

doi: 10.12452/j.fxcxb.26040708

Well-Multi-seq 实现多样品混合单细胞转录组测序

程蕊^{1,2}, 梁点一³, 许依玲⁴, 丘宇童⁴, 尹坤^{2,4*}, 杨朝勇^{2,4*}

(1. 上海大学 生命科学学院, 上海 200444; 2. 上海交通大学医学院附属仁济医院 分子医学研究所, 上海 200120; 3. 厦门大学 医学院, 福建 厦门 361005; 4. 厦门大学 化学化工学院, 福建 厦门 361005)

摘要: 单细胞转录组测序(scRNA-seq)为描绘细胞异质性提供了前所未有的分辨率, 然而传统 scRNA-seq 方法存在细胞损失率高、多细胞(Doublets)污染及成本高昂的瓶颈。为解决以上局限, 该文发展了 Well-Multi-seq 方法。通过结合尺寸排阻与准静态流体捕获的嵌套微孔芯片以及胆固醇修饰的寡核苷酸标签标记策略, 实现了高细胞捕获效率、低污染和低成本的多样品混合单细胞转录组测序。利用嵌套微孔芯片, 该文克服了泊松分布限制, 实现了高达94%的微球占有率、74%的细胞占有率及87%的细胞捕获率, 显著提升了细胞与微腔利用率。利用胆固醇分子稳定插入细胞膜的特性, 将带有样品编码的寡核苷酸链锚定于细胞表面, 实现了多样本一管混合测序。该策略不仅大幅降低试剂成本和简化实验流程, 还赋予了识别并剔除文库中多细胞污染的能力。在人鼠细胞系混合测序中, 该方法的样品标记准确率分别达92.60%(人)和95.28%(鼠), 证明了其卓越标记特异性。凭借 Well-Multi-seq 多重优势, 该方法在循环肿瘤细胞、微针活检样本等稀有细胞类型, 以及药物筛选和胚胎发育等动态过程的分析方面具有潜在的应用价值, 有望成为基础研究与临床诊断的有力工具。

关键词: 单细胞测序; 微流控芯片; 样本多重标记; 胆固醇修饰标签

中图分类号: O757.1; Q2 **文献标识码:** A **文章编号:** 1004-4957(2026)08-1644-09

Well-Multi-seq: A Method for Multiplexed Single-Cell Transcriptome Sequencing

CHENG Rui^{1,2}, LIANG Dian-yi³, XU Yi-ling⁴, QIU Yu-tong⁴, YIN Kun^{2,4*}, YANG Chao-yong^{2,4*}

(1. School of Life Sciences, Shanghai University, Shanghai 200444, China; 2. Institute of Molecular Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200120, China; 3. School of Medicine, Xiamen University, Xiamen, Fujian 361005, China; 4. College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, Fujian 361005, China)

Abstract: Single-cell RNA sequencing(scRNA-seq) provides unprecedented resolution for delineating cellular heterogeneity. However, conventional scRNA-seq approaches are bottlenecked by high cell loss rate, multicellular contamination, and high-cost. To address these limitations, we developed Well-Multi-seq, which integrates size exclusion and quasi-static hydrodynamics capture microfluidic chip and cholesterol-modified oligonucleotide tag labeling strategy. This approach enables scRNA-seq with high cell capture efficiency, low contamination, and low-cost. The microfluidic chip overcomes the Poisson distribution limit, achieving high bead (94%) and cell (74%) occupancy and high cell capture rate (87%), thereby significantly improving cell utilization. Leveraging the stable insertion of cholesterol molecules into cell membranes, sample-barcoded oligonucleotide strands are anchored onto the cell surface, allowing one-tube multiplexed sequencing of multiple samples. This strategy not only drastically reduces reagent costs and simplifies experimental workflows but also endows the system with the ability to identify and remove doublet contamination from the library. In mixed human-mouse cell line sequencing, the sample labeling accuracy of this method reaches 92.60%(human) and 95.28%(mouse), demonstrating its superior labeling specificity. Owing to its multiple advantages, Well-Multi-seq is particularly suited for multiplexed analyses of rare cell types such as circulating tumor cells and microneedle biopsy samples, as well as dynamic processes includ-

