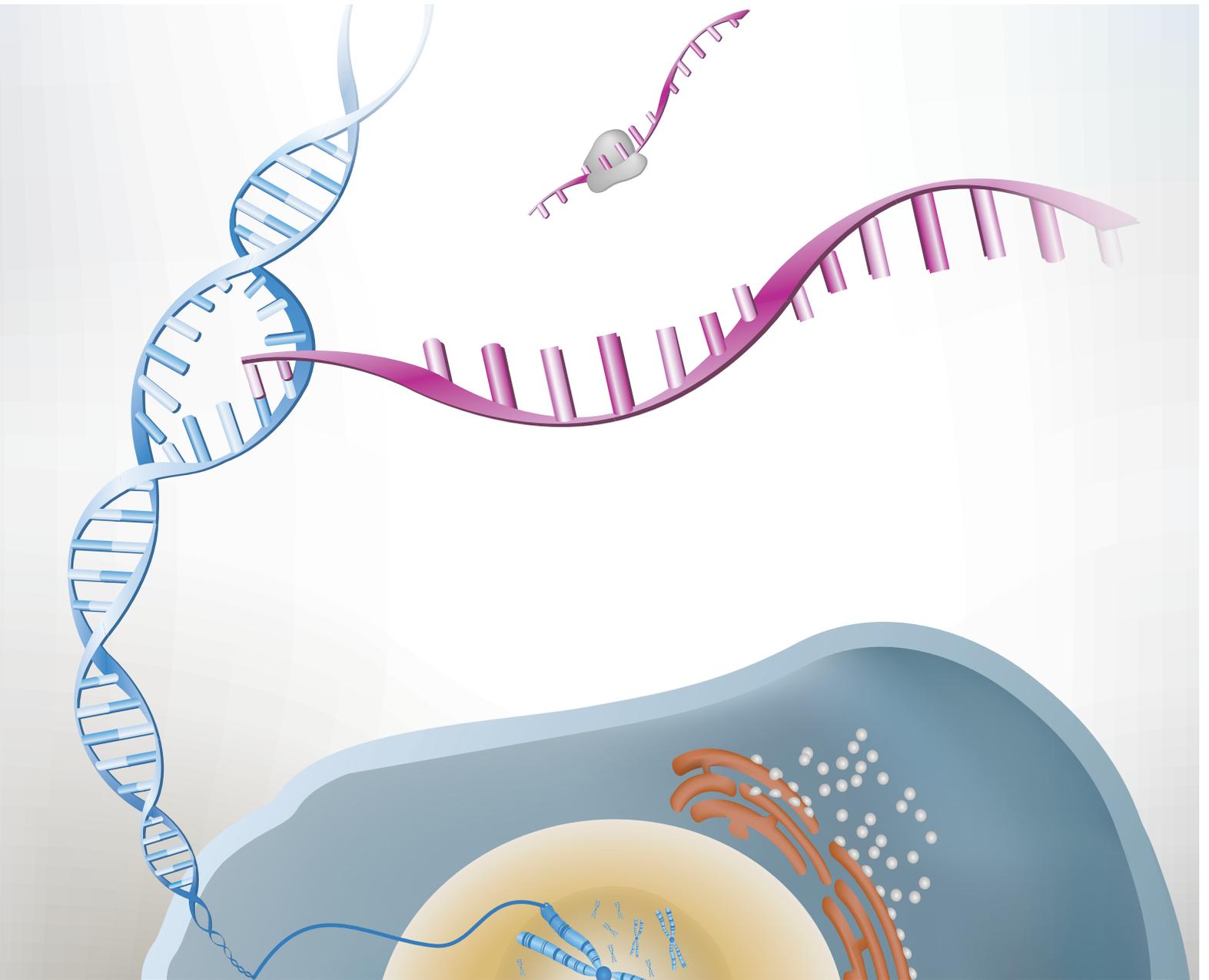


# 细胞生物学家新一代测序使用指南



# 目录

欢迎使用新一代测序	3
为什么选择新一代测序?	3
高通量技术	3
动态范围	4
细胞生物学测序应用	4
基因表达图谱	5
总 RNA 和 mRNA 测序	5
靶向 RNA 测序	5
小 RNA 和非编码 RNA 测序	5
DNA 甲基化分析	6
染色质可及性分析	6
了解 DNA- 蛋白质相互作用	6
什么是新一代测序?	7
新一代测序的基本工作流程	7
多重分析	7
测序服务和系统	8
单细胞研究	8
结语	9
术语表	10
参考文献	11

