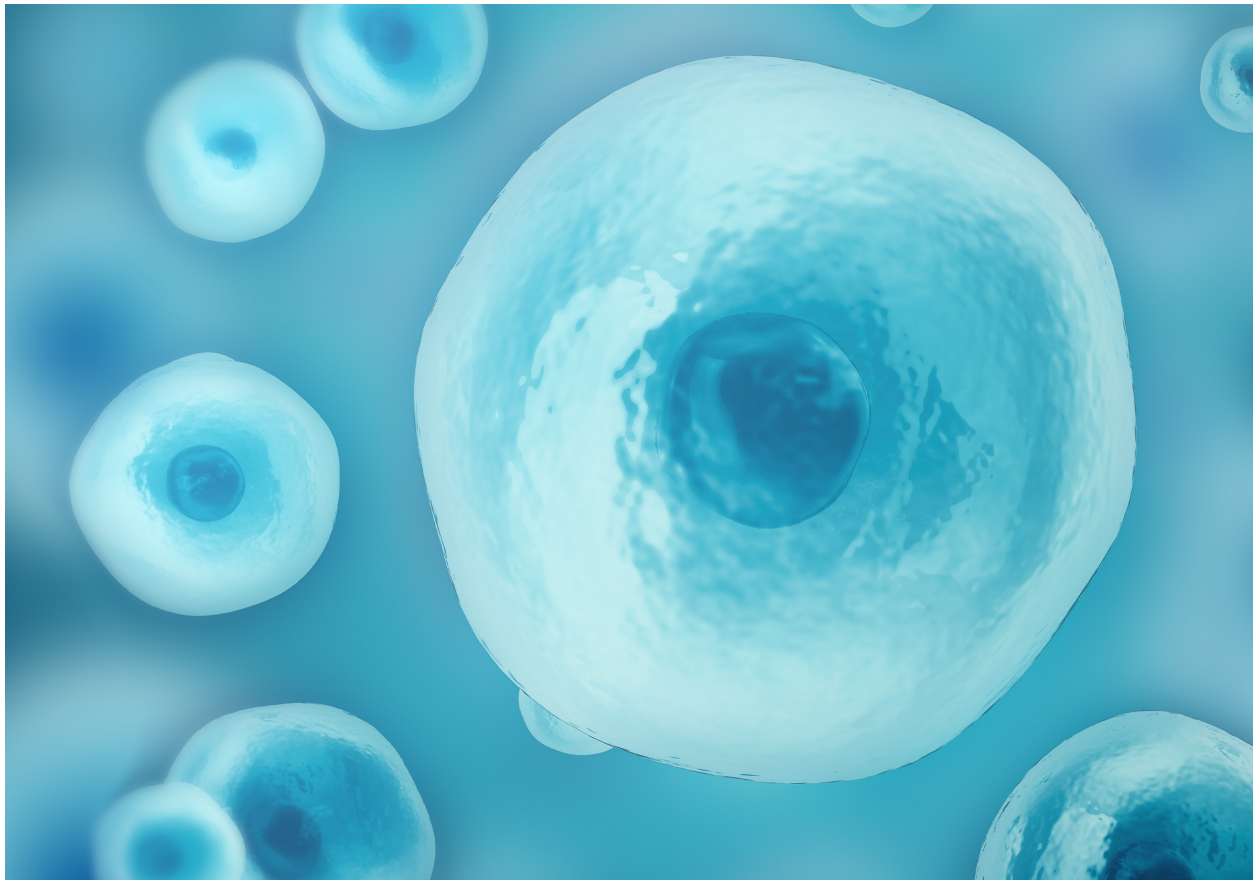


单细胞测序工作流程：关键步骤和注意事项

了解单细胞测序工作流程的每个步骤和有价值的信息，以确保实验成功。



[深入了解单细胞测序](#)

目录

1 引言	4
单细胞测序工作流程	5
2 第 1 步：单细胞制备	6
简介	6
裂解	6
富集	6
质量控制	7
目视检查	7
流式细胞术	8
关键指标	8
总结	8
3 第 2 步：单细胞分离和文库制备	9
简介	9
单细胞分离方法和平台	9
文库制备	11
制备文库的 QC	11
总结	13
4 第 3 步：测序	14
简介	14
兼容的测序系统	14
NextSeq™ 550 系统	14
NovaSeq™ 6000 系统	14
iSeq™ 100 系统	15
测序注意事项	15
实验规划	15
运行 QC	18
仪器控制软件	19
总结	19
5 第 4 步：数据分析、可视化和解读	20
简介	20
初级分析：文件转换	21
*.bcl 文件格式	21
*.fastq 文件格式	21
bcl2fastq 转换软件	21
二级分析：多重分离、比对和 QC	21
数据分析 QC 指标	22
预期文库大小和表达基因的数量	22
比对到线粒体/核糖体的 read 比例	23

Knee 曲线图	24
评估双联体	24
每个细胞的基因数量	24
跨物种分析	25
三级分析：数据可视化和解读	26
Seurat.....	26
使用 SeqGeq 软件进行高级数据可视化.....	26
总结	28
6 总结.....	29
7 了解更多.....	30
8 术语表	32
9 参考文献.....	34

1 引言

17 世纪中叶，安东尼·列文虎克（Anton van Leeuwenhoek）和罗伯特·胡克（Robert Hooke）开展了大量的微观研究，于 1665 年发现了细胞。这项开创性的工作最终促成了细胞生物学学科和细胞学说的建立。这一具有历史意义的科学理论认为，生物体是由一个或多个细胞组成，细胞是构成生物体结构和组织的基本单位，所有的细胞都来源于已存在的细胞。细胞与分子生物学、遗传学和其他领域的重大进展揭示了多细胞生物组成上的高度复杂性，实现了在单细胞水平上进行生物学研究。

活组织由多种类型的细胞组成，每种细胞类型都有不同的谱系和独特的功能，对组织和器官生物学产生影响，并最终定义机体整体的生物学功能。每个细胞的谱系和发育阶段决定细胞如何对其他细胞和微环境产生应答。此外，由于随时间产生的随机变化，相同类型细胞的亚群之间以及它们与其他细胞类型之间通常具有遗传异质性。由于存在这种复杂性，想要通过分析大量组织或细胞来深入了解细胞功能非常困难，这就突显出分离单细胞进行研究的必要性。¹

人们已经开发了多种单细胞分离和分析方法，20 世纪 60 年代末发明的流式细胞术（FACS）是一项重大突破，它对血液学、免疫学、癌症研究等领域产生了重大影响。² 这一技术可对细胞特征进行定性和定量的测量，如细胞大小、内部复杂性、DNA/RNA 含量以及多种跨膜蛋白和胞内蛋白（通过检测自体荧光或荧光素标记的抗体），另外，还可以根据不同的表达模式进行细胞分选。分选到的细胞可以进行下游分析，开展体外培养实验或体内移植研究。

定量 PCR（qPCR）除了在群体细胞水平的分析中得到广泛应用外，由于其动态范围广，工作流程成熟，且不需要专门的仪器设备，已成为单细胞下游分析的首选方法。³ 然而，qPCR 只能对少量已知序列的靶点进行检测，检测大量样本时，其工作流程会比较繁琐。高通量测序（NGS）技术的高度准确性和特异性使其成为单细胞测序的理想工具。NGS 提供了更强的发现能力来检测新的基因，无需事先了解序列信息，并且在定量罕见变异和转录本时具有更高的灵敏度，这使得它比 qPCR 更适合用于单细胞分析，特别是对于高通量研究。

单细胞测序工作流程

单细胞测序的工作流程包括 4 个关键步骤：1) 单细胞制备，2) 单细胞分离和文库制备，3) 测序和初级分析，4) 数据可视化与解读（图 1）。在整个工作流程中，有一些影响结果并决定研究能否成功的实验注意事项和关键步骤。想要获得准确的数据，得出有意义的结论，需要精心设计实验，并按要求开展实验。⁴

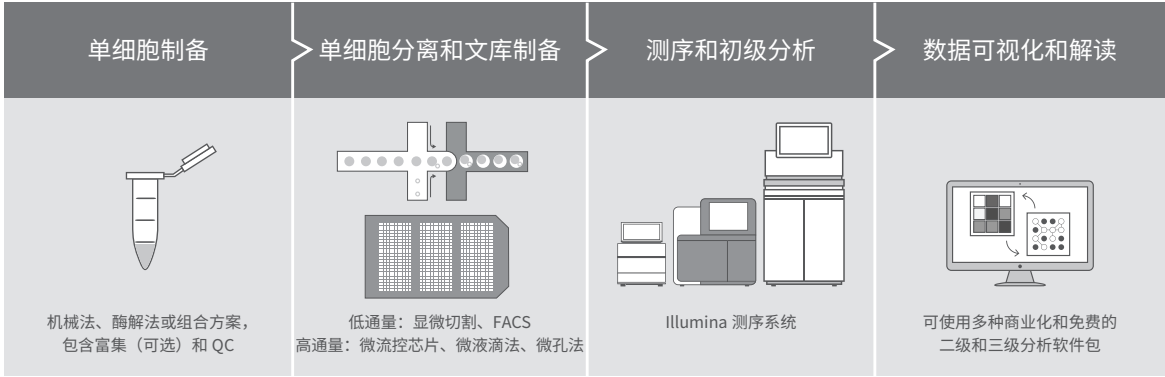


图 1：单细胞测序工作流程 - 单细胞测序工作流程首先是单细胞制备，然后是单细胞分离和文库制备，接下来是测序和初级分析，最后是数据可视化与解读。

2

第 1 步：单细胞制备

简介

大多数单细胞分离平台都需要有活性、单分散的样本，然后才能进行分离或固定。组织类型、物种和动物年龄都可能影响从组织中分离活的单细胞。本章节介绍了制备有活性单细胞悬液的一些主要注意事项。

