测序系统，研究人员能够迅速地对菌株开展测序，这让他们可以追踪病情爆发的源头和传播以及识别引起毒力增加的基因突变。¹³

外显子组测序

外显子组测序是一种广泛采用的靶向测序方法。外显子组占人类基因组的 2% 以下，但包含了大多数已知致病变异，使得全外显子组测序（WES）成为 WGS 经济实惠的替代方案。¹⁴ 采用 WES，基因组的蛋白质编码部分被选择性地捕获与测序。它能高效识别广泛应用的变异，包括群体遗传学、遗传病和癌症研究。

从头测序

从头测序是指在没有任何参考序列进行比对的情况下对某个未知基因组进行测序。测序 read 被组装为 contig，而从头测序数据的覆盖度取决于 contig 的大小和连续性（即，数据中 gap 的数量）。生成高质量从头测序的另一个重要因素是文库内包含插入片段的多样性。将短插入双端序列与长插入配对序列相结合，是实现整个基因组最大覆盖度的最有力方法（图 7）。结合插入片段能够实现范围最广的结构性变异类型的检测，并且对于准确识别更复杂的重排而言至关重要。短插入 read 以更高的深度进行测序，可以填补长插入未能覆盖住的缺口，长插入通常按较低的 read 深度进行测序。因此，混合使用这些方法可以得到更高质量的组装。随着新一代测序技术的进步，很多算法改进已出现在短读数据的测序组装仪上。研究人员可以通过实验室内现有的计算机资源采用测序 read 和可公开获得的短读长组装工具进行从头组装。

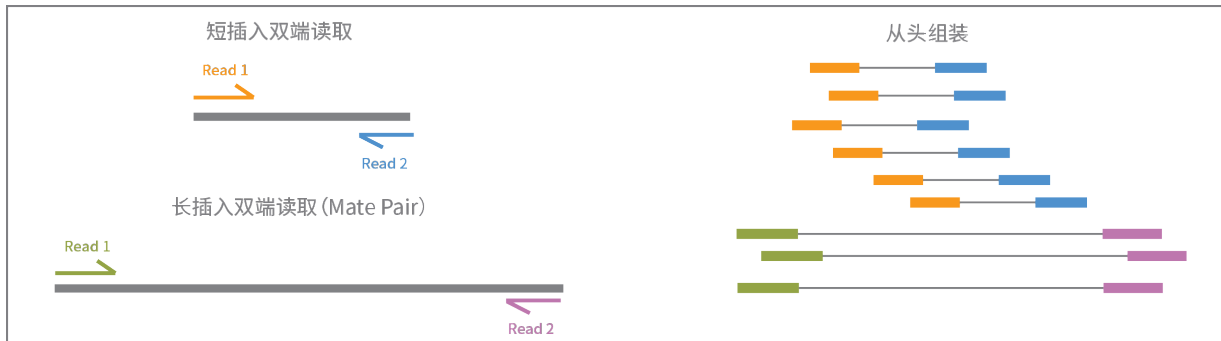


图 7 : 配对和从头组装 - 采用短插入和长插入片段与双端测序相结合可实现从头组装的最大程度的基因组覆盖度。

靶向测序

采用靶向测序，一组基因或基因组区域可被分离出来并对其进行测序。靶向测序让研究人员能够将时间、费用和数据分析集中在感兴趣的特定区域，并在更高的覆盖度水平上开展测序。例如，一个典型的全基因组测序研究能够实现每个基因组 30–50x 的覆盖度水平，而靶向测序项目能够轻易以 500–1000x 或更高的覆盖度覆盖目标区域。这种较高的覆盖度能够让研究人员识别出罕见变异，这些罕见变异通过全基因组测序或 CE 测序识别起来会非常困难且过于昂贵。

