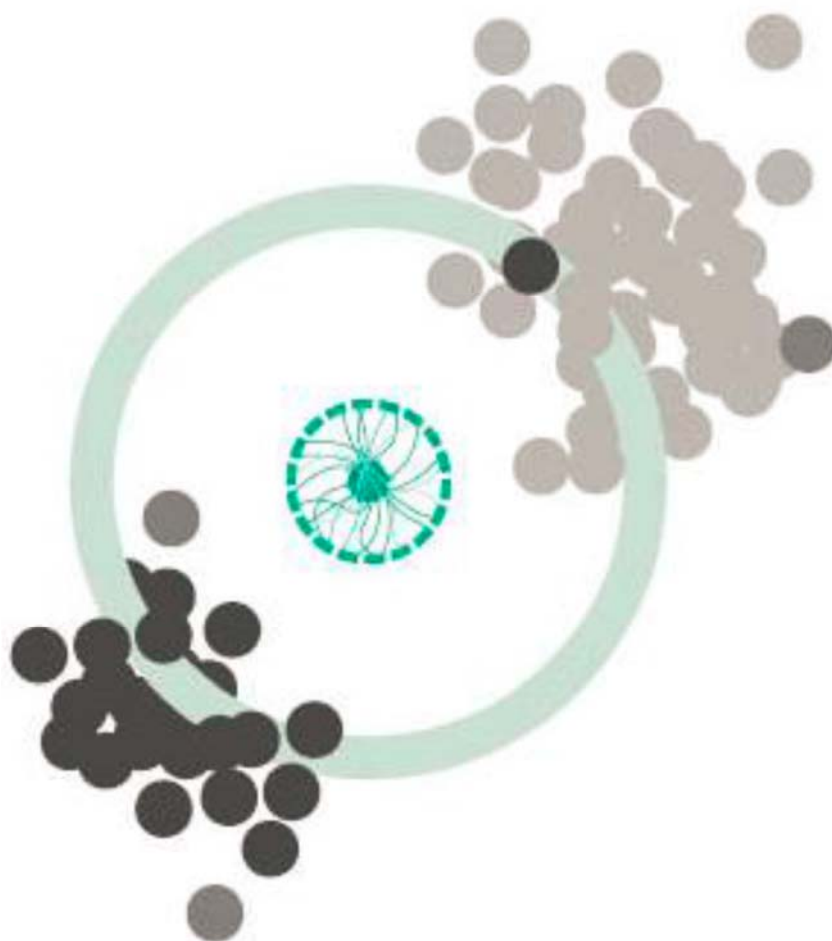


使用 Nadia 平台

对单核转录组进行高通量分析



1.摘要

Dolomite Bio 开发了 Nadia 仪器，这是一种对多个生物样品自动生成液滴，用于高通量单细胞研究的设备。本文描述了使用 Nadia 仪器进行 sNuc-Seq 的方法，就像 Habib 等人在 2017 年初描述的那样，此方法应用于产生高通量单细胞核测序库。

2.前言

Habib 最近发表的 “DroNc-Seq” 方法描述了使用细胞核作为整个细胞的代理，以产生数千个单核转录组。这可以对不易分离或固定的细胞实现高通量的单细胞测序。与 Drop-Seq 方案类似，用单个带有编码的珠子同单个核一同封装在一起（图 1）。珠子的表面分布着含有（dT）30 片段的寡核苷酸，它可以捕获多腺苷酸化的 mRNA，其在乳液破裂后被逆转录成 cDNA。每个珠子除了带有唯一的 barcode 编码,还带有不同的 UMI 位点识别编码，不仅可以识别单个细胞或细胞核，还可以识别来自细胞或细胞核的哪个转录组结合位点。与 Drop-Seq 方案一样需要作出权衡，可以使用较稀浓度的细胞核和较高浓度的捕获珠，降低双细胞核概率，或者使用较高浓度的细胞核和捕获珠，提高捕获效率，但是会增高双细胞核概率。