收稿日期: 2026-04-07; 修回日期: 2026-04-28

基金项目: 国家重大项目/重大研究计划集成项目(22293031); 国家青年科学基金项目(22404106)

* 通讯作者: 尹坤, 博士, 研究方向: 微流控技术在生物医学中的应用, E-mail: yinkun04@163.com
杨朝勇, 博士, 教授, 研究方向: 微流控技术在生物医学中的应用, E-mail: cyyang@xmu.edu.cn

网络首发日期: 2026-05-20

增强出版



ing drug screening and embryonic development, and is poised to become a powerful tool for basic research and clinical diagnosis.

Key words: single-cell RNA-sequencing; microfluidic chip; sample multiplexing; cholesterol-modified tag

单细胞转录组测序(scRNA-seq)技术的出现改变了我们对生物系统复杂性的认知。通过在单个细胞分辨率下解析基因表达谱, scRNA-seq能够揭示细胞类型的多样性、描绘发育轨迹以及解析肿瘤微环境的异质性^[1-4]。尽管该技术已成为现代生物学和医学研究的关键手段,但在面对珍贵临床样本(如循环肿瘤细胞、微针活检样本)或需要大规模平行实验(如药物筛选、胚胎发育时序分析)时,现有的主流技术平台仍面临严峻挑战^[5-6]。

当上述平台被应用于多样本队列研究时,问题进一步凸显。目前,商业化的scRNA-seq平台主要基于微滴(Droplet-based)或微孔板(Well-based)技术。然而,这两类方法均存在显著的局限性。首先,基于微滴的技术(如Drop-seq)虽然通量高,但其细胞捕获过程需遵循泊松分布(Poisson distribution)。为确保单细胞封装,通常需要将细胞浓度控制在较低水平,这导致大量细胞未被捕获而损失,细胞利用率往往低于10%^[7]。其次,在单细胞分离过程中,提高细胞浓度以增加捕获率会显著增加“双细胞”(Doublets,即两个或多个细胞被包裹在同一微单元中)的概率,从而引入严重的技术噪音和数据污染,影响下游分析的准确性^[8]。此外,高昂的试剂成本和复杂的多样本处理流程也限制了其在大规模队列研究中的广泛应用。当上述平台被应用于多样本队列研究时,问题进一步凸显,传统的多样本分析通常需要对每个样本单独建库测序,这不仅大幅增加了实验成本,还引入了批次效应(Batch effects),使得样本间的比较变得复杂^[9-10]。现有的细胞标记技术,如基于抗体标记(Cell hashing)虽在一定程度上实现了样本混合,但通常存在抗体成本较高和普适性低等问题^[11-14]。因此,开发一种既能实现高细胞捕获和低成本的多样本并行分析,且能有效识别并剔除双细胞污染的综合型平台,仍是单细胞领域亟待突破的关键瓶颈。

本课题组开发了一种集成嵌套微孔微流控芯片与化学修饰标签标记策略的单细胞转录组测序平台(Well-Multi-seq)。针对细胞捕获难题,我们设计了一种结合尺寸排阻(Size exclusion)与准静态流体捕获(Quasi-static hydrodynamic capturing)原理的嵌套微孔芯片。利用该原理,Well-Multi-seq克服了传统方法对泊松分布的依赖,实现了高微球占有率(94%)、高细胞占有率(74%)以及高细胞捕获率(87%)。利用胆固醇分子稳定插入细胞膜脂双分子层的特性,我们将带有样品编码的寡核苷酸链锚定在细胞表面,实现了多个样本在单一反应管中的混合建库与测序,这一策略大幅降低了试剂成本并简化了流程。该方法在人鼠细胞系混合测序中具有卓越的标记特异性,实现了92.60%人细胞和95.28%鼠细胞的样品识别准确率,同时成功利用标签信息鉴定出文库中的双细胞事件。凭借高细胞利用率、低成本和低污染等多重优势,Well-Multi-seq在循环肿瘤细胞、微针活检等稀有样本研究,以及药物筛选和胚胎发育等动态过程的系统解析等方面具有潜在应用价值。该平台有望为基础生物学分析提供一种高效、经济且可靠的新工具。