可选购预先设计好的靶向测序 panel，也可自定义设计。有各类靶向测序文库制备试剂盒可供选择，包括侧重于特定感兴趣区域（例如，癌症、心肌病或自闭症）的探针组的试剂盒。定制探针组可通过 DesignStudio™ 软件使用，让研究人员靶向与感兴趣研究方向相关的基因组区域。定制靶向测序，是检查特定通路中基因或者从全基因组关联分析或全基因组测序开展随访研究的理想之选。Illumina 目前支持适用于靶向测序、靶向富集和扩增子生成的两种方法（图 8）。

视文库制备试剂盒的参数而定，靶向富集可捕获 10 kb–62 Mb 之间的区域。视所采用的文库制备试剂盒而定，扩增子测序允许研究人员同时对 16–1536 个目标进行测序，跨越 2.4–652.8 kb 的总内容。这种高度多重的方法带来了发现、验证或筛查基因变异的广泛应用。扩增子测序对于发现复杂样本（例如，混合有生殖细胞系 DNA 的癌变肿瘤）中罕见的体细胞突变非常有用。^{15,16} 另外一种常见的扩增子应用是对多种物的细菌 16S rRNA 进行测序，这是系统发育和分类研究中广泛采用的方法，特别是在多样化的宏基因组学样本中。¹⁷



如需了解有关 Illumina 靶向、全基因组测序、外显子组或从头测序解决方案的更多信息，请访问：
www.illumina.com/applications/sequencing/dna_sequencing.html