## 欢迎使用新一代测序

1665年，Robert Hooke 在他的显微镜下放了一片薄薄的软木，观察到它被分成“许多小盒子”，并称之为“细胞”。他的发现和显微镜的发明造就了细胞理论，为生物学奠定了基础。作为生命的基本单位，细胞在包括发育和疾病在内的人类生理的各个方面都发挥着作用。当时，显微镜彻底改变了科学家研究细胞功能的方式，成为了研究细胞外观和行为的重要工具。然而，后来的发现表明细胞的活动不只表现在物理特征。

今天，研究疾病病因的生物学家有许多方法可以用来分析细胞——显微镜检查、流式细胞术和染色等。研究细胞生物学的许多传统方法依赖于抗体染色或其他信号来研究蛋白质的相互作用。单独使用染色法来研究蛋白质功能可能非常困难，因为并非所有蛋白质都有对应的抗体，并且可同时研究的靶点数量有限。这些方法能揭示细胞的物理属性和蛋白质功能，但无法捕捉到生物学的重要部分——决定蛋白质行为的遗传密码。越来越多的科学家发现遗传学与疾病表型之间存在着密切的联系，这也表明想要真正理解蛋白质及其行为，首先要知道其遗传学和调控机制。

了解遗传因子对于生物学的几乎所有分支都是必不可少的，而新一代测序（NGS）技术的引入已经改变了生物学的研究。<sup>1</sup>最近，DNA 元件百科全书（ENCODE）<sup>2</sup>使用 NGS 探索了转录、调控、DNA- 蛋白结合和表观遗传学，在分子水平上提供了有关人类健康的功能信息。NGS 采用序列 read 的数字化计数而不是像许多现有方法一样测量连续信号强度，得到的蛋白质翻译、基因表达和调控的定量信息比传统芯片或抗体方法的分辨率更高。了解这些过程在疾病状态下是如何改变的，可以帮助研究人员预测细胞行为的变化，并设计其他研究来评估特定功能通路中的蛋白质。NGS 为探索和理解疾病的细胞活动开辟了新的途径，科学家们因而得以开发针对导致疾病的生物学通路的靶向医疗方法。

## 为什么选择新一代测序？

作为一种无偏差的技术，NGS 可用于研究疾病生物学的基因表达和蛋白质翻译，并为进一步的功能研究提出假说。它为当前的单分析物实验分析方法提供了高通量替代方案，可在更短的时间内筛查更多的样本。生物学家可以使用 NGS 分析细胞状态和命运、研究信号转导通路、确定致病变异和检查组织特异性基因表达。

## 高通量科学

生物学研究通常会涉及单基因敲除模型，这种方法可以评估某种蛋白质的缺失或修饰如何影响特定的通路。这种方法非常耗时，而且常常无法得出确凿的结论，经常需要设计额外的实验来测试其他的假说。现在，许多科学家会使用其他方法来寻找感兴趣的靶点。NGS 能让研究人员在单一测序运行中检测整个基因组、转录组或表观基因组，并一次观察多个变化。测序结果可以通过鉴定后续研究中感兴趣的蛋白质来为实验设计提供信息，以便研究人员有效地研究目标、节省时间并且更快地发布研究结果。

## 动态范围

动态范围是可检测到的最高和最低信号之间的差值。NGS 的数字化性质支持较大的动态范围，能为基因表达分析等定量应用提供高灵敏度。通过 NGS，研究人员能以比传统芯片法更高的分辨率对 RNA 活性进行定量，这对于捕获与生物过程相关的细微基因表达变化非常重要。<sup>3</sup> 芯片测量连续的信号强度，检测范围受限于低端的噪音和高端的信号饱和，而 NGS 可以对离散的数字化测序 read 数量进行定量分析。通过增加或减少测序 read 的数量，研究人员可调节实验的灵敏度，以满足不同研究目标的需求。