1 实验部分

1.1 微流控芯片制作

双层嵌套微孔芯片模具使用软刻蚀工艺制备,借助两次套刻曝光完成掩模制作,进而形成双层柱状结构。先取硅片作为基底,在其表面旋涂SU-8 3050光刻胶,将厚度控制在50 μm ,以此制作微球捕获微孔;继续旋涂SU-8 2015光刻胶,控制厚度为25 μm ,加工得到细胞捕获微孔。完成坚膜烘烤工序后,取用体积分数为1%的1H, 1H, 2H, 2H-全氟辛基二甲基氯硅烷/HFE-7500溶液,对模具开展硅烷化钝化处理。

把聚二甲基硅氧烷(PDMS)预聚液与固化剂按照10:1的质量比充分混匀,将混合液均匀浇注在处理好的模具上,置于100 $^{\circ}\text{C}$ 环境下固化15 min,待弹性体彻底凝固后进行脱模,得到带有对应纹路的图案化PDMS模块。随后将此模块与另一块预留进样口和出样口的开槽PDMS模块紧密键合,密封拼接后形成微流道,流道厚度为500 μm ,宽度为0.5 cm,长度为1.2 cm。最后用0.5%牛血清白蛋白(BSA)溶液对芯片进行封闭处理,处理完毕后放入60 $^{\circ}\text{C}$ 环境中烘干过夜。

1.2 细胞培养

实验所用 K562 与 3T3 细胞, 均在 37 °C、5% CO₂ 条件下培养。细胞采用高糖杜氏改良 Eagle 培养基 (DMEM H-21), 加入 10% 胎牛血清 (FBS) 和青霉素/链霉素混合液 (分别为 100 U/mL 和 100 μg/mL)。

1.3 胆固醇标签的设计与合成

将 5' 端修饰疏水胆固醇的寡核苷酸链 (5'-Cholesterol-oligo) 与 3' 端修饰疏水胆固醇的寡核苷酸链 (3'-Cholesterol-oligo) 置于杂交缓冲液中进行杂交, 杂交缓冲液采用 1×PBS 缓冲液 (含 100 mmol/L NaCl, pH 7.4), 杂交在冰上进行, 时间维持 5 min; 随后, 向体系中加入第三条寡核苷酸 (multi X, 其 3' 端含 polyA 尾、中间含 8 bp 样品标签、5' 端含 PCR 柄), 继续在上述相同缓冲液及冰上条件杂交 5 min, 使 5'-Cholesterol-oligo 与 multi X 发生部分互补杂交, 最终形成目标杂交结构。寡核苷酸链均由生工生物工程 (上海) 股份有限公司合成。

1.4 胆固醇标签稳定性荧光验证

取 5×10^5 个 K562 细胞, 用含 0.04% BSA 的 1×DPBS 缓冲液稀释至 920 μL, 制备单细胞悬液。将 20 μL 10 μmol/L 的 5'-Cholesterol-oligo 与 20 μL 10 μmol/L 的 3'-Cholesterol-oligo 混匀, 冰上静置 5 min 后加入细胞悬液中, 轻柔吹打 20 次并在冰上孵育 10 min。另取 20 μL 10 μmol/L 的 multi 1 后加入上述体系, 再次吹打混匀并在冰上孵育 10 min。450×g 离心 3 min, 弃上清, 用含 0.04% BSA 的 1×DPBS 重复洗涤细胞两次。随后分别孵育 30、60、90 min 和 7 h 后, 加入 20 μL 10 μmol/L 的 polyT-Cy3 溶液, 孵育 5 min。离心清洗 3 次后, 取 1 μL 细胞悬液滴加在载玻片表面, 静置 30 s 后, 使用 10×物镜采集荧光图像并检测荧光强度, 分析孵育不同时间后胆固醇标签的稳定性。