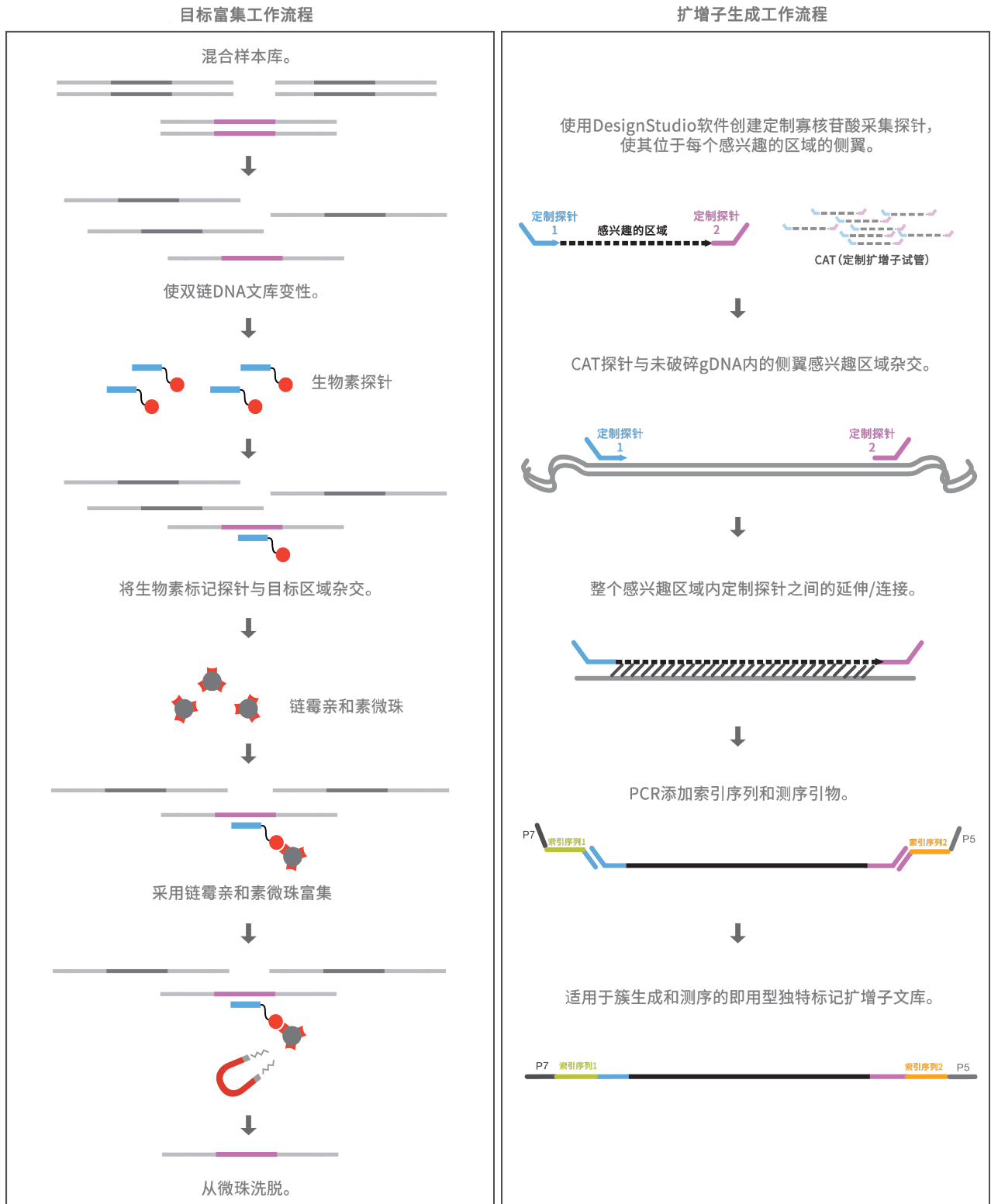


图 8 :靶向富集和扩增子生成工作流程 - 在目标富集中，感兴趣的特定区域通过与生物素标记的探针杂交而被捕获，之后通过磁性分离。扩增子测序采用高度多重 PCR 寡核苷酸组对感兴趣的区域进行扩增和纯化。

b. 转录组

RNA-Seq 的文库制备方法一般以总 RNA 样本制备开始，然后再是核糖体清除步骤。然后，在执行标准新一代测序文库制备之前，将总 RNA 样本转换为 cDNA。侧重于 mRNA、小型 RNA、非编码 RNA 或微小 RNA 的 RNA-Seq，可在 cDNA 合成之前纳入额外的分离或富集步骤而得以实现（图 9）。

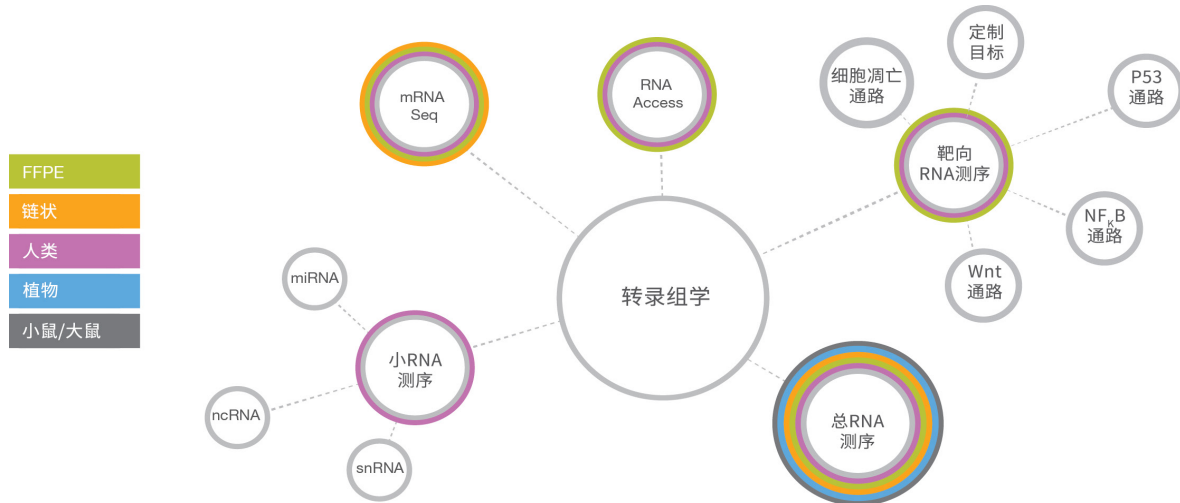


图 9：新一代测序转录组学的完整视图 - 在过去的 10 年中，采用新一代测序开展转录组学分析的各类方法不断涌现，包括总 RNA-Seq、mRNA-Seq、小型 RNA-Seq 和靶向 RNA-Seq。