裂解

在任何单细胞研究中，单细胞悬液的制备过程都是造成差异的重要因素。¹ 细胞聚集成团或细胞死亡率高的样本会影响数据质量，导致错误解读。非贴壁细胞（例如外周血单核细胞）通常比贴壁细胞或从组织中分离的细胞更容易进行单细胞处理。不同组织胞外基质（ECM）的组成和细胞结构可能存在显著的差异，应针对特定的目标组织优化实验方案。² 传统的组织裂解方案有机械分离法、酶解 ECM 法和组合方案（表 1）。

表 1：组织裂解方案

方法/方案	描述	参考方案/供应商
机械法	通过切割、切块、移液器吹打等方法机械剪切和破坏组织	从骨髓、脾脏或淋巴结中分离各种造血谱系的相关细胞
酶解法	使用胶原酶、胰蛋白酶、中性蛋白酶、弹性蛋白酶等多种酶来消化组织，裂解蛋白键	Worthington Biochemical Corporation
组合方案	在自动化系统的帮助下，机械法和酶解法可以依次或同时进行，实现更广泛的裂解	Miltenyi gentleMACS

富集

在单细胞制备中，富集特定的细胞群，或去除不需要的细胞群（包括死细胞）是一个可选的步骤，但通常也是一个关键步骤，尤其是对稀有细胞或珍贵样本。应针对每种特定的组织类型优化不同的方法（表 2）。可通过差速离心或密度梯度离心和过滤的方法，根据大小、形状和密度对细胞进行人工分离。例如，通过各种密度梯度离心，可以从外周血或骨髓中分离出单核细胞。⁵ 使用 FACS 和荧光染料可以从死细胞或凋亡细胞中标记和分离活细胞（表 3）。富集细胞亚群或稀有细胞类型时，可将标记抗体与 FACS 或基于磁珠的分离方法相结合进行阳性/阴性筛选。最终，多种因素将影响方法的选择，包括样本类型、可用抗体种类和实验设计。

表 2：富集方法

方法	描述	可用方案/供应商
离心	通过密度梯度离心，根据细胞大小、形状或密度富集细胞	Sigma-Aldrich
基于磁珠的富集	通过磁珠结合抗体的阳性/阴性筛选来富集感兴趣的细胞群（包括活细胞）	Miltenyi Biotec
FACS	通过荧光基团/荧光素结合抗体的阳性/阴性筛选来富集感兴趣的细胞群（包括活细胞）	Beckman Coulter Becton Dickinson BioLegend Bio-Rad
微流体细胞分选	利用基于荧光基团/荧光素结合抗体的阳性/阴性筛选的低压微流体，来富集感兴趣的细胞群	NanoCelect

表 3：活 / 死细胞染色试剂

试剂	工作原理	优点	缺点
经典的 DNA 染料	与 DNA 结合的染料（例如 PI、7-AAD），无法穿透完整的细胞膜，不会进入活细胞	价格便宜，使用方便	与细胞内染色不兼容
胺染料	与蛋白质氨基结合的染料，无法穿透完整的细胞膜，不会进入活细胞	与细胞内染色兼容，有多种染料可供选择	比其他染料更昂贵，必须在无游离蛋白的情况下进行标记
活体染料	可穿透完整的细胞膜，只有在有代谢活性（活）的细胞中分裂后才会发出荧光	价格便宜，使用方便	用于细胞内染色时具有一定的困难和限制

缩写：PI，碘化丙啶；7-AAD，7-氨基放线菌素 D
链接：expertcytometry.com/3-reagents-for-identifying-live-dead-and-apoptotic-cells-by-flow-cytometry/

质量控制

单细胞测序实验需要投入大量的时间、金钱、样本材料和资源。在进行单细胞分离、文库制备和测序之前，几个简单的质量控制（QC）措施可以确保实验的高质量。

目视检查

在显微镜下对细胞悬液进行目视检查是很有必要的，它可以快速识别可能会使下游步骤变得复杂的碎片、细胞双联体和细胞聚集物（图 2）。更重要的是，准确的细胞计数对于实现后续单细胞分离过程的目标细胞通量至关重要。可以使用显微镜和血细胞计数器进行手动细胞计数。也可以使用自动化系统进行准确的细胞计数，采集细胞悬液的明场图像，根据细胞特征（例如大小、亮度和圆度）生成直方图进行更详细的检查。商业化的自动化细胞计数器有 Countess II Automated Cell Counter（Thermo-Fisher）、TC20 自动细胞计数器（Bio-Rad）和 Auto 1000 Bright Field Cell Counter（Nexcelom Bioscience）等。

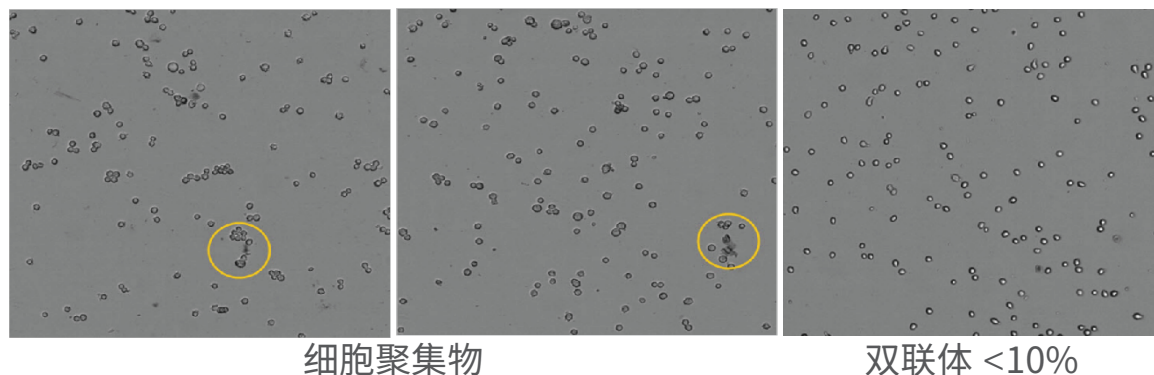


图 2：目视检查细胞悬液 - 用明场显微镜目视检查组织裂解后制备的细胞悬液，可以发现样本中存在碎片、细胞双联体和较大的细胞聚集物（黄色圆圈）。图中展示了一个细胞双联体少于 10% 的样本（右）。

流式细胞术

流式细胞术是一种非常有价值的质量控制工具，可以同时评估多种指标，包括细胞大小、活率和双联体或聚集物。此外还可以将抗体标记作为分析的一部分，评估是否存在感兴趣的细胞群以及其比例是否适当。

关键指标

可以检测几个关键的 QC 指标，来确定单分散细胞悬液的制备是否成功，这些指标包括：

- **细胞活率**：死亡或受损的细胞会将核酸释放到细胞悬液中，这些核酸会在后续步骤中保留下来，可能会影响结果。建议细胞活率大于 85%。
- **细胞大小分布**：直方图可以检查多种峰的存在，这些峰能指示细胞碎片（较小的峰）、双联体/聚集物（两倍标准细胞大小的峰）或大碎片（较大的峰）。
- **细胞浓度**：理想的细胞上样浓度取决于分离方法。合适的浓度至关重要，上样量不足或过多都会影响单细胞分离或数据的质量。

总结

利用单细胞测序来进行单细胞水平的复杂生物系统研究，需要将组织恰当地裂解成单分散活细胞悬液。可供选择的方法很多，应根据目标组织选择具体的方案并进行优化。应考虑纳入富集步骤和关键的 QC 指标，以确保在保持活率的同时提高单细胞的产量。优化单细胞制备方案后，研究人员可以更加放心地进行单细胞分离和文库制备。

3

第 2 步：单细胞分离和文库制备

简介

有多种方法可用于单细胞的捕获和分离，应主要根据研究问题和样本类型来选择最佳方法。同样，有多种技术可用于分析分离细胞的基因组、转录组、表观基因组和蛋白组，选择的方法将决定文库制备、测序和下游分析的方法。本章节讨论了可用的单细胞分离方法，重点介绍了对分离细胞进行全局表征的技术。

单细胞分离方法和平台

单细胞分离方法可以根据通量来区分。低通量方法包括手工操作或细胞分选/分离技术（例如 FACS），每次实验能处理几十个，几百个甚至几千个细胞（表 4）。微流体技术的进展使高通量单细胞分析成为可能，利用该技术，研究人员可以在一次实验中经济有效地检测成千上万个细胞（表 5）。⁶

表 4：低通量单细胞分离方法

方法/平台	描述	优点	缺点	商业化产品/方法示例
梯度稀释	梯度稀释细胞悬液至每孔一个细胞	方法简单；不需要专业设备	耗时；分离得到的细胞可能为多个细胞	Corning 梯度稀释方案
口吸移液	使用玻璃毛细管分离单细胞	方法简单	操作困难，具有随机性	N/A
自动显微操作	用自动微量移液管分离单细胞	可在指定位置放置细胞	需要专业设备	一种用于高通量单细胞培养的自动化系统。使用自动化系统进行单细胞沉积和图案化。
激光捕获显微切割	用激光从组织切片中分离单细胞	保留了空间关系	技术上有难度；对 DNA/RNA 有潜在 UV 损伤	激光捕获显微切割从复杂组织分离单细胞。
FACS	使用电荷分离含有单细胞的微滴	根据细胞大小、形态、内部复杂性和蛋白表达（通过抗体标记）准确筛选细胞类型	需要昂贵的专业设备；细胞处于高压条件下	Beckman Coulter Becton Dickinson Bio-Rad