1.5 细胞样本标记

取 5×10^5 个 K562 和 3T3 细胞, 分别用 940 μL 含 0.04% BSA 的 1×DPBS 重悬。将 20 μL 10 μmol/L 的 5'-Cholesterol-oligo 与 20 μL 10 μmol/L 的 3'-Cholesterol-oligo 混合均匀, 冰上孵育 5 min 后均分为两份, 分别加入两种细胞悬液中, 轻柔吹打混匀, 冰上孵育 10 min。随后向 K562、3T3 细胞中分别加入 20 μL 10 μmol/L 的 multi 1、multi 2, 轻柔吹打后冰上孵育 10 min。孵育后细胞经 450×g 离心 3 min, 使用含 0.04% BSA 的 1×DPBS 清洗 2 次。最后取 30 μL 两种处理后的细胞混匀备用。

1.6 单细胞 RNA 捕获

向芯片中通入 1×DPBS, 超声去除气泡。单细胞捕获过程中, 先吸出芯片内的液体, 再缓慢加入细胞悬液, 静置 5 min 待细胞捕获, 吸出内部细胞悬液并加入含 0.04% BSA 的 1×DPBS 溶液冲洗芯片, 直至镜检显示边缘无未捕获细胞。随后进行编码微球捕获, 用含 0.04% BSA 的 1×DPBS 重悬微球后快速加入芯片, 静置铺展微球, 待微球铺满后吸净多余悬液, 用 Prelysis buffer (20×Tris-EDTA, 含 50 mmol/L DTT) 冲洗芯片至边缘无游离微球。接着向芯片中加入矿物油, 再缓慢加入细胞裂解液 (0.05 g 十二烷基肌氨酸钠分散在 1 mL 矿物油中, 超声混匀) 并室温静置 10 min 完成细胞裂解。之后切开芯片通道, 加入 6×SSC 吹打回收微球, 随后分别使用 6×SSC 清洗微球 2 次, 使用 1×RT buffer 清洗 1 次。

1.7 RNA 文库构建

向微球中加入逆转录反应试剂 (含 10 U/μL Maxima H Minus 反转录酶) 充分混匀, 42 °C 旋转孵育 90 min; 逆转录反应结束后, 微球离心, 去除上清液, 依次用 TE-SDS (10 mmol/L Tris-HCl (pH 8.0), 1 mmol/L EDTA, 0.1% SDS)、TE-TW (10 mmol/L Tris-HCl (pH 8.0), 1 mmol/L EDTA, 0.05% Tween 20) 及 10 mmol/L Tris-HCl (pH 8.0) 溶液洗涤。随后向微球中加入外切酶混合液 (含 1 U/μL EXO I), 37 °C 旋转孵育 45 min 完成酶切反应, 去除微球上未捕获到 RNA 的编码序列; 反应结束后分别使用 TE-SDS、TE-TW 及 Tris-HCl 清洗微球以去除残留试剂。然后加入二链合成混合液 (含 0.5 U/μL Klenow 片段), 在 37 °C 条件下旋转孵育 60 min 完成二链合成, 反应结束后清洗微球。随后加入扩增反应液 (含 1×HiFi 扩增预混液), 设置 PCR 程序完成扩增反应。扩增结束后采用磁珠法纯化并分选 PCR 产物, 分为 cDNA 产物与胆固醇标签, 并通过 Qubit 定量确定产物浓度, 确保满足后续实验需求; 最后 cDNA 产物使用诺唯赞建库试剂盒构建测序文库, 胆固醇标签通过扩增加入测序接头构建测序文库, 分别纯化两种文库。最后通过二代测序分别检测两个文库的序列信息, 后续进行生物学分析。