总 RNA 和 mRNA 测序

转录组测序是基因表达研究的一项重大进步，因为它允许对整个转录组而非预先确定的一组基因提供快照。全转录组测序提供给定生物学时刻的细胞转录谱综合概览，极大地提高了 RNA 研究方法的能力。与任何一种测序方法一样，几乎无限的动态范围可实现对常见和罕见转录本进行识别和定量分析。附加功能包括在可变剪接点比对测序 read 以及检测同种型、新型转录本和基因融合。支持链特异性的文库制备试剂盒既可用于总 RNA 测序方法又可用于 mRNA 测序方法。

靶向 RNA 测序

靶向 RNA 测序是一种测量感兴趣转录本的方法，适用于检测表达差异、等位基因表达、基因融合、同种型、cSNP 和可变剪接功能。Illumina TruSeq® Targeted RNA Sequencing Kit 包括预先配置、实验可验证的 panel，侧重于特定细胞通路或疾病状态，例如细胞凋亡、心脏毒性、NFκB 通路等。可针对感兴趣的特定基因分析设计和订购定制内容。靶向 RNA 测序是用于研究感兴趣的特定通路，验证基因表达芯片或全转录组测序结果的强有力方法。

小 RNA 和非编码 RNA 测序

小 RNA、非编码 RNA 或 microRNA 是一些短至 18–22 bp 的核苷酸，它们通常作为基因抑制因子或抑制子在基因表达调控中发挥作用。由于 microRNA 在转录和转译调控中发挥的作用越来越明显，对其开展的研究也逐步发展。^{18,19}



如需了解有关小 RNA（非编码 RNA）、靶向 RNA、总 RNA 以及 mRNA 测序的 Illumina 解决方案，请访问：
www.illumina.com/applications/sequencing/rna.html。

c. 表观基因组学

基因组学涉及到 DNA 序列中遗传性或获得性变化的研究，而表观基因组学研究的却是由 DNA 序列变化之外的其他机制引起的基因活动遗传性变化。表观遗传活动的机制包括 DNA 甲基化、小型 RNA 介导的调控、DNA-蛋白质相互作用、组蛋白修饰等等。

甲基化测序

表观遗传学比较关键的一个焦点在于，研究整个调控特定区域（例如，启动子或异染色质）的胞嘧啶甲基化（5mC）状态。胞嘧啶甲基化会明显改变时间和空间基因表达以及染色质重塑。²⁰ 尽管存在多种研究基因甲基化的方法，但甲基化测序在单一核苷酸水平评估甲基化状态的同时，还充分利用了新一代测序技术和全基因组分析的优势。这两种甲基化测序方法已被广泛应用：全基因组亚硫酸氢盐测序（WGBS）和简化表观亚硫酸氢盐测序（RRBS）。采用 WGBS 时，亚硫酸氢钠测序化学将非甲基化胞嘧啶转换为尿嘧啶，然后再转换为测序读段或数据产出中的胸腺嘧啶。在 RRBS 中，DNA 被一种不受甲基化状态影响的限制酶 MspI 消化。大小范围为 100–150 bp 的片段被分离出来，用以富集含 DNA 区域的 CpG 和启动子。然后采用标准新一代测序实验方案构建测序文库。



如需了解有关甲基化测序解决方案的更多信息，
请访问：

www.illumina.com/techniques/sequencing/methylation-sequencing.html

ChIP 测序

蛋白质-DNA 或蛋白质-RNA 相互作用对很多生物学过程和病状都会产生显著影响。这些相互作用可以通过将染色质免疫沉淀（ChIP）实验分析方法与 NGS 方法相结合采用新一代测序进行检验。ChIP-Seq 实验分析方法从染色质免疫沉淀步骤开始（由于 ChIP 实验方案必须特定于物种、组织类型和实验条件，所以这些方案千差万别）。