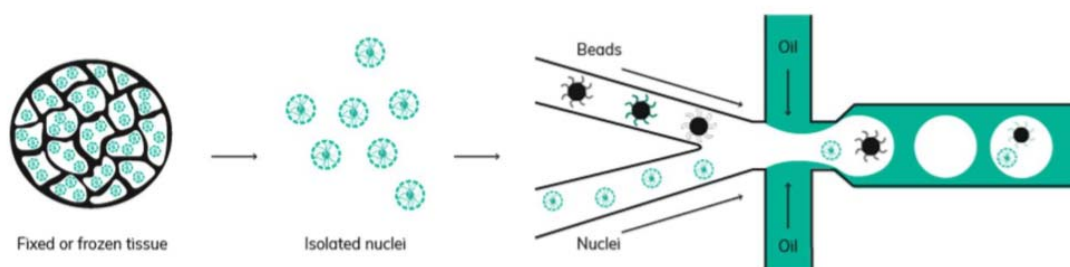


图 1 固定或冷冻组织样品分离的单个细胞核同编码捕获珠封装的示意图，成千上万的单核被唯一条形码的珠子封装。

细胞核在液滴内裂解，并且它们的 mRNA 被捕获珠捕获。

自从 Habib 的文章发表以来，细胞核封装技术越来越多地用于难以分离的细胞或储存的组织，并且正在迅速从工程演示转变为研究工具。最初使用的细胞核封装系统是由研究实验室的工程师定制的实验系统，然而生物学家在日常研究工作中使用这种技术并不是一件容易的事。在本文中，我们将描述在 Dolomite Bio Nadia 自动化平台上非常方便简单的实现 sNuc-Seq 方法。

3.材料和方法



图 2 Nadia 仪器

液滴系统： Dolomite Bio 的 Nadia 仪器（图 2）用于使用液滴微流控技术对单细胞和单细胞核进行高通量分析。Nadia 仪器使用三个独立的超稳定的压力泵来产生高度单分散的液滴。Nadia 软件采用引导提示模式一步一步的指导用户执行完成所选的应用方法。在整个实验过程中，Nadia 可以对样品进行冷却和搅拌，保持细胞转录组的状态，并且确保细胞和捕获珠均匀悬浮，最大化的降低双细胞的概率。用户也可以对 Nadia 上的程序做出定制化的更改，例如可以更改样品体积、样品数量和样品搅拌速度以适应不同的实验要求。

捕获珠和细胞核的制备和封装。从人体 HEK 293T 和小鼠 3T3 细胞的混合物中提取细胞核。本文中使用的材料和方法的详细描述可以在“sNuc-Seq on Nadia protocol”中找到。

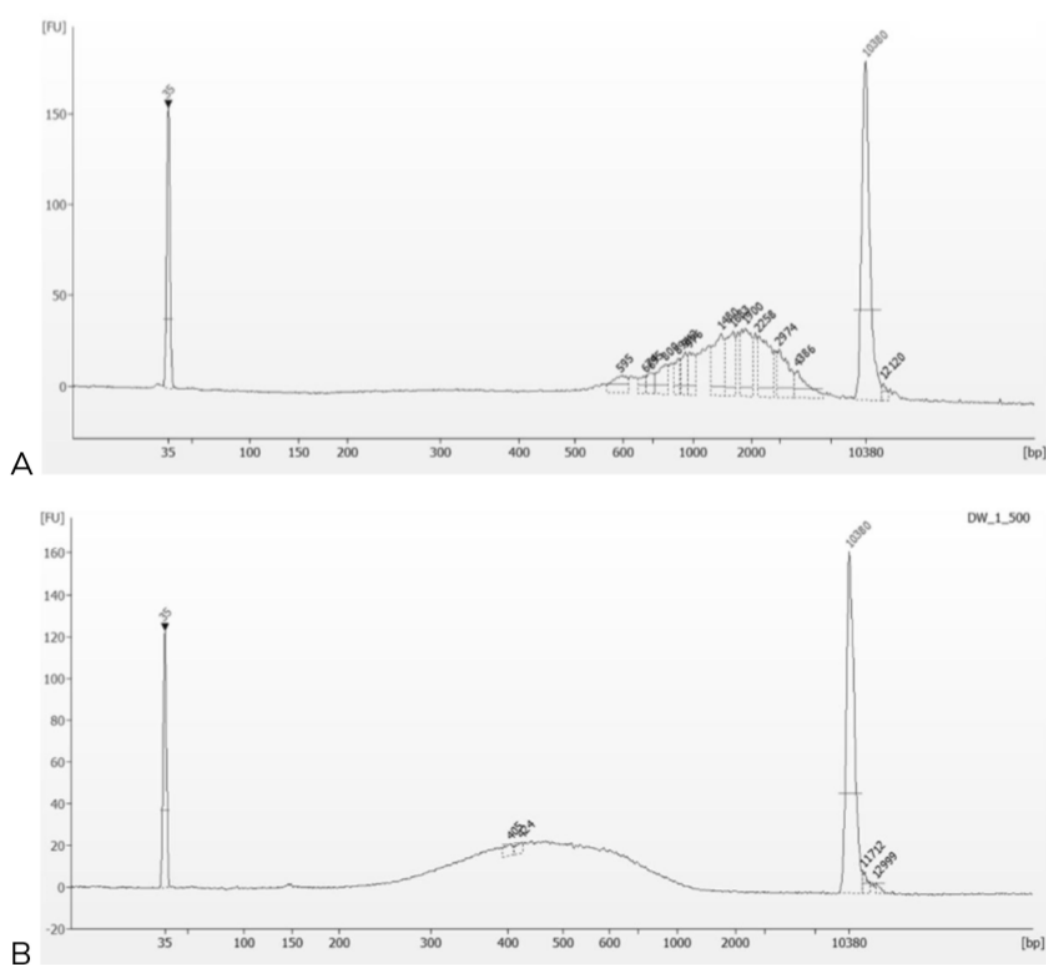
单核 DNA 库的制备和 NGS 分析。通过 Nadia 进行包裹之后的步骤请参考“sNuc-Seq on Nadia protocol”方案中所述来进行。在 4 个 PCR 反应中使用 5,000 个捕获珠，扩增总共 20,000 个捕获珠，使用 BioAnalyzer 测定 PCR 扩增和标记后的 DNA 质量，选出 200-250 个高质量细胞核数据做进一步的分析。