1.8 数据分析

具体细节见补充文件。

2 结果与讨论

2.1 Well-Multi-seq 设计与工作流程

当细胞与微球在重力作用下完成高效配对后, 细胞裂解释放的 mRNA 与含有样本标签的寡核苷酸链将被同步捕获至同一编码微球上。随后统一进行微球回收, 并依次经逆转录、外切酶消化、二链合成及 PCR 扩增等步骤, 分别构建转录本文库与样本标签文库, 纯化分离后进行高通量测序, Well-Multi-seq 平台通过创新性结合尺寸排阻与准静态流体的嵌套微孔芯片与胆固醇修饰的寡核苷酸链标记策略, 构建了高细胞捕获率、低污染和低成本的多样品混合单细胞转录组测序新方法。Well-Multi-seq 平台通过创新性结合尺寸排阻与准静态流体的嵌套微孔芯片与胆固醇修饰的寡核苷酸链标记策略, 构建了高细胞捕获率、低污染和低成本的多样品混合单细胞转录组测序新方法。为了实现多样品混合测序, 我们首先利用胆固醇分子稳定插入细胞膜脂双分子层的特性^[15-16], 设计了胆固醇修饰的寡核苷酸标签, 赋予每个样本独特的样品标签。为了增加标记的稳定性, 将标签设计为三链结构(如图 1A)。利用碱基互补策略, 将胆固醇修饰且长短不一的两条寡核苷酸链组装, 在底部形成双分子胆固醇结构。然后, 加入含有引物结合序列、样品标签和 PolyA 尾巴的寡核苷酸链与上述长寡核苷酸链互补组装, 最终形成三链结构。对每个样本标记后, 将多管细胞样本混合并通入微流控芯片中。通过尺寸排阻的嵌套式微孔保证上下微孔只允许捕获单个细胞与编码微球, 并利用准静态流体力学防止细胞逃逸, 从而突破泊松分布限制实现细胞累积捕获。当细胞与微球在重力作用下完成高效配对后, 细胞裂解释放的 mRNA 与含有样本标签的寡核苷酸链将被同步捕获至同一编码微球上, 随后统一进行微球回收、反转录、cDNA 扩增及文库构建。在后续的数据分析中, 该方法通过样本标签精准地划分了样本归属。利用样本标签的共现分析结合 PDF 策略, 成功识别并剔除了文库中的异源双细胞事件, 当细胞与微球在重力作用下完成高效配对后, 细胞裂解释放的 mRNA 与含有样本标签的寡核苷酸链将被同步捕获至同一编码微球上。随后统一进行微球回收, 并依次经逆转录、外切酶消化、二链合成及 PCR 扩增等步骤, 分别构建转录本文库与样本标签文库, 纯化分离后进行高通量测序, 从而实现高通量、低成本且低污染的多样本混合单细胞测序分析(图 1B)。

2.2 芯片性能与捕获效率表征

为评估 Well-Multi-seq 平台的物理捕获性能, 本实验首先对嵌套式微孔芯片的结构特征与流体行为进行详细表征。在结构设计上, 芯片由数万个位于液体通道下方的同心堆叠双层微孔组成(图 2A)。由于大多数细胞的直径小于 25 μm , 因此我们将下层微孔设计为边长 25 μm 的方形结构; 上层微孔则设计为内切圆直径与高度均为 50 μm 的六边形结构, 以适配直径为 30~50 μm 的功能性编码微球。这种基于尺寸排阻的双层设计, 从物理空间上确保了每个微孔单元仅能容纳单个细胞与单个微球, 为后续的一对一配对奠定了基础。此外, 由于微流控通道内的流动主要为层流, 当流体流经芯片表面时, 流速在垂直方向上呈现显著的梯度分布(图 2B)。具体而言, 位于主通道下方的上层微孔区域受主流体剪切力影响较小, 流速急剧下降; 而位于更深层的下层细胞捕获孔, 其内部流体几乎处于静止状态(准静态)。这种流速的逐级骤降不仅有效防止已捕获细胞的逃逸, 还同时支持细胞在多次加载过程中的累积捕获, 突破了传统单次加载的效率限制。