如需了解更多关于 ChIP-Seq 的信息，请访问：

www.illumina.com/techniques/sequencing/dna-sequencing/chip-seq.html

核糖体分析

核糖体分析是一种基于被核糖体保护的 mRNA 片段的深度测序方法。对这些片段的纯化和测序，可提供具体时间点某一细胞内处于活动状态的所有核糖体的“快照”。此信息能确定哪些蛋白质正在细胞内被积极地翻译，并且对于研究翻译控制、测量基因表达、确定蛋白质合成率或预测蛋白质丰度都很有用。核糖体分析能对细胞转译流程进行系统性监控并预测蛋白质丰度。确定转录本中正在被转译的区域有助于定义复杂生物的蛋白质组。采用新一代测序时，核糖体分析能够对蛋白质合成进行详尽而准确的体内分析。



如需了解更多关于 Illumina 核糖体分析的信息，
请访问：

www.illumina.com/applications/sequencing/rna.html

III. Illumina 从 DNA 样本到数据的新一代测序解决方案

a. Illumina 新一代测序工作流程

Illumina 为新一代测序工作流程提供从文库制备到数据分析的全面解决方案（图 10）。所有新一代测序方法，包括全基因组测序、外显子组测序、靶向测序、RNA 测序等，均有文库制备试剂盒可供选择。Illumina 文库制备实验方案可适应各类通量需求，从小型实验室的手动实验方案到更大型实验室或基因组中心的全自动化文库制备工作站，均可适用。同样，Illumina 还能提供全面的测序平台产品组合，从台式的 MiniSeq 和 MiSeq[®] 系统到工厂规模的 HiSeq X 和 NovaSeq 系列测序系统，均能提供与各类实验室或测序中心契合的速度、性能和成本。对于新一代测序工作流程的最后一个步骤，Illumina 提供了用户友好的生物信息学工具，可通过网络轻松访问，或通过现场服务器在仪器上访问。



图 10 :Illumina 从 DNA 样本到数据的解决方案 - Illumina 为包括文库制备、测序和最终数据分析在内的新一代测序工作流程的所有步骤提供完全集成的 DNA 数据化解决方案以及技术和支持。

b. 整合数据分析

来自任何一套 Illumina 测序系统的数据都可以无缝导入 BaseSpace[®] Sequence Hub，这是一款用户友好型基因组云计算平台，能够提供精简的数据管理、分析测序工具和数据存储。经优化的 BaseSpace Sequence Hub 可自动处理生成的大量数据。研究人员将能够享受到一个丰富的生态系统，包括商用工具和开源工具，这些数据分析工具源自 Illumina 和第三方开发者，包括比对和变异检测、注释、可视化、判读以及体细胞变异检出。BaseSpace Onsite Sequence Hub 是 BaseSpace Sequence Hub 的本地版本，它可以通过一台安装好的本地服务器实现现场数据存储和分析。可在仪器上直接访问的 BaseSpace Sequence Hub 能够实现多个工作流程步骤的整合，包括采用 BaseSpace Prep 进行文库制备计划，[‡] 运行设置和测序化学验证以及将数据实时、自动传输至 BaseSpace 计算环境。

然后采用 BaseSpace App，通过比对和后续的数据分析步骤流畅地开展新一代测序工作流程。BaseSpace App 可提供各类不同的分析途径，包括从头组装、单核苷酸多态性和插入缺失分析、RNA 表达谱、16S 宏基因组学、肿瘤-正常组织比较、表观遗传学 / 基因调控分析以及更多。Illumina 与商业和学术软件开发者密切合作，创建内容丰富的数据分析工具系统环境，以满足各类研究目标的需求。在新一代测序工作流程的最后阶段，数据可与合作者共享或立即交付至世界各地的客户。[§]