## 细胞生物学测序应用

NGS 平台支持多种应用，研究人员可在该平台上研究任何细胞类型或生物。在疾病病例中，有多种因素可以影响表型，包括基因突变、不同时间和不同环境条件下基因表达的变化，以及表观遗传调控的差异。细胞生物学家在寻找导致复杂性状或疾病的分子通路时，必须考虑这些潜在的变化。研究人员可以利用 NGS 的灵活性，使用单个技术来对信号通路中的基因表达进行定量（图 1）。



图 1：NGS 转录组学的完整视图——在过去的 10 年中，采用 NGS 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 RNA-Seq 和靶向 RNA-Seq。

## 基因表达和调控图谱

评估基因表达和调控的传统方法包括芯片和定量聚合酶链式反应 (qPCR)，这些方法通常需要事先了解探针设计。RNA 测序 (RNA-Seq) 是一种更灵敏的方法，可以通过测量离散的数字化测序 read 来进行 RNA 定量。<sup>4</sup> RNA-Seq 的文库制备通常先制备总 RNA 样本，然后将其转化为 cDNA，最后进行标准的 NGS 文库制备。侧重于 mRNA、小 RNA、非编码 RNA 或 microRNA 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 1)。

## 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，大动态范围可实现对常见和罕见转录本的识别和定量分析。其他功能包括在剪接点比对测序 read 以及检测异构体、新型转录和基因融合。<sup>5</sup> 能够精确检测链特异的文库制备试剂盒可用于总 RNA-Seq 和 mRNA-Seq 方法。

## 靶向 RNA 测序

靶向 RNA 测序是一种测量感兴趣转录的方法，适用于检测差异表达、等位基因表达、基因融合、异构体、cSNP 和剪接点。

## 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沉默因子在基因表达调控中发挥作用。由于 microRNA 在转录和翻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。<sup>6,7</sup>

了解有关小 RNA (非编码 RNA)、靶向 RNA、总 RNA 以及 mRNA 测序 Illumina 解决方案的更多信息，请访问 [www.illumina.com/applications/sequencing/rna.html](http://www.illumina.com/applications/sequencing/rna.html)

表观遗传学是研究由 DNA 序列变化以外的其他机制引起的基因活动遗传性变化。最近的研究表明，生活方式和环境因素可能会造成 DNA 表观遗传的变化，从而导致或加剧疾病。表观遗传活动的机制包括 DNA 甲基化、DNA-蛋白质相互作用、染色质可及性、组蛋白修饰等。

## DNA 甲基化分析

异常的 DNA 甲基化及其对基因表达的影响与许多疾病相关，包括阿茨海默病。<sup>8</sup> 胞嘧啶核苷酸的甲基化影响着包括基因表达、RNA 加工和蛋白质功能在内的各种细胞活动。这些过程可以通过在特定时间激活或抑制某些基因来影响细胞生物学，从而影响发育、疾病、衰老和免疫防御。现有两种广泛应用的甲基化测序方法：全基因组亚硫酸氢盐测序 (WGBS) 和简化代表性亚硫酸氢盐测序 (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 CpG 和启动子的 DNA 区域。此外，靶向 Methyl-Seq 在全基因组亚硫酸氢盐测序和甲基化芯片之间提供了一种经济高效的选择。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](http://www.illumina.com/techniques/sequencing/methylation-sequencing.html)