表 5：高通量单细胞分离方法

方法/平台	描述	优点	缺点	商业化产品/方法示例
微流控系统	微流控芯片分离流动槽中的细胞	高度灵敏的技术，兼容小体积细胞，灵活的定制试剂	要求细胞大小一致，耗材昂贵	Fluidigm C1 System Fluidigm Polaris System
微液滴平台 ⁷⁻¹⁰	利用微液滴制备装置将单细胞分离到微液滴中，然后裂解并捕获目标 DNA/RNA	独特分子标签（UMI）和细胞条形码使细胞和基因特异性识别成为可能，分离每个细胞的成本较低，供应商提供广泛的支持	需要专业设备，技术上有难度	1CellBio inDrop System 10X Genomics Chromium Controller Bio-Rad ddSEQ Single-Cell Isolator Instrument Dolomite Bio Nadia Instrument Mission Bio Tapestry Platform
微孔 ^{11,12}	通过芯片中的微孔捕获单细胞	支持细胞成像和短期培养，非常适合贴壁细胞	商业化解决方案较少	BD Rhapsody Single-Cell Analysis System CellMicrosystems CellRaft AIR System Celsee Genesis System
组合标签	经过两轮随机分配至微孔板，结合转座酶进行标记，可以使完整细胞核标记上独特的条形码	方法成本低，可分析大量细胞，兼容多种方法	商业化解决方案较少	使用组合细胞标签法进行单细胞染色质可及性分析。

文库制备

单细胞测序工作流程的下一个关键步骤是文库制备。有多种可用的方法，但选择的细胞分析方法和特定的测序方法是重要的考虑因素（表 6）。具体方法的选择在很大程度上将由研究的问题决定。

制备文库的 QC

准确地评估文库质量和浓度对于最大限度地提高测序数据的质量和产出非常重要。Fragment Analyzer (Advanced Analytical) 是一个成熟的解决方案，可以在 Illumina 测序工作流程的文库制备过程中同时对 DNA 和 RNA 进行定性和定量（图 3）。¹³ Bioanalyzer (Agilent Technologies) 是另一个文库 QC 选项。与 qPCR 等传统方法相比，这两种方法的优势包括：精确而灵敏的 DNA/RNA 定量、片段大小测定和潜在污染物的检测。¹⁴

无论选择何种方法对文库进行定量和定性评估，只有高质量的文库才能进行测序，才能确保产生可靠、高质量的数据。计算基因组质量指数 (GQN) 可以评估基因组 DNA 样本，该数值与用户自定义的、特定应用程序的“高质量 DNA”阈值有关。类似地，RNA 质量指数 (RQN) 或等效的 RNA 完整性指数 (RIN) 是两种广泛使用的 RNA 样本质量评估指标。¹³ 对文库进行定性和定量评估后，可根据测序平台和流动槽，加入适量的文库进行测序。

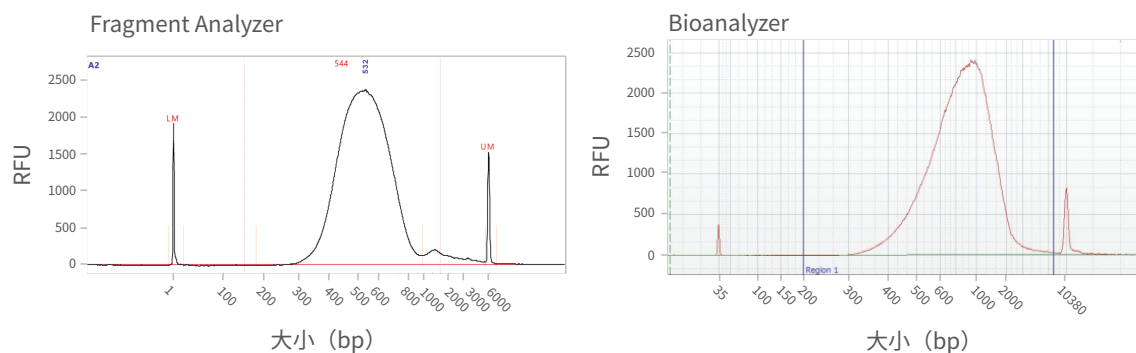


图 3：文库 QC - Fragment Analyzer（左）和 Bioanalyzer（右）的文库 QC 曲线显示了高质量的测序文库。

表 6：单细胞分析的扩增技术

转录组		
方法	描述	商业化产品 / 方法示例
全长 RNA-Seq	RNA 转录本 5' 末端转换机制（SMART）技术可以进行全长 cDNA 的扩增	Takara SMARTer cDNA Synthesis Kit
mRNA 末端标签扩增 (3' WTA 或 5' WTA)	通过 3' 末端多聚腺苷酸（poly(A)）尾链	10X Genomics Chromium 单细胞基因表达解决方案（3' WTA） 10X Genomics Chromium 单细胞免疫分析解决方案（5' WTA） SureCell WTA 3' Library Prep Kit for the ddSEQ System
靶向 panel	各种预先设计的单细胞靶向 RNA 测序 panel，能够对 IR、T 细胞和乳腺癌等进行分析。	BD Rhapsody Single-Cell Analysis
IR-Seq	免疫组库测序（IR-Seq）是一种用于 B 细胞或 T 细胞抗原受体组库组分定量的靶向测序方法。	10X Genomics Chromium 单细胞免疫分析解决方案
基因组		
方法	描述	商业化产品/方法示例
MALBAC	多次退火环状循环扩增技术（MALBAC）可通过单细胞进行整个基因组的准线性扩增。	使用 MALBAC 进行单细胞转录组扩增 Yikon Genomics
DOP-PCR	简并寡核苷酸引物 PCR（DOP-PCR）利用寡核苷酸的部分简并序列进行全基因组扩增。	简并寡核苷酸引物 PCR：使用单一简并引物对目标 DNA 进行普通扩增 使用简并寡核苷酸引物 PCR（DOP-PCR）进行全基因组扩增
靶向 panel	各种预先设计的单细胞靶向 DNA 测序 panel，能够对血液恶性肿瘤、实体瘤、拷贝数变异（CNV）等进行分析。	10X Genomics Chromium 单细胞 CNV 解决方案 MissionBio Tapestri 定制化的单细胞 DNA Panel Mission Bio Tapestri 单细胞 DNA Panel
表观基因组		
方法	描述	商业化产品/方法示例
ATAC-Seq	染色质转座酶可及性测序分析（ATAC-Seq）可通过转座酶将测序接头插入染色质的开放区域中，在全基因组范围内评估染色质可及性。	10X Genomics Chromium 单细胞 ATAC 解决方案 Abcam ATAC-Seq 实验方案 Bio-Rad SureCell 单细胞 ATAC-Seq 建库试剂盒
HiC	HiC 结合了染色质构象捕获（3C）和 NGS，可无偏差地分析整个基因组的染色质相互作用。	Hi-C：捕获基因组构象的综合技术。全面的长片段相互作用图谱揭示了人类基因组的折叠原理。
蛋白质检测		
方法	描述	商业化产品/方法示例
AbSeq	DNA 标记抗体可以通过 NGS 进行蛋白质分析	BD Abseq antibody-oligonucleotide conjugates
CITE-Seq	转录组和表位的细胞标签测序（CITE-Seq）可使用寡核苷酸标记的抗体将蛋白质检测转化为 NGS 定量分析	同时测定单细胞的表位和转录组。 cite-seq.com

总结

分离单细胞是单细胞测序工作流程的一个关键步骤。可供选择的方法很多，应根据研究的问题选择具体的方案并进行优化。应考虑纳入制备文库的 QC 评估。制备好高质量的单细胞文库后，研究人员可以更加放心地进行测序。如果您想与我们讨论各种单细胞测序方法，以及如何将它们整合到您的研究中，请联系当地的 Illumina 代表。

4

第 3 步：测序

简介

分离出活性单细胞，提取感兴趣的遗传物质，制备好文库后，就可以进行关键的测序步骤了。所有 Illumina 测序平台均采用边合成边测序（SBS）技术，全球 90% 以上的测序数据都是基于该技术产生的。¹⁵ Illumina SBS 技术是一种专利方法，能以大规模并行的方式在单个碱基掺入不断延伸的 DNA 链时对其进行检测。Illumina 测序系统单次运行可以输出的数据范围从 30 万碱基到数万亿碱基，取决于仪器类型和配置。本章节介绍了适用于单细胞研究的 Illumina 测序系统，讨论了确保测序成功运行的重要注意事项。

兼容的测序系统

虽然 Illumina 测序系统都能够对单细胞文库进行测序，但是单细胞测序实验选择的测序系统主要取决于研究的问题和规模。推荐使用以下系统进行单细胞测序研究（图 4）。

NextSeq™ 550 系统

NextSeq 550 系统是台式 NGS 系统，可以快速、简单和经济地进行高通量测序。NextSeq 550 系统适用于研究实验室，不需要专门的设备。它支持中通量至高通量的测序应用，是小规模单细胞测序研究的理想选择。



如需了解更多关于 NextSeq 550 系统的信息，请访问

www.illumina.com/systems/sequencing-platforms/nextseq.html

NovaSeq™ 6000 系统

NovaSeq 6000 系统代表了 Illumina 迄今为止最强大、操作简便、可扩展且可靠的高通量测序平台，可提供高品质的数据。它提供了多种流动槽类型和运行配置，从 SP 流动槽的 8 亿条 read 到 S4 流动槽（单端测序模式）的 100 亿条 read。NovaSeq 6000 系统前所未有的数据产出和通量使其成为大规模筛选研究的理想选择，例如药物筛选、细胞图谱研究和其他大规模实验。



如需了解更多关于 NovaSeq 6000 系统的信息，请访问

www.illumina.com/systems/sequencing-platforms/novaseq.html

iSeq™ 100 系统

小型 iSeq 100 系统结合了互补金属氧化物半导体（CMOS）技术与经过验证的 Illumina 边合成边测序（SBS）化学技术，可在实现快速检测的同时提供高度准确的数据，它是 Illumina 产品组合中最小巧、经济的测序系统。iSeq 100 系统非常适合在 NovaSeq 6000 系统上进行全面的测序运行之前进行文库质量控制，这样可以得到更一致的结果，有助于确保实验成功。