如需了解更多关于 BaseSpace Sequence Hub 的信息，请访问：
www.illumina.com/basespace。

[‡] 目前仅限与 MiniSeq 系统和 NextSeq 500/550 系统配合使用。HiSeq 系统和 MiSeq 系统可使用 Illumina Experiment Manager (IEM) 以实现相同的计划和验证功能。

[§] 仅限基于云计算的环境。BaseSpace Onsite Sequence Hub 仅限在本地用户间进行数据共享。

IV. 术语表

接头：与测序文库中每个 DNA 片段的 5' 和 3' 端相连的寡核苷酸。这些接头与 Illumina 测序流动槽表面的寡核苷酸互补。

桥式扩增：在 Illumina 的流动槽表面发生的扩增反应。在流动槽生产过程中，它的表面包被了两种不同的寡核苷酸，通常被称为“P5”和“P7”。在桥式扩增的第一步，将单链测序文库（带有互补的接头末端）上样到流动槽中。当文库沿着表面“流动”时，文库中的单个分子与互补的寡核苷酸结合。当连接片段的另一端弯曲，与表面的另一条互补寡核苷酸“形成桥”时，完成启动。反复的变性和延伸循环（与 PCR 类似）使得单个分子在局部被扩增成数百万个独特的克隆簇。这个又被称为“成簇”的过程发生在自动化的流动槽仪器（称为 cBot 系统）中，或 NGS 仪器内置的簇模块中。

簇：与流动槽表面结合的模板 DNA 的克隆分组。每个簇被单独的 DNA 模板链接种，通过桥式扩增来进行克隆扩增，直到每个簇大约有 1000 个拷贝。流动槽上的每个簇产生单条测序 read。例如，流动槽上 10000 个簇将产生 10000 条单端 read 和 20000 条双端 read。

Contig：一串连续的序列，通过比对重叠的测序 read 以计算机模拟方式生成。

覆盖水平：与参考 DNA 的每个碱基匹配的测序碱基的平均数量。例如，以 30 倍覆盖度测序的全基因组意味着基因组中的每个碱基平均被测序 30 次。

流动槽：视仪器平台而定，包括一个、两个或八个物理隔离道的玻片。每条道都包被有一层结合在表面并与接头互补的寡核苷酸。视应用参数而定，每条道可以测序单独的一个文库或最多 96 个文库的组合。

索引 / 条形码 / 标签：与测序文库中的片段相连接的独特 DNA 序列，便于下游的计算机模拟分类和识别。索引通常是接头或 PCR 引物中的组分，在测序文库的制备阶段与文库片段相连接。Illumina 的索引通常为 8–12 bp。带有独特索引的文库可混合在一起，上样到测序流动槽的一个通道中，并在同一运行中测序。之后通过生物信息学软件识别 read 并分类。此过程被统称为“多重分析”。

插入：在文库制备阶段，样本 DNA 被片段化为特定大小（通常为 200–500 bp，但可能会更大）被连接或“插入”到两个寡核苷酸接头之间。原来的样本 DNA 片段也被称为“插入片段”。

环化配对文库：大小范围为 2–5 kb 的长插入片段、通常作为双端文库运行的测序文库。序列对之间的较长 gap 长度对于创建从头测序、插入缺失识别和其他方法内的 contig 而言较为有用。

多重检测：请参阅“索引 / 条形码 / 标签。”

read：新一代测序使用精密仪器来确定 DNA 或 RNA 样本的核苷酸序列。一般来说，序列“read”指的是与 DNA 或 RNA 样本相对应的 A、T、C 和 G 碱基的数据串。在 Illumina 的技术中，单次测序运行产生数百万条 read。

参考基因组：参考基因组是一种已完全测序并组装好的基因组，可发挥 Scaffold 的作用，以连接和比较新的测序 read。一般而言，作为数据分析的第一步，从测序运行生成的 read 将与参考基因组进行比对。