为进一步验证该芯片的单细胞分离与配对效率, 本课题组以 K562 细胞系和 30~50 μm 编码微球为模型进行测试。显微镜观察与统计分析显示(图 2C-D), 嵌套微孔芯片实现了极高的单细胞占用率(约 74%)与微球占用率(约 94%), 最终产生高达 74.3% 的细胞-微球有效配对率。与传统的基于泊松分布的微孔或微滴平台相比, 该方法的配对效率提升约 5~8 倍, 提高了细胞的分析通量。此外, 得益于累积捕获效应, 经过 3 次连续的细胞沉降与重悬步骤后, 芯片对投入细胞的总捕获率高达 87%(图 2E)。这一结果表明, Well-Multi-seq 芯片不仅能高效实现细胞与微球配对, 还能高效捕获投入的细胞, 使其在循环肿瘤细胞等稀有样本分析中展现出巨大的应用潜力^[17]。

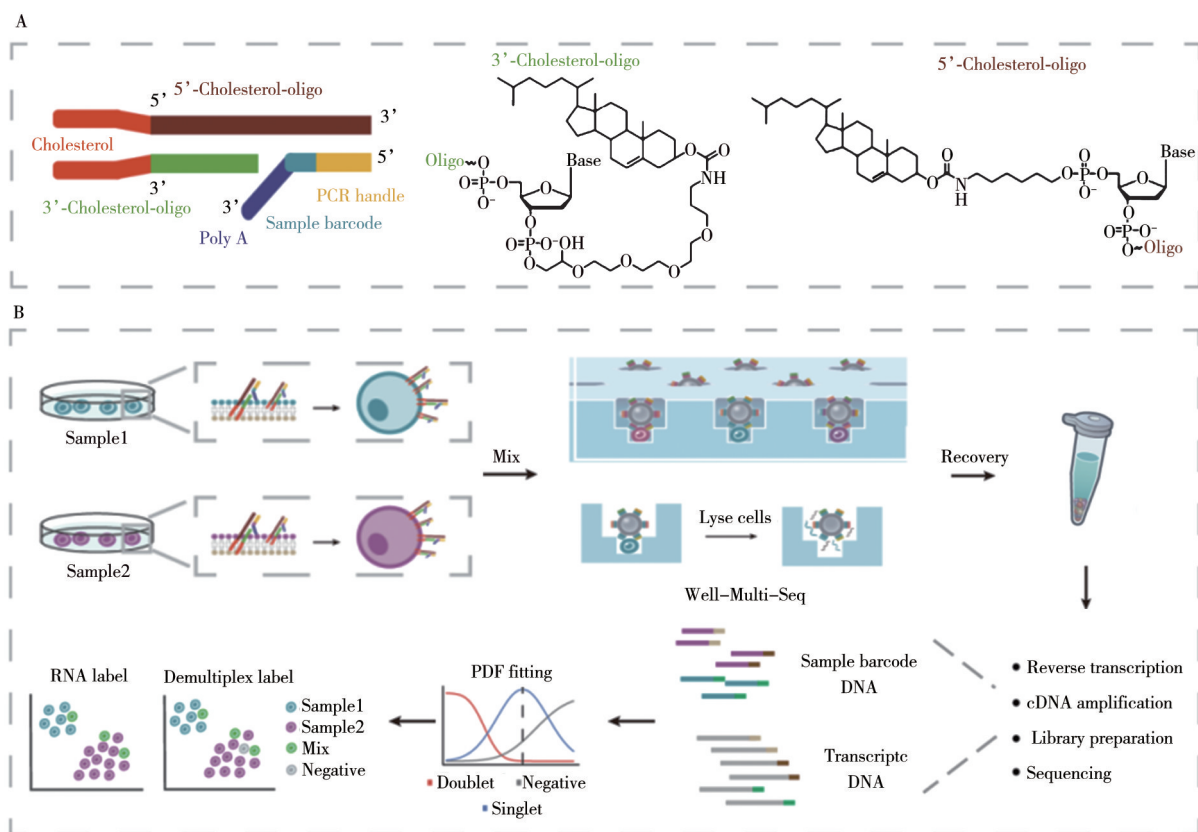


图 1 Well-Multi-seq 混样拆分技术实验的流程示意图

Fig. 1 Schematic diagram of the experimental workflow for sample pooling and demultiplexing in Well-Multi-seq