如需了解更多关于 iSeq 100 的信息，请访问

www.illumina.com/systems/sequencing-platforms/iseq.html

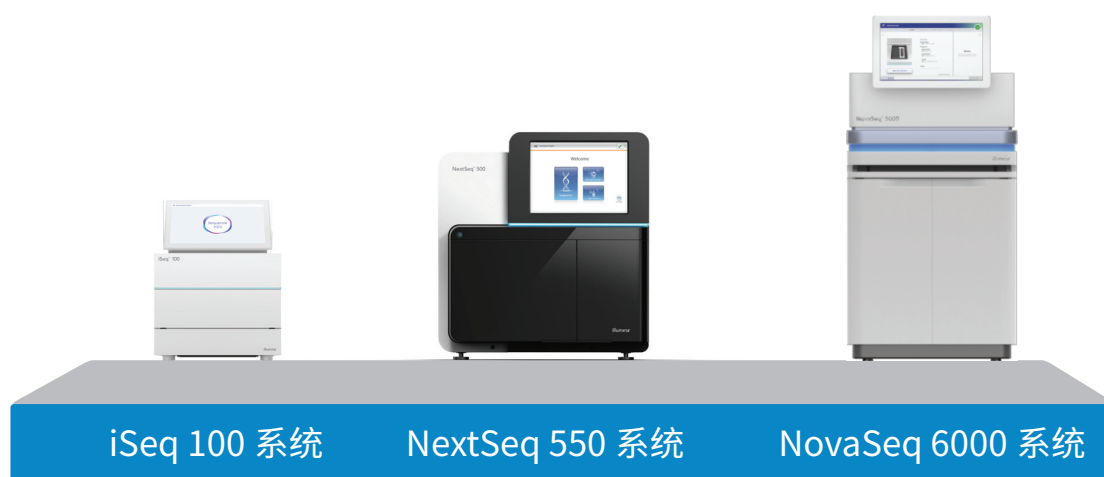


图 4：适用于单细胞测序的 Illumina 测序系统 - Illumina NGS 系统具有灵活的通量和简便、一体化的工作流程，适用于任何规模的单细胞测序实验，可提供高度准确的数据。

测序注意事项

实验规划

read 深度

传统样本或群体细胞样本的测序覆盖度是指比对到或“覆盖”已知参考碱基序列的 read 平均数量。NGS 覆盖度水平常常决定了特定碱基位置的变异发现是否具有某种水平的可信度。对于测序覆盖度的要求因不同的应用而有所不同。在较高的覆盖度下，每个碱基被较多的比对序列所覆盖，或“深度”更深，因此碱基检出也有着较高的可信度。¹⁶

对于各种单细胞测序应用，read 深度不是指每个碱基的 read 数，而是指每个细胞的 read 数。单细胞测序运行所需的测序深度取决于多个因素，包括样本类型、分析的细胞数量、实验目标等。对于单细胞 RNA-Seq，已有报道表明，在混合了不同细胞类型的细胞群中，每个细胞只需 10,000 到 50,000 条 read，就可以实现无偏差的细胞类型分类。¹⁷ 如果实验目的只是鉴定稀有细胞群或发现细胞中混合的细胞群，这种较低的 read 深度比较经济实用。然而在研究更多同质细胞群时，这种 read 深度可能就不够，而且它不能提供给定细胞的基因表达详细信息。在这种情况下，可能需要更深的测序来改进细胞鉴定和低表达基因的检测。已有报道表明，每个细胞测序 50 万条 read 就足以检测到细胞中大多数表达的基因，而每个细胞测序 100 万条 read（接近测序饱和）就可以估计基因表达的平均值和方差。^{18,19} 最终所需的测序深度将在很大程度上取决于样本类型和实验目标，需要针对每个研究进行优化。

双端测序与单端测序

单端测序仅从一端开始 DNA 测序，是最简单的 Illumina 测序方法。单端测序可提供大量高质量的数据，比双端测序更快、更便宜。²⁰ 对于某些方法，例如小 RNA 测序或染色质免疫沉淀测序（ChIP-Seq），单端测序是一个很好的选择。相反，双端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。这能实现更好的 read 比对，特别是在难以测序的、重复的区域上。所有 Illumina NGS 系统都能够进行双端测序（图 5）。

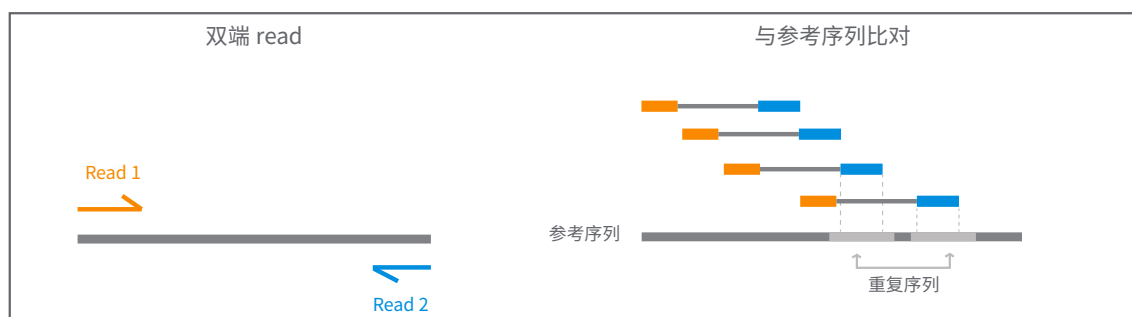


图 5：双端测序和比对 - 双端测序技术能够对 DNA 片段的两端进行测序。因为每对 read 之间的距离是已知的，比对算法可以利用此信息更精确地在重复区域上来标示 read。

在相同的时间内、相同文库制备工作的情况下，双端测序除了可以产生两倍数量的 read 以外，作为 read 对进行比对的序列还可以检测出单端测序数据无法检出的插入缺失（indel）变异。²¹ 此外，双端测序也有助于检测基因组重排，例如插入、删除和倒位。双端 RNA 测序可用于探索应用，如检测融合基因、新的转录本和新的可变剪接体。²²

簇密度

Illumina 的大规模平行测序是通过流动槽表面的簇生成实现的。在以往的簇生成过程中，接头连接的文库元件会与流动槽表面的互补寡核苷酸杂交。每个连接的文库片段作为“种子”，通过桥式扩增过程，扩增生成包含数千个相同片段的克隆簇（图 6）。簇生成完成后，流动槽表面会有数百万到数十亿个簇。

簇扩增

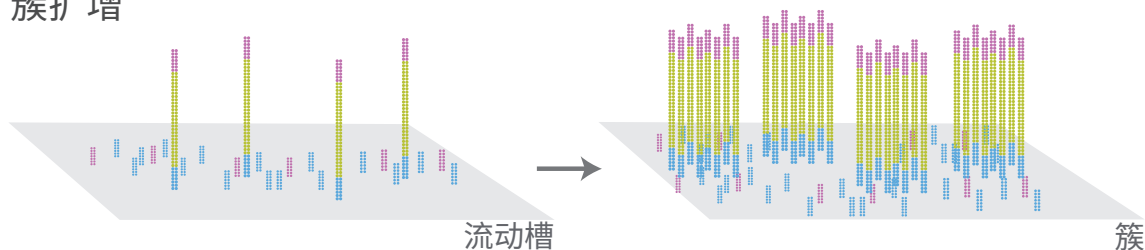


图 6：簇生成 - 文库片段加载到流动槽并杂交到流动槽表面。通过桥式扩增，将每个结合的片段扩增成不同的克隆簇。

理想情况下，各个簇的大小相似，彼此之间间隔良好，能在成像过程中实现准确的分辨率。但实际上，DNA 簇会随机分布在这些“非图案化”流动槽上，许多簇与邻近的簇非常接近，尤其是当样本过饱和时，这样会很难区分各个簇。NextSeq 550 系统使用的是非图案化流动槽。

为了更有效地利用流动槽表面的空间，Illumina 创建了图案化流动槽。图案化流动槽具有蚀刻到表面的图案化纳米孔。每个纳米孔包含 DNA 探针，用于在簇生成过程中采集制备 DNA 链进行扩增。纳米孔之间的区域没有 DNA 探针。这一流程确保 DNA 簇只在纳米孔中形成，相邻簇之间的分布更为均匀，实现了成像时簇的精确分辨率。这种设计能最大限度地利用流动槽表面，使生成簇的总量增多。²³ NovaSeq 6000 系统使用的是图案化流动槽。

流动槽上簇密度会显著影响数据质量和产量，特别是对于非图案化流动槽，它也是测量测序性能的一个关键指标。它会影响运行质量、通过过滤的 read、Q30 分值和总数据产出。想要达到运行的最佳簇密度，需要在成簇不足和过度成簇之间找到平衡。其目标是以足够高的密度进行测序，最大程度地提高总数据产出，同时密度也不能过高，以避免过度成簇。NextSeq 550 系统的推荐簇密度为 170–220 K/mm²。²⁴ 图案化流动槽已经提供了最佳的簇密度，它们不太容易出现成簇不足和过度成簇的情况。但为了实现最佳的性能，仍应该以推荐的浓度上样。大多数商业化的单细胞文库制备试剂盒都针对每个 Illumina 测序系统提供了簇密度建议。

运行 QC

通过过滤的百分比

通过过滤的百分比 (%PF) 是一个重要的测序 QC 指标，它是指通过过滤并且会保留下来进行下游分析的簇数量。使用非图案化流动槽时，Real-Time Analysis 软件会在测序运行早期的模板生成过程中进行图像分析时评估各个簇。去除所有信号较暗或低质量的簇可作为有效的预过滤步骤，以便实现相对较高的 %PF 值。对于图案化流动槽，固定的簇位置不需要在模板生成时进行评估，因此没有针对不理想的簇进行预过滤的步骤。相反，该方法会在后期的纯度过滤中过滤掉不理想的簇。纯度的定义是最亮的碱基信号强度除以最亮的和第二亮的碱基信号强度之和的比值。如果前 25 个测序循环中没有或只有一个碱基检出的纯度值低于 0.6，则簇“通过过滤”。这种过滤过程会从图像分析结果中去除最不可靠的簇。因此，图案化流动槽的 %PF 指标低于非图案化流动槽，但这不会影响性能或数据质量 (图 7)。²⁵