## 染色质开放性分析

真核生物基因组被包装成染色质，染色质的包装方式在基因调控和最终的细胞表型中发挥着关键的作用。<sup>9</sup> 转录因子与 DNA 上的调控序列的结合需要染色质处于开放构象。传统评估开放染色质状态的方法，例如 DNAase I 超敏分析，非常耗时并且需要大量的细胞。转座酶开放性染色质测序 (ATAC-Seq) 是一种快速可靠的方法，可对单个细胞或大量样本进行分析。<sup>10,11</sup> 在这种分析方法中，基因组 DNA 暴露于 Tn5 (一种高活性转座酶)，该转座酶优先插入到开放染色质位点并添加测序引物。测序的 DNA 即代表开放的染色质，我们即可通过数据分析深入了解基因调控。ATAC-Seq 现已用于更好地了解胚胎发育、T 细胞激活和癌症中的基因调控。<sup>12,13</sup>

## 了解 DNA-蛋白质相互作用

染色质免疫沉淀测序或 ChIP-Seq 可用于检测蛋白质、DNA 和 RNA 之间的相互作用。使用 NGS 的 ChIP-Seq 能让研究人员识别整个基因组中多个蛋白质靶点的结合位点，包括转录因子和组蛋白。DNA-蛋白质相互作用分析可供深入了解许多生物过程和疾病状态中重大事件的调控。目前用于转录因子分析的方法，例如芯片和 qPCR，范围有限，而且只提供了一小部分基因的信息。NGS 能进行广泛的分析，可以识别和研究许多在疾病状态中可能被转录因子激活的基因。通过 ChIP-Seq，研究人员可以更好地了解染色质修饰和局部结构改变会如何影响细胞和信号通路中转录因子的活动。

如需了解有关 ChIP-Seq 的更多信息，请访问 [www.illumina.com/techniques/sequencing/dna-sequencing/chip-seq.html](http://www.illumina.com/techniques/sequencing/dna-sequencing/chip-seq.html)

## 什么是新一代测序？

NGS 是一种先进的技术，通过荧光技术提供基因组、转录组或表观基因组的碱基视图。Illumina 测序化学过程，边合成边测序（SBS），是世界范围内应用最广泛、成果发表最多的 NGS 技术。<sup>14</sup> SBS 可以在单个碱基掺入延伸的 DNA 链时对其进行检测。

### 新一代测序的基本工作流程

NGS 的原理与桑格（毛细管电泳）测序相似。每个 DNA 或 RNA 片段沿着单链模板重新合成时，即可根据发射出的信号按顺序识别片段上的碱基。NGS 扩展了这一过程；数百万个反应以大规模并行的方式进行，而不是局限于一个或几个片段。这一进展使大量基因组信息的快速测序成为可能。

我们用单个 DNA 或 RNA 样本展示这一过程。首先，核酸会被片段化，形成可高效测序的小片段文库。识别出的新碱基串被称为 read，以已知的参考基因组为支架重新组装 read（重测序）。所有比对 read 即代表了样本的整个序列（图 2）。

如需观看 Illumina 测序的详细动画，请访问 [www.illumina.com/SBSvideo](http://www.illumina.com/SBSvideo)。