 A: special triple-stranded structure carrying sample barcodes (带有样品标签的特殊三链结构); B: overall workflow of Well-Multi-seq, encompassing sample labeling, cell-microsphere pairing, co-capture, library construction, and high-throughput sequencing (Well-Multi-seq 整体工作流程, 包括样品标记、细胞-微球配对、共捕获、文库构建与高通量测序)

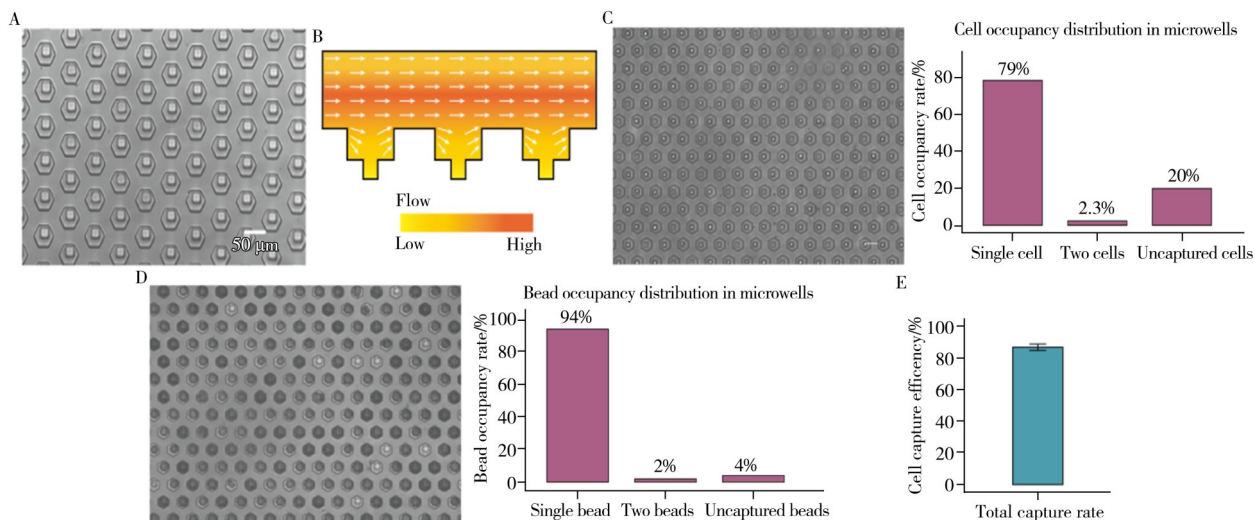


图 2 Well-Multi-seq 芯片的性能与捕获效率表征

Fig. 2 Characterization of chip performance and capture efficiency in Well-Multi-seq