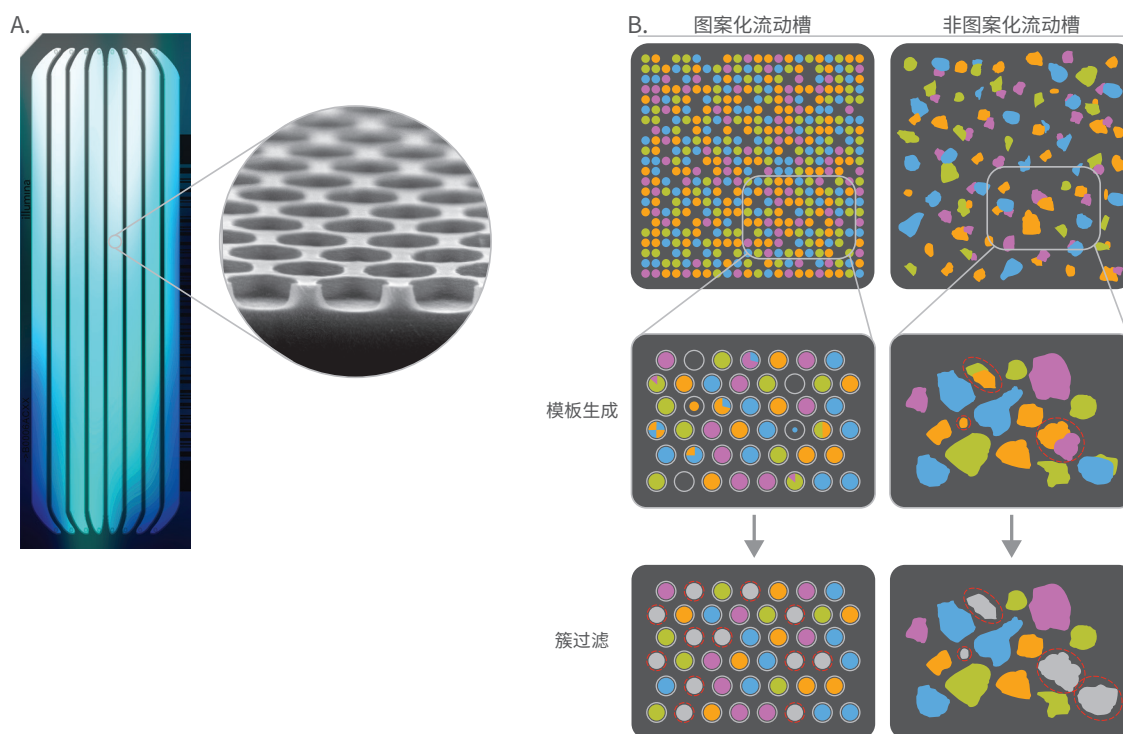


图 7：图案化和非图案化流动槽上通过过滤的簇 - 图案化流动槽表面有蚀刻的纳米孔 (A)。对于非图案化流动槽，模板生成期间会过滤掉质量低或信号较暗的簇 (B)。对于图案化流动槽，会在后期的纯度过滤中过滤掉空孔和不理想的簇，这导致其 %PF 指标较低 (B)。

≥Q30 的碱基百分比

测序质量分值可用于评估碱基的错误检出率。利用 SBS 技术，read 中的每个碱基会得到基于类 phred 算法^{26,27}（与最初为桑格测序实验开发的算法类似）的质量分值。指定碱基的测序质量分值 Q 是按照如下公式定义的：

$$Q = -10\log_{10}(e)$$

其中 e 是预估的碱基错误检出率。Q 值越高，错误率越小（表 6）。Illumina SBS 测序化学过程提供较高的精确度，绝大多数碱基得分在 Q30 及以上（%≥Q30）。然而，Q 值也许并不是评估单细胞文库测序结果最合适的指标，由于文库制备技术、条形码设计和样本制备方面的差异，Q30 分值也会出现明显的差异。大多数商业化的单细胞文库制备试剂盒为评估高质量实验的关键指标提供了指导，包括 Q30 分值、有效的条形码、估算的细胞数量、细胞中 read 片段和检测的总基因数。^{28,29}



如需了解更多关于测序质量分值的信息，请参阅以下技术说明：

- [新一代测序的质量分值](#)
- [了解 Illumina 质量分值](#)

仪器控制软件

所有 Illumina 测序系统都预先安装了仪器控制软件。控制软件会手把手的指导用户装载流动槽和试剂，同时提供质量数据统计概览以便在运行时进行监测。该软件能够自动进行图像分析和碱基检出，生成碱基检出的质量报告。

总结

Illumina 测序系统的数据准确性高，通量灵活，为各种规模的单细胞测序研究提供了成熟的 NGS 解决方案。为了平衡成本和能够达到实验目标的最佳测序参数，应在进行测序之前考虑所需的 read 深度，无论单端测序还是双端测序都需要考虑，这一点非常重要。在测序之前，还应该考虑其他测序指标，例如簇密度、%PF 和 %≥Q30（以及替代指标），测序后也应评估这些指标，以帮助确保获得成功的结果。获得高质量、可靠的测序数据后，研究者即可进行数据分析、可视化和解读。

5

第 4 步：数据分析、可视化和解读

简介

单细胞测序运行完成后即可进行下游分析。一般来说，单细胞测序实验的分析流程包括三个阶段：初级分析（碱基检出）、二级分析（多重分离、比对和遗传鉴定）和三级分析（数据可视化和解读）（图 8）。目前还没有针对单细胞测序实验统一的分析流程。流程中的每个步骤都有许多方法和软件程序可供选择。研究目标、单细胞分离平台和一般的实验室考虑因素将在很大程度上决定使用的特定分析流程。本章节概述了单细胞测序分析的步骤和一些可用的工具。



图 8：单细胞测序分析流程示例 - 包括初始文件转换、初级分析、二级分析和三级分析的单细胞测序分析流程示例。

初级分析：文件转换

*.bcl 文件格式

Illumina 测序系统生成的原始数据文件为二进制碱基检出（BCL）格式。这种测序文件格式包含碱基检出和每个循环中每个簇的碱基检出质量。BCL 文件格式是在测序系统中使用的文件格式，需要转换为 FASTQ 格式，才能用于用户开发的或第三方数据分析工具。

*.fastq 文件格式

FASTQ 是基于文本的测序数据文件格式，可储存原始测序数据和质量分值。FASTQ 文件已成为储存 Illumina 测序系统的 NGS 数据的标准格式，可用作多种二级数据分析解决方案的输入。

bcl2fastq 转换软件

bcl2fastq 软件能将 BCL 文件转换为 FASTQ 文件用于下游分析，第一个 read 完成测序后即可开始这个过程。如果需要对样本进行多重分析，那么生成 FASTQ 文件的第一步就是多重分离。多重测序使多个单独的样本可以在一个流动槽的一个通道中运行，极大地提高了系统的产出。多重分离可以根据簇的标签序列，将簇分配到一个样本。多重分离完成后会根据每个样本组装好的序列生成 FASTQ 文件。如果样本不进行多重分析，则不用进行多重分离步骤，每个流动槽通道中的所有簇都会分配到一个样本。³⁰

二级分析：多重分离、比对和 QC

单细胞测序数据可以在 BaseSpace™ Sequence Hub（Illumina 云端基因组计算环境）中安全地传输、存储和分析。BaseSpace Sequence Hub 提供了大量 BaseSpace App。这些商业化工具和开源工具支持各种常见的测序数据分析需求，如比对、变异检出等。这些 App 具有直观的按键式用户界面，无需生物信息学专业知识。

数据分析的第一步通常是将 read 定位和比对到参考基因组中。有多种软件应用可以使用，包括 BWA Aligner BaseSpace App 使用的 Burrows-Wheeler 比对（BWA）³¹ 算法，以及 RNA-Seq Alignment BaseSpace App 使用的剪接转录本比对至参考基因组（STAR）³² 算法（表 7）。

表 7：BaseSpace 初级分析和二级分析 App

BaseSpace App	描述
	BWA Aligner BWA Aligner App 使用了 BWA-MEM 比对器将样本（包括 FASTQ 文件）与参考基因组（包括通过导入的 FASTA 文件创建的定制参考基因组）进行比对。
	RNA-Seq Alignment RNA-Seq Alignment 的工作流程：使用 STAR 比对器进行 read 定位，使用 salmon 对参考基因和转录本进行定量，使用 Strelka 变异检出软件进行变异检出（SNV 和小插入缺失），使用 Manta 进行融合检出，还使用了来自 Picard 和其他来源的 QC 指标。
	SureCell RNA Single-Cell Single Cell RNA App 适用于分析使用 SureCell Whole Transcriptome Analysis 3' Library Preparation Kit 制备的样本。该应用程序可用于 Illumina Bio-Rad 单细胞测序解决方案的细胞和基因计数、数据过滤和计算，并生成综合指标报告。

有多种 BaseSpace 应用程序和第三方案程序可用于细胞鉴定、计数和绘制感兴趣的遗传构成图谱，如下所示。

- **RNA-Seq**：通过细胞条形码和 UMI（如果有使用）进行多重分离来构建每个细胞的基因表达模型
- **ATAC-Seq**：在每个细胞中识别插入到开放染色质区域的接头序列
- **CITE-Seq**：通过抗体、UMI 和细胞条形码进行多重分离来绘制每个细胞的蛋白表达谱
- **靶向 DNA 测序**：通过细胞条形码进行多重分离来构建每个细胞的基因组变异模型