Illumina 测序和生物信息学流程。在 Illumina 的 NextSeq 500 仪器上对标记的 DNA 库进行测序，使用 2x75bp 配对端运行。读取长度配置如下：read 1 为 26 bp，index read 为 8 bp，read 2 为 116 bp。对于数据分析，“dropSeqPipe” (<https://github.com/Hoohm / dropSeqPipe>) 是由哈佛大学 McCarroll 实验室开发的计算流程，分析如 Macosko 等人 2015 文中所描述的方法进行。

4.结果

为了评估 Nadia 的性能以及整体工作流程的效率，使用“sNuc-Seq on Nadia protocol”方案进行了一系列混合物种实验。影响整体工作流程效率和数据质量的因素包括 PCR 重复率，基因捕获率和重现性。PCR 重复率由检测到的每个细胞核数据中相同 UMI（唯一分子标识符）的数量决定。基因捕获率由在给定测序深度检测到的基因数定义，基因捕获率代表了整个方案的效率。此外，比较不同用户在不同日期进行的实验的结果将给出重现性的信息。最后，比较同样从 Nadia 获得的 sNuc-Seq 和 scRNA-Seq 数据，对来自两种方法的转录组结果的相关性和差异性进行评估。

用于 NGS 分析的单核 cDNA 库的制备。从 Nadia 中回收乳液后，按照“sNuc-Seq on Nadia”方案中所述方法流程处理捕获珠。在 PCR 扩增（图 3A）和标记（图 3B）之后，使用 BioAnalyzer 分析一部分 cDNA 库，用来评估尺寸分布和总体质量。PCR 产物的尺寸分布（图 3A）是哺乳动物转录组的典型分布。测序标记库（图 3B）具有 300 至 1000bp 的预期尺寸范围。



Nadia 仪器 sNuc-Seq 方法的表现。 基于 knee-和 Barnyard 图以及样品中检测到的核糖体和线粒体 DNA 的百分比评估测序数据的质量。knee 图可以估计出每个样品的有效珠子的数量（即捕获有来自细胞核的 mRNA 的珠子）。每个珠子可以与具有唯一性的条形码和相应数量的 NGS 读数相关联。然后将条形码按读数的降序排列，并与读数的累积分数一起显示。在两个自制样品的 Nadia 单核测序数据库中，考虑到测序量的表现，有 210 个（图 4）和 250 个细胞核被识别出来。

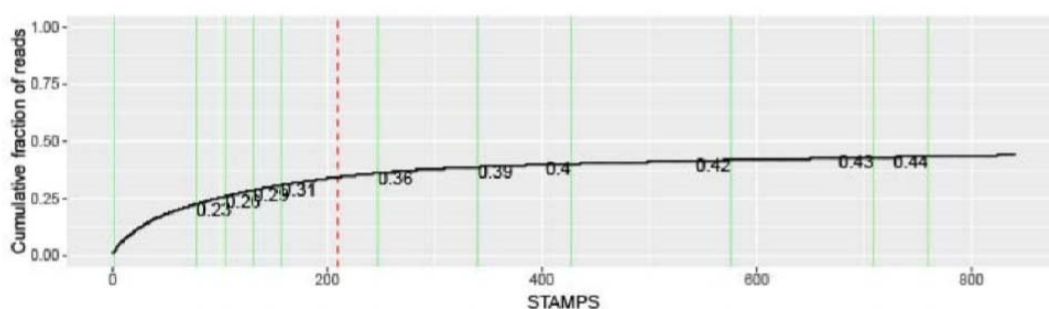


图 4 以读数的降序排列细胞条形码以及读数的累积分数的 knee 图

为了确定自制样品混合物种实验中的双重包裹率，使用了 Barnyard 图。该图将细胞核描述为单个数据点，每个数据点与一定数量人或小鼠转录本相关联（图 5）。数据显示，其中 91 个细胞核仅与小鼠转录本相关，113 个仅与人类转录本相关，而 6 个则带有来自两个物种的转录本。这相当于 5.7% 的双重包裹率。该双重包裹率与 Habib 等人 2017 年发表的混合物种数据相当，他们 DroNcSeq 数据的双重包裹率为 4.7%。