边合成边测序（SBS）：SBS 技术利用四种荧光标记的核苷酸对流动槽表面的数千万个簇进行平行测序。在每个测序循环中，单个标记的 dNTP 被添加到核苷酸链中。这个核苷酸标记作为聚合的“可逆终止子”：在 dNTP 掺入后，荧光染料可通过激光激发和成像来识别，然后用酶切除，以便下一轮的掺入。在每个循环反应中直接根据信号强度进行碱基检测。

V. 参考文献

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *PNAS*. 1977;74(12):5463–5467.
2. Collins FS, Morgan M, Patrino A. The human genome project: lessons from large-scale biology. *Science*. 2003;300 (5617):286–290.
3. Davies K. 13 years ago, a beer summit in an English pub led to the birth of Solexa. *BioIT World*. September 28, 2010. (www.bio-itworld.com/2010/issues/sept-oct/solexa.html).
4. Illumina. HiSeq X Ten Series of Sequencing Systems. 2014. (www.illumina.com/documents/products/datasheets/datasheet-hiseq-x-ten.pdf).
5. Fallows J. When will genomics cure cancer?. *The Atlantic*. January 2014. (www.theatlantic.com/magazine/archive/2014/01/when-will-genomics-cure-cancer/355739/).
6. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013;14 (5):R51.
7. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nat*. 2008;456(7218):53–59.
8. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One*. 2013;8(10):e77910.
9. Illumina. Nextera DNA Library Preparation Kits data sheet. 2014. (www.illumina.com/documents/products/datasheets/datasheet_nextera_dna_sample_prep.pdf).
10. Illumina. Nextera XT DNA Library Preparation Kit data sheet. 2014. (www.illumina.com/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf).
11. Illumina. TruSeq DNA PCR-Free Library Preparation Kit data sheet. 2013. (www.illumina.com/documents/products/datasheets/datasheet_truseq_dna_pcr_free_sample_prep.pdf).
12. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 2012;2513–2524.
13. Grad YH, Lipsitch M, Feldgarden M, et al. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. *PNAS*. 2012;109(8):3065–3070.
14. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;(9):418–426.
15. McEllistrem MC. Genetic diversity of the pneumococcal capsule: implications for molecular-based serotyping. *Future Microbiol*. 2009;4(7):857–865.
16. Lo YMD, Chiu RWK. Next-generation sequencing of plasma/serum DNA: an emerging research and molecular diagnostic tool. *Clin Chem*. 2009;55:607–608.
17. Ram JL, Karim AS, Sandler ED, Kato I. Strategy for microbiome analysis using 16S rRNA gene sequence analysis on the Illumina sequencing platform. *Syst Biol Reprod Med*. 2011;57(3):117–118.
18. Wang Y, Kim S, Kim IM. Regulation of metastasis by microRNAs in ovarian cancer. *Front Oncol*. 2014;10:143.
19. Dior Up, Kogan L, Chill HH, Eizenberg N, Simon A, Revel A. Emerging roles of microRNA in the embryo-endometrium cross talk. *Semin Reprod Med*. 2014;32(5):402–409.
20. Phillips T. The role of methylation in gene expression. *Nat Education*. 2008;1(1):116.

Illumina 中国

上海办公室 · 电话 (021) 6032-1066 · 传真 (021) 6090-6279
北京办公室 · 电话 (010) 8455-4866 · 传真 (010) 8455-4855
技术支持热线 400-066-5835 · techsupport@illumina.com · www.illumina.com.cn

仅供研究使用。不得用于诊断。

© 2017 Illumina, Inc. 保留所有权利。Illumina, BaseSpace, DesignStudio, HiSeq, HiSeq X, MiniSeq, MiSeq, NextSeq, Nextera, TruSeq 和南瓜橙色是 Illumina, Inc. 的商标或注册商标。本文档包含的所有其他品牌和名称均为其各自所有者的财产。Pub. No. 770-2012-008-B



@illumina_china



@illumina

illumina[®]