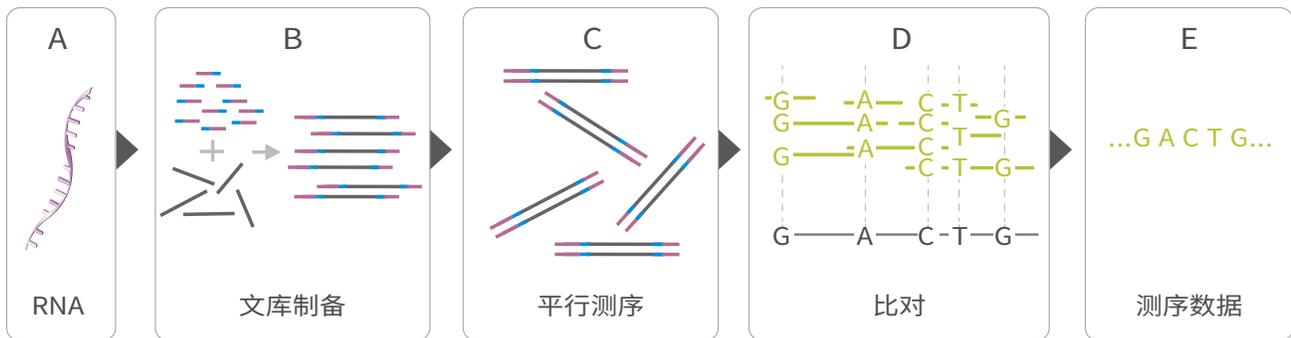


图 2：NGS 的概念。

- 提取核酸（DNA 或 RNA）。如果对 RNA 进行测序，需从 RNA 模板逆转录得到 cDNA。
- 文库制备步骤中，将 DNA 片段化并在末端添加接头。
- 对文库中的每个片段进行平行测序。
- 将单个序列 read 与参考序列（例如参考基因组或已知基因）进行比对。
- 生成一致的比对 read，用于后续的变异检出。

### 多重分析

除了在 NGS 仪器上生成的数据量增加以外，每次运行的样本通量也随着时间的推移而增加。多重分析可以将大量或大批样本合并起来，在单一测序运行中同时测序（图 3）。采用多重样本时，唯一标签序列（或“条形码”）在文库制备过程中被添加到每个 DNA 片段，以便在最终数据分析前识别每一个 read 并对其分类。通过使用 NGS 进行多重分析，研究人员可以快速处理大量样本，并利用其统计学分析能力准确地检测表达谱。多重分析大大缩短了多样本研究获得数据的时间，研究人员能够更快、更轻松地从实验中得到答案。

## 测序服务和系统

服务实验室是研究人员刚开始使用 NGS 时的一种高性价比选项。它遍布全球，能提供多种样本类型的测序服务。供应商可针对各种研究设计提供服务，包括 RNA-Seq、亚硫酸氢盐测序、ChIP-Seq、小 RNA 分析等。一些服务供应商还提供分析后的数据，缩短研究人员分析数据的时间，让其可以专注于下一次发现。

如需查找附近的服务供应商，请联系您当地的客户经理或通过 [customerservice@illumina.com](mailto:customerservice@illumina.com) 联系 Illumina 客户服务部门。

考虑到 NGS 带来的优势，许多研究人员会选择购买 NGS 系统。现有的系统包含可满足不同研究规模的仪器，从可以在单次运行中处理许多个样本的高通量仪器到适合小规模研究的桌面式测序仪。请访问 [www.illumina.com/sequencer](http://www.illumina.com/sequencer)，查找最适合您实验室的测序仪。

如需查看 Illumina 文库制备试剂盒的完整列表，请访问 [www.illumina.com/products/by-type/sequencingkits/library-prep-kits.html](http://www.illumina.com/products/by-type/sequencingkits/library-prep-kits.html)

## 单细胞研究

没有一项技术比流式细胞技术更能拓展免疫学领域。流式细胞术让研究人员能够在单细胞水平上鉴定免疫细胞表型，从而使我们能够了解 T 细胞和 B 细胞亚群以及它们在免疫、耐药性和肿瘤发生中的具体作用。<sup>15</sup> 相比之下，基因组分析技术如 NGS、qPCR 和芯片通常在大量细胞样本上进行。单细胞 NGS 是一种新兴的方法，可检测单个细胞的基因组、转录组或表观基因组，提供组织中细胞间变异的高分辨率视图。单细胞测序可以在周围环境中识别细胞，让研究人员能够单独评估细胞，而不是依赖于整个细胞群的平均信号。目前，在 ILMN 技术的基础上已开发有超过 75 种<sup>16</sup> 单细胞 NGS 方法。这些方法让研究人员能以高通量的方式分析单个细胞的转录组<sup>17</sup>、表观基因组<sup>11</sup> 和基因组<sup>18</sup>。这些研究的数据正在重新定义我们对造血细胞生成<sup>19</sup>、大脑发育<sup>20</sup> 和肿瘤发生<sup>21</sup> 的理解。