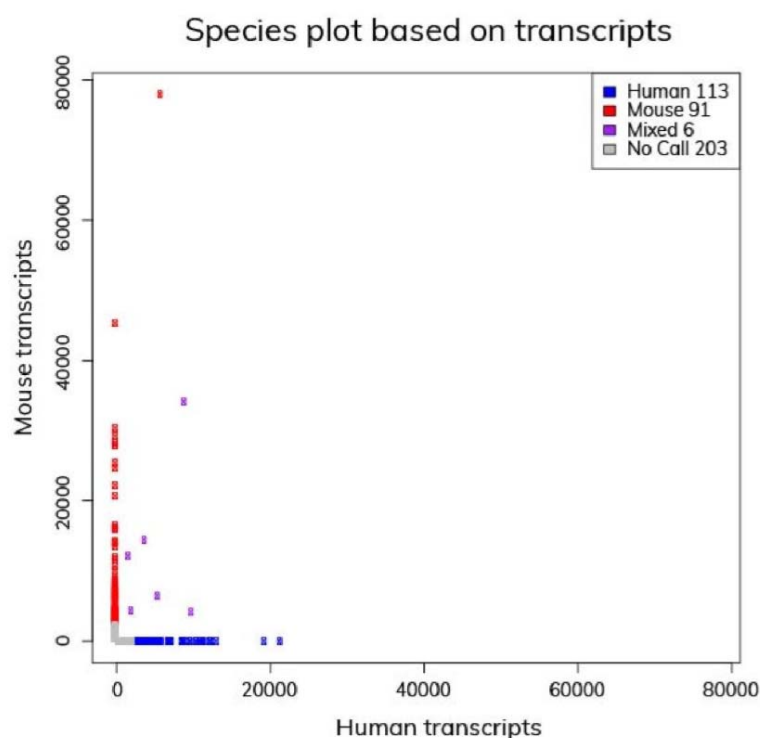


图 5 Barnyard 图描绘了细胞核作为单独的数据点，每个数据点与人或小鼠转录本关联

捕获的核糖体和线粒体 RNA 的数量是决定 sNuc-Seq 数据质量的重要因素，数量低表示提取细胞核的纯度高。从 Nadia 仪器获得的样品中核糖体和线粒体 RNA 读数的百分比低于 0.5%，并且与 Habib 等人 2017 年获得的值相当。(图 6)。

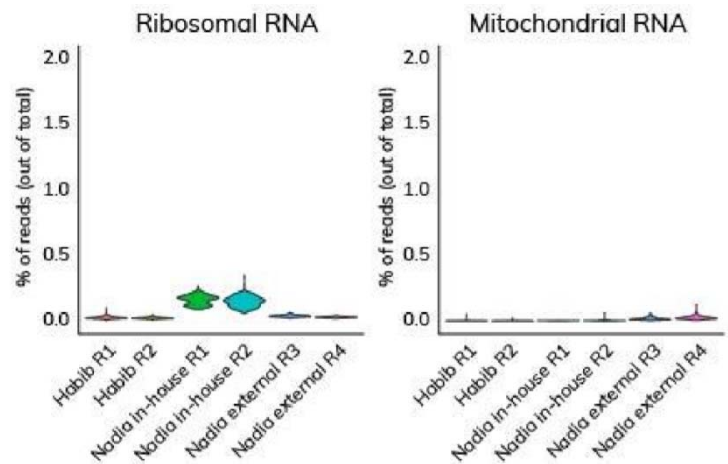


图 6 检测到的核糖体和线粒体 RNA 的百分比

PCR 重复率和基因捕获效率的评估

为了评估 PCR 重复率和基因捕获效率，将一组四个独立的 sNuc-Seq Nadia 运行数据与 Habib 等人 2017 年发表的 DroNc-Seq 数据进行比较。数据集由两组两个重复组成，每组的 200 个核，一组为自制样品，另一组为外来样品。对所有 Nadia 数据集进行测序，测序深度为大约 85,000 个读数/细胞核。表 1 中给出了比较的概况，该表总结了用 knee 图确定的细胞核数，每个细胞核的 NGS 读数和检测到的基因和 UMI（转录本）的中位数。