对于 RNA-Seq，SureCell RNA Single-Cell App 支持 Illumina Bio-Rad 单细胞测序解决方案的数据分析。SureCell RNA Single-Cell App 可进行简化的数据分析，具有测序质量控制指标，可将唯一转录本分配至单细胞，也可以鉴定亚群和差异表达基因。

数据分析 QC 指标

在进行下游分析之前，应评估多个 QC 指标，帮助确定单细胞测序数据集的质量，过滤掉质量较低的数据点/细胞。

预期文库大小和表达基因的数量

每种细胞类型都有一个预期的文库大小，对于 RNA-Seq，也有一个标准的表达基因数量。位于标准的预期范围之外的细胞（过低或过高）可能代表无需进行下游分析的低质量“细胞”，或者也有可能代表值得进一步研究的不寻常的细胞（图 9）。

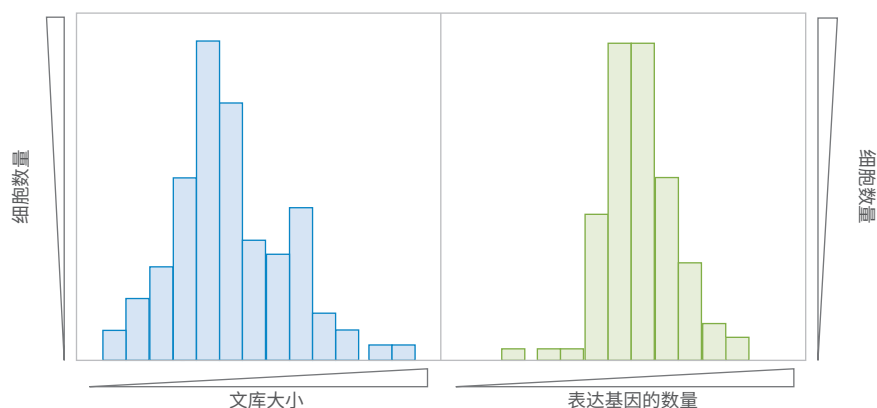


图 9：通过文库大小或表达基因的数量来过滤细胞 - 按文库大小（左图）或表达基因数量（右图）进行分类的细胞分布图。每种细胞类型的每个参数都有一个标准的预期值。超出预期范围的细胞有可能是质量低的细胞、细胞碎片或不寻常的细胞。

比对到线粒体/核糖体的 read 比例

另一个 QC 指标是定位到线粒体基因组基因的 read 比例，或者定位到核糖体 RNA 的 read 比例（图 10）。定位到线粒体和核糖体的比例高表明细胞质量较低，这可能是细胞凋亡增加所致，可以不进行下游分析。

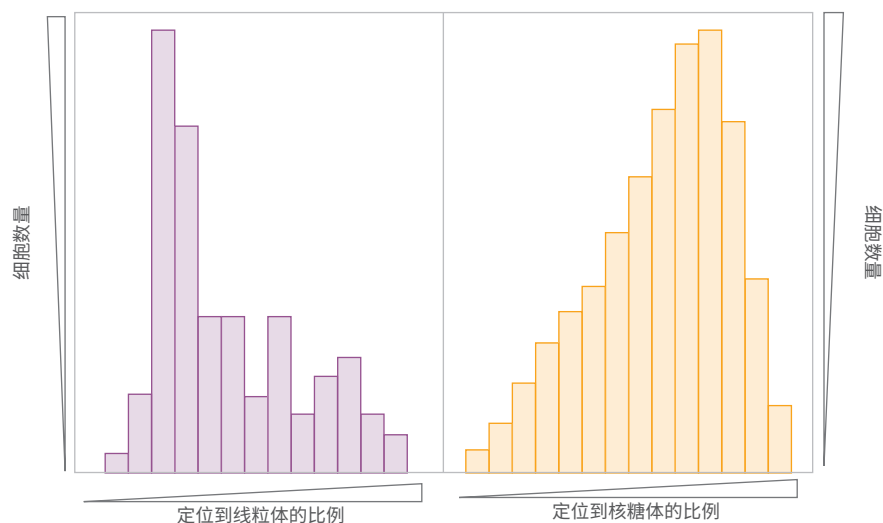


图 10：通过定位到线粒体或核糖体的比例过滤细胞 - 按定位到线粒体基因组（左）或核糖体（右）的 read 比例进行分类的细胞分布图。定位到线粒体或核糖体的比例高的细胞可能质量较低。

Knee 曲线图

绘制基因 UMI 计数与细胞条形码（按基因 UMI 计数降序排列）的曲线图，可以从统计水平上识别“真正”的细胞，排除非细胞条形码（图 11）。阈值以上的细胞条形码（knee 曲线图左侧）的基因 UMI 计数代表了真正的细胞，而阈值以下（knee 曲线图右侧）的细胞基因 UMI 计数低于特定细胞的预期值。

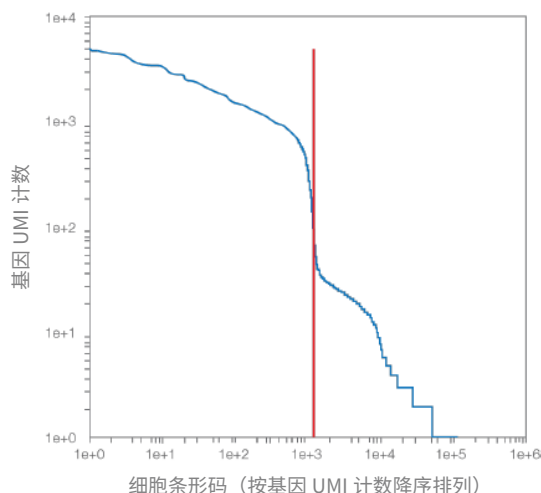


图 11：过滤非细胞条形码 - 阈值（垂直红线）左侧的细胞条形码有数千个基因 UMI 计数，代表真正的细胞。阈值右侧的细胞条形码基因 UMI 计数为 1–100，低于完整的活细胞标准的预期值，代表空的磁珠。

评估双联体

每个细胞的基因数量

任何给定的细胞类型都有一个标准的预期表达基因数量。这个指标已被用于在下游分析中检测和排除双联体。²⁹ 然而，尽管每个细胞使用的基因数可以用于具有同质细胞群的单细胞测序实验，例如培养的细胞系，但它用于复杂的异质组织时可能会有问题。事实上，大多数活的单细胞会落在预期的表达基因数量范围之内 (n)，分布在范围之外的细胞（例如大约是该数量的两倍， $2n$ ）有可能代表感兴趣的细胞，值得进一步研究和鉴定，例如血液样本中的循环肿瘤细胞（图 12）。最后，由于缺乏可靠的双联体检测计算方法，研究人员应该通过实验设计最大限度地降低双联体出现的概率。³³

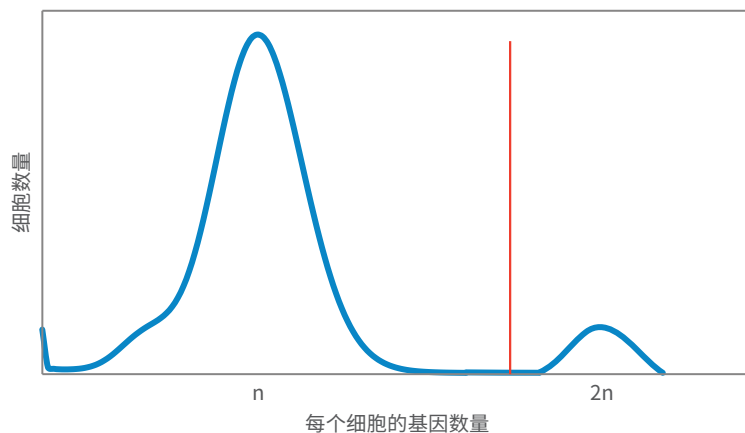


图 12：过基因数量排除双联体 - 按基因数量进行分类的细胞分布图。阈值（垂直红线）右边的每个“细胞”的基因数量是预期的两倍，很可能是双联体

跨物种分析

串扰代表给定实验中液滴或微孔内为双联体细胞的百分比。将两个不同物种的细胞按 1:1 的比例混合在一个样本中是一种检测细胞串扰的有效方法。如果分析的样本中包含了两个不同物种的细胞，检测到一个细胞具有两个物种的 UMI 时，即代表该细胞为双联体（图 13）。

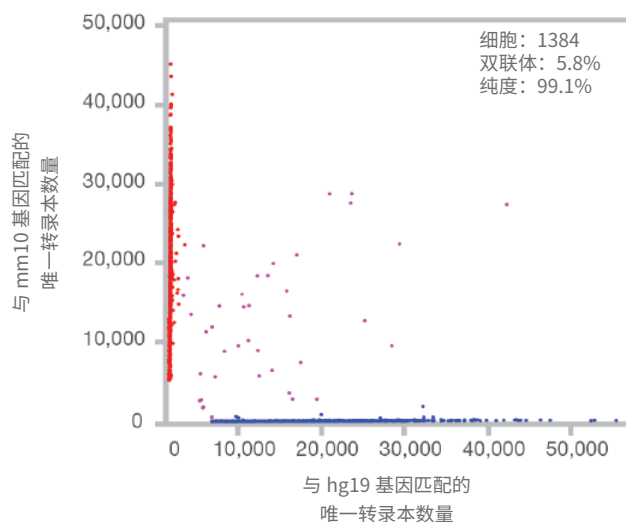


图 13：通过物种特异性的 UMI 排除双联体 - 在用两个不同物种的细胞进行细胞混合实验时，检测到细胞同时定位到两个物种的 UMI（紫色点）即代表其为双联体。

三级分析：数据可视化和解读

read 比对到参考基因组，并且完成了二级分析（包括数据质量控制以去除非细胞条形码和 / 或质量较低的细胞）后，就可以对质量较好数据集进行可视化和深入探索，从而深入了解所研究细胞的生物学信息。有许多免费的和商业化的软件程序可以使用（表 8 和表 9）。