## 单细胞研究接上页

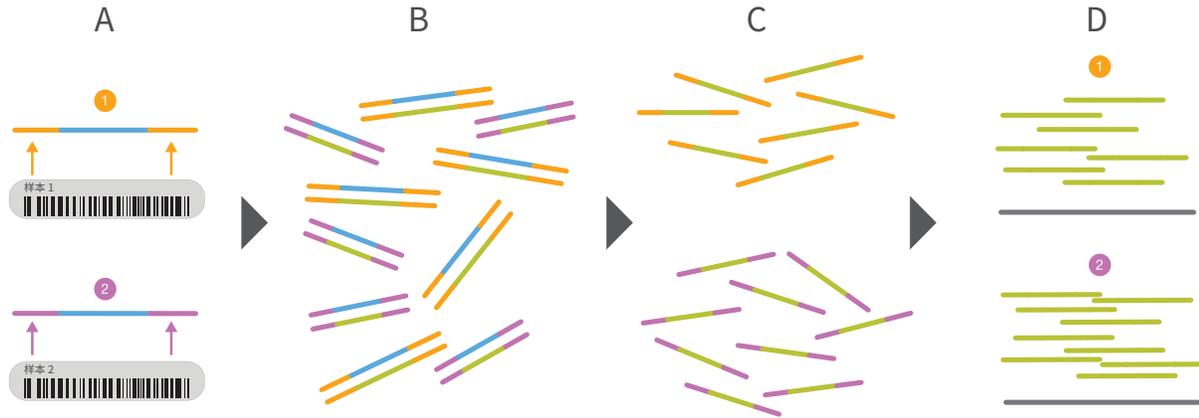


图 3：样本多重分析概述。

- 来自两个独特样本的两个具有代表性的 DNA 片段分别连接到一个特定的条形码序列上，该条形码序列可识别样本来源。
- 混合每个样本的文库并进行平行测序。每个新 read 都包含了片段序列以及能够识别样本的条形码。
- 通过条形码序列将各个样本区分开。
- 每组 read 与参考序列进行比对。

## 结语

过去十年中，NGS 技术的进展提高了人们对基因组学的认知，这反过来又催生出了研究细胞功能和变异的新方法。研究 DNA 和 RNA 序列可深入了解蛋白质功能和调控，这对疾病研究意义重大。

有了 NGS，科学家可以进行多重分子分析，而不是依次分析单个分子，可以为研究建立无偏差的起点，加快研究进度，最终更快地发布研究结果。结果是通过 NGS 的高分辨率和定量分析发现更多蛋白质前体的信号。随着越来越多的实验室采用 NGS，和越来越多的研究不断地持续绘制特定表型的遗传图谱，越来越多的生物学家如今已能够使用强大的遗传工具来指导其实验设计。Illumina 致力于提供业界最高的数据质量，这体现为 Illumina 在所有 NGS 技术公司中拥有最多的仪器安装基数<sup>22</sup> 以及与许多研究领域的领导者保持着良好的关系。总而言之，我们带来了 NGS，让人们对人类生物学和疾病能有更深层次的了解。

如需查看 Illumina 文库制备试剂盒的完整列表，请访问

[www.illumina.com/products/by-type/sequencingkits/library-prep-kits.html](http://www.illumina.com/products/by-type/sequencingkits/library-prep-kits.html)

## 术语表

**接头：**与测序文库中每个 DNA 片段的 5' 和 3' 端相连的特定寡核苷酸。这些接头序列与 Illumina 测序流动槽表面结合的寡核苷酸互补。

**桥式扩增：**在 Illumina 流动槽表面发生的扩增反应——也称为簇生成。流动槽表面被两种不同的寡核苷酸覆盖。重复的变性和延伸循环（类似于 PCR）导致单个片段局部扩增至数千个相同的片段。数百万至数十亿个独特的克隆簇覆盖于流动槽表面。Illumina NGS 的簇生成发生在测序仪器或 cBot（单独的流控技术仪器）中。