表 1 比较一组四个独立的 sNuc-Seq 在 Nadia 上的运行数据与已发表的 DroNc-Seq 数据

样本	核数	NGS 读数/核数	中位数	
			基因	UMIs
Nadia 自制 R1	210	86,761	2,420	4,397
Nadia 自制 R2	250	88,348	1,714	2,804
Nadia 外来 R3	200	85,672	1,921	3,306
Nadia 外来 R4	200	86,967	2,041	3,546
Habib R1	3499	54,424	1,226	1,831
Habib R2	2000	50,580	1,528	2,352

为了评估基因捕获效率，将六个数据集分别对基因数和读数作图（图 7），图 7 显示在 Nadia 仪器上获得的所有数据与 Habib 等人 2017 年发布的数据非常相似。

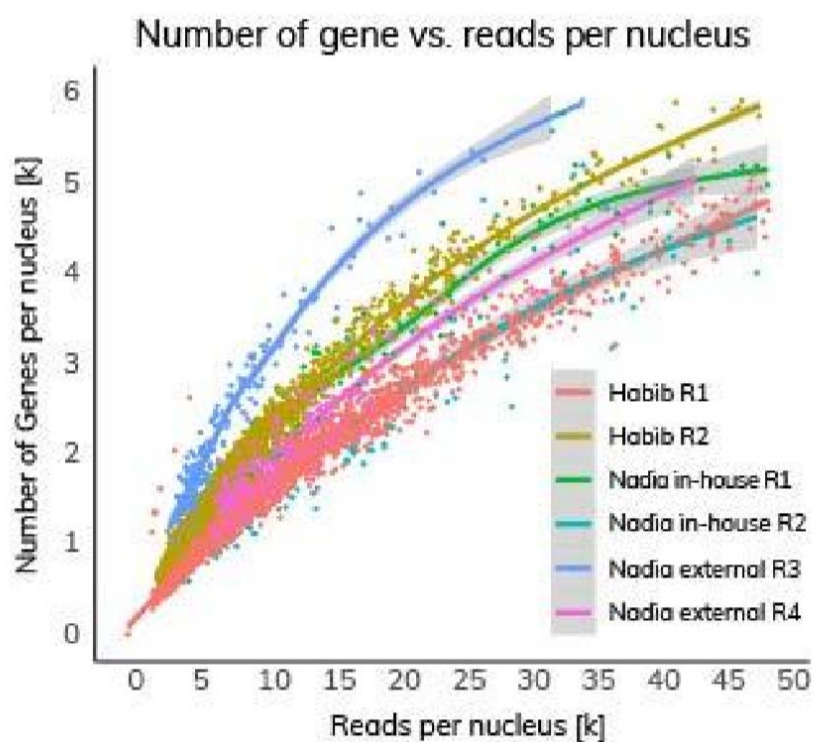


图 7 将检测到的基因数量与所有样本的读取深度进行比较

对 UMI 数量和细胞核读数作图，可以看出 cDNA 扩增过程中的 PCR 复制率和检测到的转录本数量（图 8）。该图中的黑线反映了 1 UMI/read 有的理想比率。Nadia 数据集上的所有 sNuc-Seq 都与此理想线非常接近。