Seurat

Seurat 是一个基于 R 语言的单细胞 RNA-Seq 分析软件，可使用归一化、降维方法、绘图、热图和数据集成等工具来评估细胞异质性。³⁴ Seurat 使用了降维方法将多维数据（例如数千个细胞，而且每个细胞都有数千个表达基因）转换为可以理解的形式，通过数学方法将维数减少到二维或三维来表示。由此得到的细胞聚类对应于具有不同特征的特定细胞状态或类型。（图 14）。

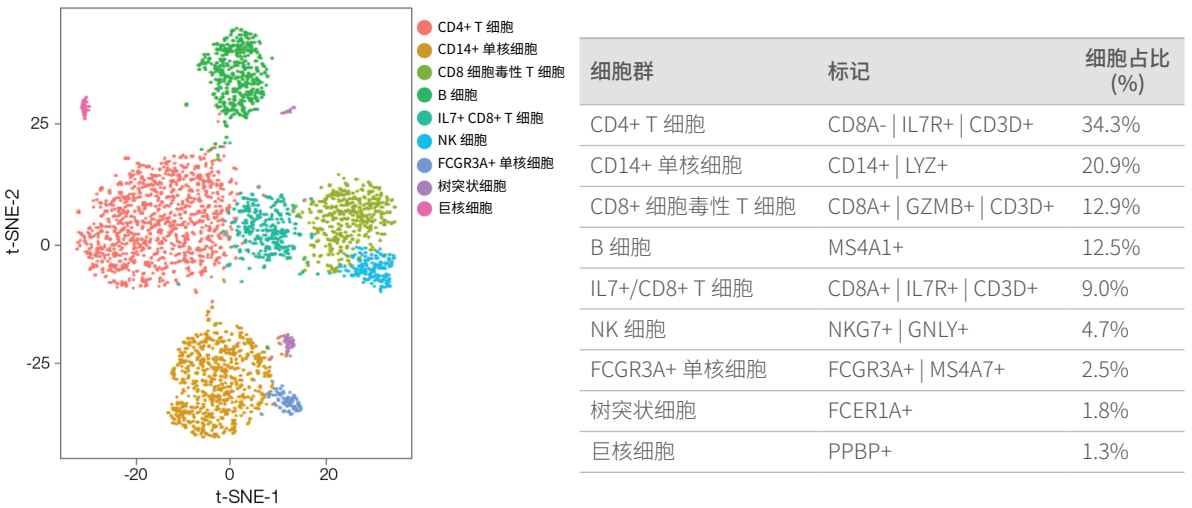


图 14：在 Seurat 中对 PBMC 进行无偏差的聚类分析 - 样本被分为 9 个细胞群，包含 3354 个细胞，降低至约每个细胞 70,000 条 read，分辨率设置为 0.80 和界限值为 100 个基因。表中列出了 Seurat 识别出的细胞。

使用 SeqGeq 软件进行高级数据可视化

SeqGeq 软件是单细胞基因表达数据的高级数据分析、探索和可视化桌面应用程序，由 FlowJo, LLC（现属于 BD Biosciences）开发。SeqGeq 软件具有强大的数据降维和类群识别功能。该软件与 BaseSpace Sequence Hub 直接整合，能生成单细胞统计颜色映射图和汇总热图，具有拖拽报告编辑器，能可视化和分析基因表达数据（图 15）。

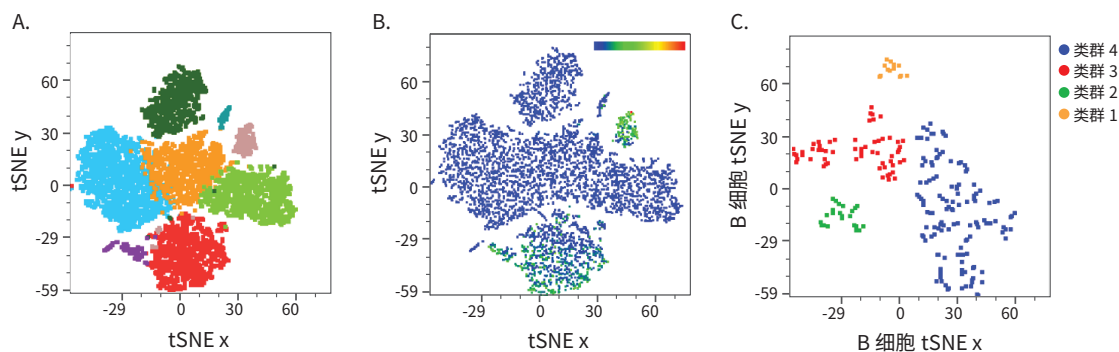


图 15：使用 SeqGeq 简便的聚类分析 - (A) 基于细胞特异性基因/标记的差异表达对 PBMC 进行无偏差的聚类分析。(B) 基于 B 细胞相关基因的表达水平，利用基因集表达丰度鉴定 B 细胞（绿色）的示例（热图中蓝色代表低表达，红色代表高表达）。(C) 在 (B) 中鉴定的 B 细胞内使用 PCA 指导 t-SNE 进行进一步的无偏差聚类分析，鉴定 B 细胞亚群。

表 8：开源的三级分析软件

软件	提供者	描述
Seurat	Satija 实验室 satijalab.org/seurat	Seurat 是一个用于 QC、分析和探索单细胞 RNA-Seq 数据的 R 包。 ³³ Seurat 旨在使用户能够识别和解释单细胞转录组指标异质性的来源，并整合成不同类型的单细胞数据。
t-SNE	Van der Maaten 实验室 lvdmaaten.github.io/tsne	t-Distributed Stochastic Neighborhood Embedding (t-SNE) 是一种通过在二维或三维图中给每个数据点分配一个位置来可视化高维数据的计算技术。 ³⁵ t-SNE 经常用于通过单细胞测序数据来可视化亚群。
UMAP	Git Hub github.com/lmcinnes/umap rdrr.io/cran/Seurat/man/RunUMAP.html	Uniform Manifold Approximation and Projection (UMAP) 是一种分析高维数据的算法，是 t-SNE 的一种替代方法，计算时间更短。 ³⁶
Monocle	Trapnell 实验室—— cole-trapnell-lab.github.io/monocle-release	Monocle 是一个基于 R 语言的单细胞 RNA-Seq 分析软件，专为确定细胞发育轨迹而设计。Monocle 是已知开始和结束细胞状态的实验的理想选择。
人类细胞图谱计划	Broad 研究所—— www.humancellatlas.org	人类细胞图谱计划是一个合作项目，它将构建一个数据协调平台，包含四个关键组成部分：数据提交的接收服务、跨越多个云端的同步数据存储、标准化的二级分析流程和数据访问门户网站、三级分析和可视化。

表 9：商业化三级分析软件

软件	供应商	描述
SeqGeq	BD Biosystems	SeqGeq 软件是单细胞基因表达数据的高级数据分析、探索和可视化桌面应用程序。SeqGeq 提供了强大的数据降维和类群识别功能。
Partek Flow	Partek	Partek Flow 是一款针对 NGS 数据应用的软件分析解决方案。它拥有可靠的统计算法、信息丰富的可视化和前沿的基因组工具，使不同技能水平的研究人员都能够放心地进行数据分析。
CytoBank Platform	CytoBank, Inc.	Cytobank 是一个基于云端的平台，可以同时多个单细胞数据集进行分析和可视化。
Loupe Cell	10X Genomics	Loupe Cell Browser 旨在使用户能够以交互的方式快速地在单个细胞数据中找到重要的基因、细胞类型和亚结构。
Tapestri Insight	MissionBio	Tapestri Insight 是一款用于单细胞 DNA 分析的软件解决方案。它可以进行序列导入、数据分析和可视化。



如需了解更多关于单细胞测序分析流程步骤的信息，请参阅[单细胞 RNA 数据分析工作流程技术说明](#)

如需查看单细胞测序实验的二级和三级分析示例，请参阅[外周血单核细胞的单细胞测序应用说明](#)

总结

单细胞测序实验的分析流程是灵活多变的，可以根据研究目标自行选择。一般来说，该流程包括初级、二级和三级分析阶段，这几个阶段分别会进行序列的比对、遗传构成的鉴定和数据的可视化和探索。如果您想与我们讨论各种单细胞测序分析方法，以及如何将它们整合到您的研究中，请联系当地的 Illumina 代表。

6

总结

在过去的十年中，随着细胞分离新技术和单细胞测序新方法和新应用的发展，单细胞鉴定和研究领域取得了重大进展。这些进展为单细胞测序工作流程的每个步骤（从单细胞制备到数据分析）催生了大量可用的商业化解决方案。随着单细胞分离和检测的选项越来越多，实验方案的多样性显著增加，每种方案都有其固有的优缺点。因此，研究人员面临着诸多决策问题，例如细胞通量、测序深度、所需转录本长度、是否应纳入表观遗传或蛋白水平测量等。

想要充分利用单细胞测序来阐明复杂的生物系统，精心设计实验和优化工作流程的每个步骤至关重要。研究人员必须有明确的生物学目标和合理的实验设计，才能选择出针对其研究问题的最佳方法。在这里，我们概述了单细胞测序工作流程的每个步骤，讨论了每个步骤的重要注意事项和潜在挑战，介绍了商业化产品，并为设计和开展成功的单细胞研究提供了建议。Illumina 致力于利用 NGS 技术进行单细胞测序，加深人们对细胞与分子生物学、复杂疾病和环境对人类健康的影响的了解。