**簇：**与流动槽表面结合的模板 DNA 的克隆分组。每个簇都来源于单条 DNA 模板链，通过桥式扩增生成大约 1,000 个拷贝。流动槽上的每个簇都产生单条测序 read。例如，流动槽上 100 万个簇将产生 100 万条 read。

**流动槽：**带有 1-8（取决于仪器平台）条物理隔离道的载玻片。每条道都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，每条道可以运行单个样本或最多 384 个多重样本的混池。

**标签序列：**也称为条形码或标签，这些独特的序列通常长 8-12 个碱基对，会与测序文库中的片段连接，用于后续数据分析步骤中的识别。在文库制备阶段添加标签序列（通常是接头的一部分）。

**多重分析：**带有唯一标签的多个样本可以合并在一起，上样到同一流动槽中，在单一测序运行中同时测序。根据应用类型和所用测序仪器的不同，可以合并 10-384 个样本。

**Read：**流动槽上的单个簇产生的独特序列。序列 read 的长度取决于仪器运行时设置的测序循环数。例如，150 个循环数的测序运行会产生 150 个碱基对的 read，流动槽上的 100 万个簇将产生 100 万个独特 read。测序运行完成后，所有的序列 read 都将导出生成数据文件。

**参考基因组：**已知或先前已测序的基因组。新的序列 read 会比对到充当骨架参考基因组上（重测序）。在没有参考基因组的情况下，必须使用 contig 组装（从头测序）来构建基因组。

**边合成边测序（SBS）：**SBS 技术利用 4 种荧光标记的寡核苷酸来平行测序流动槽表面的数百万至数十亿个簇。在每个测序循环中，单个带有标记的 dNTP 被添加到核酸链中。核苷酸标记作为“可逆终止子”用于聚合。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。通过测定每个循环的信号强度来直接进行碱基检出。

## 参考文献

1. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods*. 2008;5:16-18.
2. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol*. 2011;9:e1001046.
3. Su Z, Labaj PP, Li S, et al. A comprehensive assessment of RNA-Seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotech*. 2014 32:903-914.
4. 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:e77910.
5. Wang Z, Gerstein M, Snyder M. RNA-Seq and revolutionary tool for transcriptomics. *Nature Rev Genet*. 2009; 10:57-63.
6. Wang Y, Kim S, Kim IM. Regulation of metastasis by microRNAs in ovarian cancer. *Front Oncol*. 2014;10:143.
7. 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.
8. Lunnon K, Smith R, Hannon E, et al. Methyloomic profiling implicates cortical deregulation of *ANK1* in Alzheimer's disease. *Nat Neurosci*. 2014;17:1164-1170.
9. Kornberg, RD, Lorch Y. Chromatin structure and transcription. *Annu Rev Cell Biol*. 1992; 8:563-87.
10. Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNDNA-binding proteins and nucleosome position. *Nat Methods*. 10:1213-20.
11. Cusanovich DA, Daza R, Ade A, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*. 2015; 348:910-14.
12. Cusanovich DA, Reddington JP, Garfield DA, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018; 555:538-542.
13. Gate RE, Cheng CS, Aiden AP, et al. Genetic determinants of co-accessible chromatin regions in activated T-cells across humans. *Nat Genet*. 2018; 50:1140-1150.
14. 基于 Illumina 公司 2017 年存档数据计算。
15. Practical Flow Cytometry. Fourth Edition. Shapiro, HM. 2013.
16. Uber Research, Dimensions analysis
17. Macosko EZ, Basu A, Satija R, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161:202-14.
18. Navin NE. The first five years of single-cell cancer genomics and beyond. *Genome Research*. 2018 25:1499-1507
19. Paul F, Ya'ara A, Giladi A, et.al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* 2015 163:1-15
20. Pollen AA, Nowakowski, TJ, Shuga J, et. al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotech*. 2014 32:1053-61
21. Puram SV, Tirosh I, Parikh AS, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*. 2018;172:1-14.
22. 基于 Illumina 公司 2017 年存档数据计算。

## 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。

© 2018 Illumina, Inc. 保留所有权利。所有商标均为 Illumina 公司或其各自所有者的财产。

Pub. No. 1070-2014-006 QB 6452 Current as of 25 September 2018



@illumina\_china

@illumina

illumina®