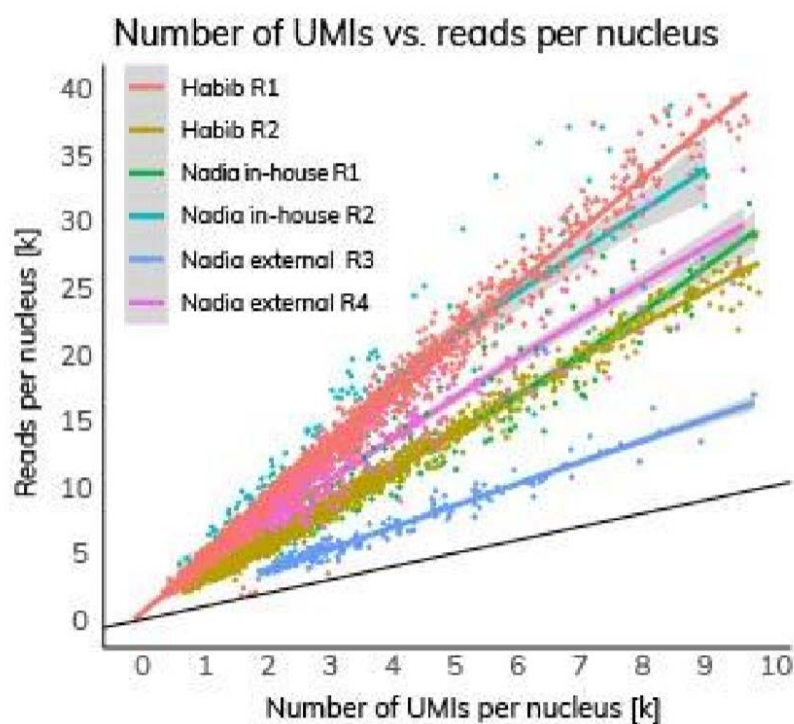


图 8 将检测到的 UMI 数量与样本中每个细胞核的读取深度进行比较。黑线代表 1 UMI/read 的理想比率，表明没有 PCR 重复。

单核和单细胞测序获得数据的比较

将在 Nadia 上的使用 sNuc-Seq 方法获得的自制样品数据与 Nadia 上的 scRNA-Seq 方法获得的数据进行比较，表明了数据的总体相似性，即不同数据中细胞和细胞核标记基因的表达以及内含子和外显子的百分比。散点图用于比较在 Nadia 上获得的 sNuc-Seq 和 scRNA-Seq 之间的所有条形码上的基因的平均表达。与单细胞相比，单核表达谱的平均 Pearson 相关系数是 0.92 (图 9A)。图 9 中以蓝色和红色分别代表富集的细胞核和细胞的基因。在 sNuc-Seq 和 scRNA-Seq 数据之间比较这些基因的表达水平。