7 了解更多

我如何开始进行单细胞测序？

您可以利用以下资源来规划单细胞测序实验：

- [简单定制的 RNA-Seq 工作流程选择指南](#)
- [新一代测序系统选择指南](#)
- [单细胞 RNA 数据分析工作流程技术说明](#)
- [文库制备和芯片试剂盒选择器](#)

如果我在测序运行或数据分析过程中需要帮助怎么办？

无论您是需要立即处理的基本数据分析问题，还是需要深入咨询的复杂问题，Illumina 都能为您提供帮助。除了即时的电话和电子邮件支持，Illumina 客户服务和支持团队还提供了一整套便捷的方案，从初期培训，到仪器支持、个性化咨询和 NGS 的持续教育。Illumina 客户支持服务包括：

Illumina 技术支持

美洲、欧洲和亚太地区全球每周 5 天每天 24 小时电话和电子邮件支持。

Illumina 技术支持专家可用 GoToAssist 进行桌面共享（GoToAssist 是一个具有桌面共享功能，可通过电话快速识别和诊断问题的强大工具）。为获得更快的案例处理，在主电话菜单输入案例编号，直接引导到处理您案例的技术支持专家。

Illumina 大学培训

- 在您指定的机构开展专家指导的培训
- 在 Illumina 培训中心开展专家指导的培训
- 在线课程
- 网络研讨会

Illumina 咨询服务

- 提供仪器和文库制备检测的概念验证服务
- 提供设计协助和产品优化的 Concierge 定制设计服务
- 提供生物信息学咨询和培训的 Illumina 生物信息学专业服务
- 提供基因组 IT 解决方案的 Illumina 基因组 IT 咨询服务
- 安装验收（IQ）、运行验收（OQ）、性能验收（PQ）

我可以向谁了解更多关于单细胞测序的信息？

如需向 Illumina 代表了解单细胞测序解决方案，请致电 Illumina 客户解决方案中心（北美地区：1.800.809.4566；其他地区：1.858.202.4566），立即开始规划您的单细胞测序实验。

8

术语表

纯度过滤：从图像分析中过滤掉图案化流动槽上不理想的簇的过程。纯度的定义是最亮的碱基信号强度除以最亮的和第二亮的碱基信号强度之和的比值。如果前 25 个测序循环中没有或只有一个碱基检出的纯度值低于 0.6，则簇“通过过滤”。

簇密度：流动槽上簇的数量和分布。簇密度是测序性能的一个重要指标，特别是对于非图案化流动槽，它会显著影响测序运行的数据质量和产量。

簇生成：在此过程中，文库会上样至流动槽中，片段会被一层结合在表面并与接头互补的寡核苷酸所捕获。通过桥式扩增，将每个片段扩增成不同的克隆簇。每个簇包含多达 1000 条样本链，长度通常为 120–170 bp

互补金属氧化物半导体（CMOS）技术：CMOS 技术支持单通道测序，不仅可以获得高度准确的数据，而且能进一步降低小型系统的测序成本。

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。

荧光活化细胞分选（FACS）：该技术可以定性和定量检测多种细胞特征，例如大小、内部复杂性、DNA/RNA 内容以及多种膜结合蛋白和胞内蛋白（通过检测自体荧光或荧光素结合的抗体）。

基因组质量指数（GQN）：由 Advanced Analytical Technologies, Inc. (AATI) 开发的一种计算方法，适用于 Fragment Analyzer，可评估 DNA 样本的质量。

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。

多重分析：一种可以提高测序系统通量的技术，可将大量带有唯一标签的文库混合在一起，上样到测序流动槽的一个通道中，并在同一运行中测序。之后再行多重分离，通过生物信息学软件识别 read 并分类。

新一代测序（NGS）：一种基于非桑格法的高通量 DNA 测序技术。与桑格测序相比，NGS 平台可对数十亿条 DNA 链进行平行测序，大幅度提高通量，并最大限度的降低基因组片段克隆方法（桑格测序常用）的使用。

图案化流动槽：流动槽的固定位置上有数十亿个纳米孔，能使簇间隔均匀且大小均一，实现极高的簇密度。

双端测序：指在同一运行中对 DNA 片段的两端分别进行测序，然后将正向和反向 read 作为 read 对进行比对的过程。

通过过滤的百分比（%PF）：通过过滤的百分比（%PF）是一个重要的测序 QC 指标，它是指通过过滤并且会保留下来进行下游分析的簇数量。

≥Q30 的碱基百分比：Q30 是一个质量分值，对于质量分值为 30（Q30）的碱基检出，预测每 1000 个碱基检出中有一个不正确。≥Q30 的碱基百分比是指质量分值大于等于 Q30 的碱基所占的百分比。

质量分值（Q 值）：是对碱基检出中错误发生几率的预测。

定量聚合酶链式反应（qPCR）：一种能够测量样本核酸量的应用。该方法使用聚合酶来扩增感兴趣的核酸。实时检测 PCR 循环中扩增产物积累的水平。这些数据可用于推断起始核酸量。

read 深度：参见“测序覆盖度”。另外在单细胞测序中，read 深度不是指每个碱基的 read 数，而是指每个细胞的 read 数。

RNA 完整性指数（RIN）：一种衡量 RNA 完整性的算法，这种算法基于 RNA 电泳结果，并结合了各种提供 RNA 完整性信息的特征，得到了一种更通用的衡量标准。

RNA 质量指数（RQN）：AATI 开发的用于 Fragment Analyzer 的专有算法，可评估 RNA 样本的质量，相当于 RIN。

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。

测序覆盖度：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。测序覆盖率也可以称为“read 深度”。

9

参考文献

1. Tung PY, Blischak JD, Hsiao CJ, et al. [Batch effects and the effective design of single-cell gene expression studies](#). *Sci Rep*. 2017;7:39921.
2. Herzenberg LA, Parks D, Sahaf B, et al. [The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford](#). *Clin Chem*. 2002;48(10):1819–1827.
3. Jones GM, Busby E, Garson JA, et al. [Digital PCR dynamic range is approaching that of real-time quantitative PCR](#). *Biomol Detect Quantif*. 2016;10:31–33.
4. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. [Experimental considerations for single-cell RNA sequencing approaches](#). *Front Cell Dev Biol*. 2018;6:108.
5. Sigma-Aldrich. [Centrifugation](#). *Biofiles*. 2011;6(5):4–14.
6. Valihrach L, Androvic P, Kubista M. [Platforms for Single-Cell Collection and Analysis](#). *Int J Mol Sci*. 2018;19(3):DOI 10.3390.
7. Macosko EZ, Basu A, Satija R, et al. [Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets](#). *Cell*. 2015;161(5):1202–1214.
8. Klein AM, Mazutis L, Akartuna I, et al. [Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells](#). *Cell*. 2015;161(5):1187–1201.
9. Zheng GX, Terry JM, Belgrader P, et al. [Massively parallel digital transcriptional profiling of single cells](#). *Nat Commun*. 2017;8:14049.
10. Pellegrino M, Sciambi A, Yates JL, Mast JD, Silver C, Eastburn DJ. [RNA-Seq following PCR-based sorting reveals rare cell transcriptional signatures](#). *BMC Genomics*. 2016;17:361.
11. Retig JR, Folch A. [Large-scale single-cell trapping and imaging using microwell arrays](#). *Anal Chem*. 2005;77(17):5628–5634.
12. Han X, Wang R, Zhou Y, et al. [Mapping the mouse cell atlas by microwell-seq](#). *Cell*. 2018;172(5):1091–1107.
13. Illumina. (2017) [Scalable Nucleic Acid Quality Assessments for Illumina Next-Generation Sequencing Library Prep](#). Accessed May 2019.
14. Agilent Technologies. (2000) Comparing the Agilent 2100 Bioanalyzer performance to traditional DNA analysis techniques.
15. Data calculations on file. Illumina, Inc., 2017.
16. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. [Sequencing depth and coverage: key considerations in genomic analyses](#). *Nat Rev Genet*. 2014;15(2):121–132.
17. Haque A, Engel J, Teichmann SA, Lönnberg T. [A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications](#). *Genome Med*. 2017;9:75.
18. Streets AM, Huang Y. [How deep is enough in single-cell RNA-seq?](#) *Nat Biotechnol*. 2014;32(10):1005–1006.
19. Rizzetto S, Eltahla AA, Lin P, et al. [Impact of sequencing depth and read length on single cell RNA sequencing data of T cells](#). *Sci Rep*. 2017;7(1):12781.
20. National Genomics Infrastructure. [Comparison of PE and SE for RNA Seq](#). *SciLifeLab*. 2016;1–3.
21. 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.
22. Wang Z, Gerstein M, Snyder M. [RNA-Seq: a revolutionary tool for transcriptomics](#). *Nat Rev Genet*. 2009;10:57–63.
23. Illumina. (2015) [Patterned Flow Cell Technology Technical Spotlight](#). Accessed May 2019.

24. Illumina. (2019) [Cluster Optimization Overview Guide](#). Accessed May 2019.
25. Illumina. (2017) [Calculating Percent Passing Filter for Patterned and Nonpatterned Flow Cells Technical Note](#). Accessed May 2019.
26. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998;8(3):175–185
27. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998;8(3):186–194
28. 10X Genomics. (2017) [Chromium Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina Sequencers](#). Accessed May 2019.
29. AlJanahi AA, Danielsen M, Dunbar CE. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Mol Ther Methods Clin Dev.* 2018;10:189–196.
30. Illumina. (2013) [bcl2fastq Conversion User Guide](#). Accessed May 2019.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–1760.
32. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
33. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc.* 2018;13(12):2742–2757.
34. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotech.* 2018;36:411–420.
35. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J of Machine Learning Res.* 2008;9:2579–2605.
36. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotech.* 2019;37:38–44.

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。

© 2019 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。关于具体的商标信息，请访问 www.illumina.com/company/legal.html。Pub. No. 770-2019-007-A QB8344



Illumina Academy



@Illumina

illumina[®]