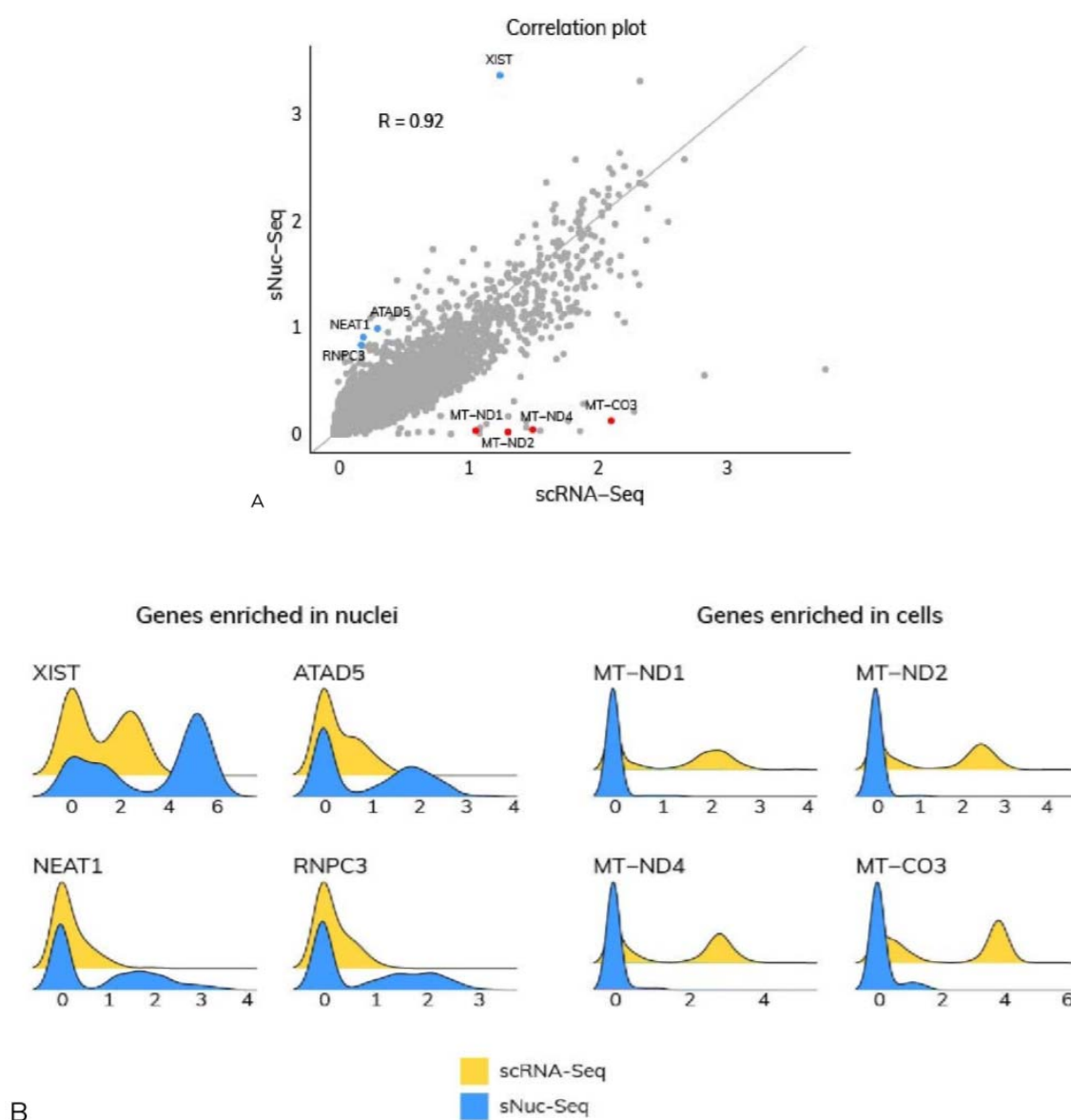


图 9 (A) 比较自制样品 scRNA-和 sNuc-Seq 数据的所有条形码的基因表达的散点图，细胞核（蓝色）或细胞（红色）(B) 图显示在 Nadia 获得的细胞或细胞核的基因 (x-axes) 的表达水平。sNuc-Seq (蓝色) 和 scRNA-Seq (黄色) 数据比较。

把通过 Nadia 获得的 scRNA-和 sNuc-Seq 数据与 Habib 等人 2017 年获得的 DroNc-Seq 数据的外显子和内含子的读数相对于检测到的总读数的比例进行比较（图 10）。细胞的外显子的读数比例接近约 90%，而对于细胞核，外显子读数占总读数的 50-60%。而细胞内含子的读数比例低于 10%，在细胞核中的内含子高达约 40%。

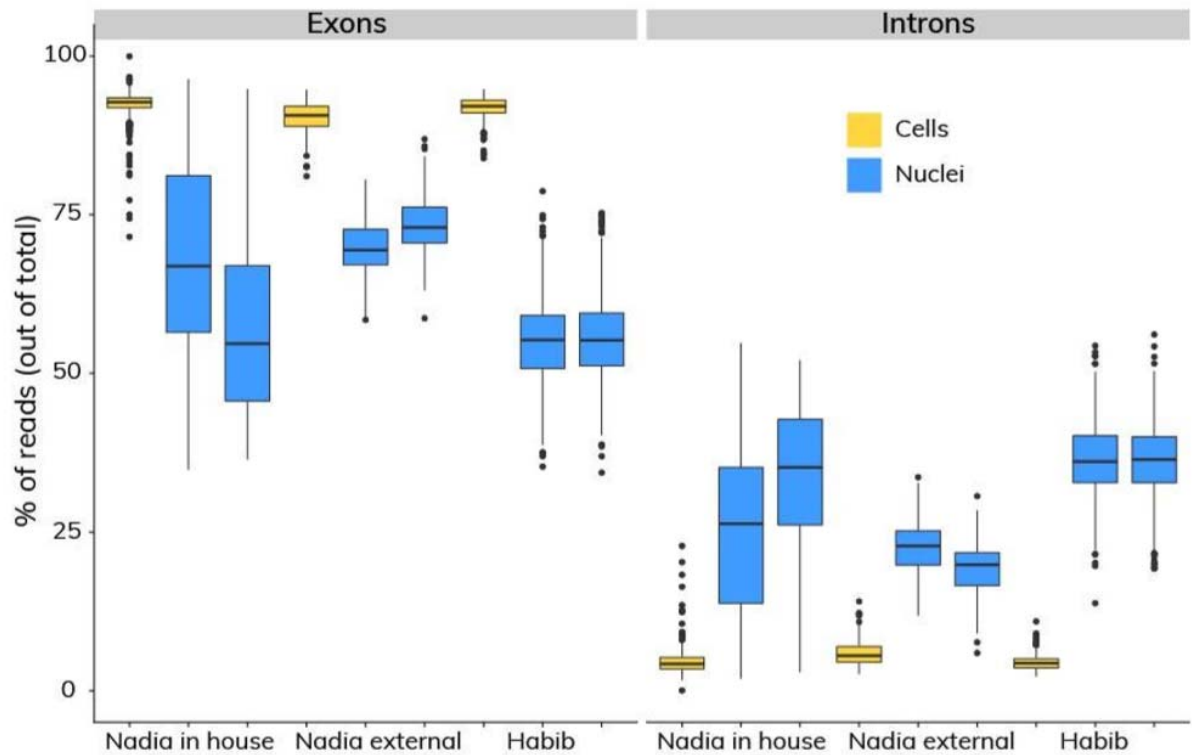


图 10 在 Nadia 获得的 scRNA-和 sNuc-Seq 数据与 Habib 等人 2017 年获得的 DroNc-Seq 数据之间比较

5.讨论

高通量单细胞核分析的发展,使得我们第一次可以分析固定的或冷冻的或不易分离的细胞（如神经元细胞和植物细胞）的基因表达谱。

本文中的数据论证了“sNuc-Seq on Nadia”方法的效率。通过分析双重包裹率，基因和转录本捕获效率，检测的核糖体和线粒体 DNA 的量以及通过比较从细胞和细胞核获得的数据来评估方案的性能。所有上述内容均在 Nadia 自制样品和外来样品获得的不同数据库之间进行了比较，以及和 Habib 等人 2017 年发布的数据库进行了比较。

Habib 等人 2017 年提出，分析一定数量的细胞核所需的珠子确切数量将高度依赖于所讨论的细胞类型，每个细胞核所需的读数的数量以及要进行的测序的量。该文章的研究人员测试了每个

样品使用不同数量的珠子，范围从 20,000 到 160,000 个珠子。为了将我们的结果与公布的数据进行比较，比较的数据来自 Nadia 使用自制样品和 20,000 颗珠子组合的结果。根据测序量的表现，平均有 200 个高质量细胞核能够被列入进一步分析的候选名单。与 Drop-Seq 相比，其中每个 PCR 反应 2,000 个珠子是最佳条件，Habib 等人 2017 年建议每个反应使用 5,000 个珠子用于 DroNc-Seq，该数量也适用于本文。然后通过计算双重包裹率并比较检测到的核糖体和线粒体 DNA 的量来评估在 Nadia 上产生的 sNuc-Seq 库的质量。两种情况的结果都获得了高质量的数据库并与 Habib 等人在 2017 年发表的论文中获得的结果相似。

也使用了每个细胞核被捕获的基因和转录组 (UMI) 的数量来评估 sNuc-Seq 方法在 Nadia 上的表现性能。在 Nadia 上获得的数据平均每个核检测到 2,000 个基因和 3,500 个转录组，读取深度为每个核约 80,000 个读数。并对这些结果与公布的 DroNc-Seq 数据进行了比较 (Habib2017)。相比之下，根据测序深度和分析的细胞类型的不同，使用 scRNA-Seq 方法在 Nadia 上进行单细胞 scRNA-Seq 测序，获得的检测基因的数量范围为每个细胞 3,000 至 5,500 个基因。考虑到与完整细胞相比，细胞核通常含有较低量的 RNA，因此细胞中捕获的基因和转录组的数量高于细胞核的数量，这种情况是正常的。

本文的另一个目的是比较使用相同细胞类型 (HEK 293T 和 3T3 细胞) 从 sNuc-Seq 和 scRNA-Seq 实验获得的转录组。文献表明，在很大程度上，从细胞核捕获的转录组代表了细胞的转录组 (1,3)。所以当使用细胞核代表细胞时，能够识别不同的细胞类型和分类。根据该假设，可以检测到 Nadia scRNA-和 sNuc-Seq 数据之间的强相关性。虽然发现一些基因在细胞或细胞核中特别多，但与细胞核相比，细胞中获得的基因通常多 5 倍。这可以通过以下事实来解释：scRNA-Seq 方案中使用的细胞裂解液也裂解细胞核。因此，可以预料到 scRNA-Seq 数据捕获来自细胞质以及细胞核的 mRNA 的信息。为此，从单细胞中获得的基因包括许多线粒体 RNA，例如 MT-ND1 和 MT-CO3，而单核转录本包括反义 RNA，例如 XIST。来自细胞核的转录组的特征是大量的新生 RNA。通过分析来自 Nadia 平台 sNuc-Seq 数据中的内含子和外显子读数的比例证实了这一点。正如所料，scRNA-Seq 数据包含更高比例的外显子读数和相应较低比例的内含子读数。总之，通过 Nadia 平台获得的 sNuc-Seq 数据中外显子与内含子读数比率，同 Habib 等人 2017 年的 DroNc-Seq 方法获得的数据作对比，是非常具有可比性的，证实了分离细胞核的纯度以及来自 Nadia 实验数据的高质量性。

6.总结

使用 Nadia 平台的 sNuc-Seq 方法，Dolomite Bio 获得了高质量的单核 cDNA 库，并进一步通过分析来评估整个工作流程的性能和效率。本工作评估了双重包裹率，PCR 重复率和基因捕获率，还比较了 sNuc-Seq 和 scRNA-Seq 数据，重点关注对细胞核或细胞特异的基因表达差异，以及内含子-外显子-读数比率，以确定细胞核的新生 mRNA 特征量。事实证明，Nadia 平台的 sNuc-Seq 方法不仅具有低的双重包裹率，而且具有高的基因捕获率和低的 PCR 重复率。这些结果证明 Nadia 平台的 sNuc-Seq 方法是分析单核转录组的有效方法。

7.参考文献

- [1] Habib et al. "Massively parallel single-nucleus RNA-seq with DroNcseq." Nature methods 14.10 (2017): 955.
- [2] Macosko E., et al. "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." Cell 2015 161:1202
- [3] Lake et al. "A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA" Scientific Reports volume 7(2017): 6031