A: microscopic image of microwells on the chip, scale bar: 50 μm (芯片微孔的显微成像图); B: schematic diagram of fluid flow within microfluidic channels. The color gradient and streamline thickness jointly represent the fluid flow velocity. Thicker streamlines and warmer colors indicate higher flow velocity, while thinner streamlines and colder colors indicate lower flow velocity. The color scale at the bottom represents the flow velocity ruler (微流控通道内的流动流体示意图, 图中颜色梯度与流线粗细共同表征流体流速。流线越粗及颜色越暖, 代表流速越高; 流线越细、颜色越冷, 代表流速越低。下侧色标为流速标尺); C: microscopic images of captured cells and bar plot of cell occupancy rate (细胞被芯片捕获的显微镜图像和细胞占有率条形图); D: microscopic images of captured microspheres and bar plot of microsphere occupancy rate (微球被芯片捕获的显微镜图像和微球占有率条形图); E: bar plot of cell capture efficiency (细胞捕获效率条形图)

2.3 胆固醇修饰的三链结构细胞标记稳定性表征

为探究胆固醇修饰的三链结构标记细胞的稳定性,将三链结构插入细胞膜上,然后加入含有荧光的互补链表征细胞膜插入分子数,通过表征细胞荧光随时间的变化,即可探究该三链结构的稳定性。首先对细胞进行标签标记,分别在30、60、90 min及7 h四个时间点进行孵育,随后加入荧光互补链,利用荧光显微镜采集细胞明场与荧光图像(图3A)。结果表明,各时间点下,细胞均呈现出清晰且均匀的荧光信号且荧光强度随孵育时间延长无明显降低(图3B)。该结果证实,胆固醇修饰的三链结构在插入细胞膜后,能够牢牢锚定在细胞膜上,不易发生脱落解离,可长期维持稳定的细胞结合状态,具备优异的细胞标记稳定性^[15-16]。

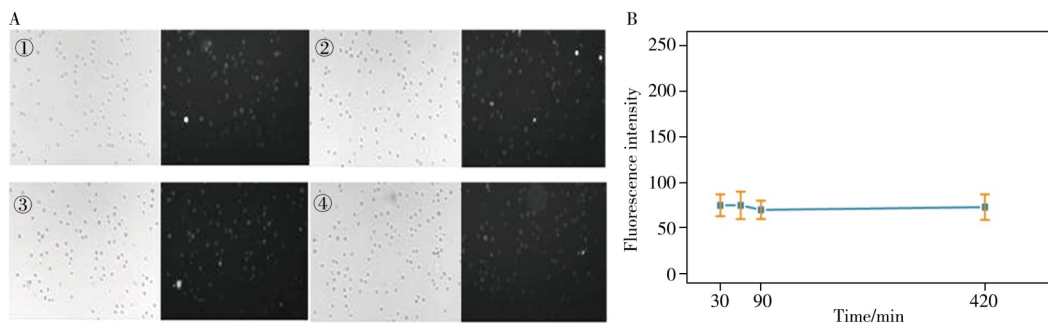


图3 Well-Multi-seq 三链结构细胞标记稳定性表征

Fig. 3 Characterization of cell labeling stability of the triple-stranded structure in Well-Multi-seq

A: comparison of bright-field and fluorescence images of cells; numbers ①-④ correspond to cell imaging results at staining durations of 30 min, 60 min, 90 min, and 7 h, respectively (细胞明场成像与荧光成像对比图, 序号①-④依次对应染色时长30 min、60 min、90 min和7 h的细胞成像结果); B: line plot of the trend of cellular fluorescence intensity varying with labeling time (细胞荧光强度随标记时间变化趋势折线图)

2.4 测序数据质量与转录组捕获效率评估

在单细胞高通量测序中,数据质量直接决定了后续样本拆分的可靠性与生物学分析的可解释性。本文在引入新型胆固醇标签标记体系后,首先表征了此外源化学修饰是否会影响数据的质量。转录组文库的质控结果显示,有效比对到转录组的 Reads 占比高达81.63%,而非目标片段(如接头污染或降解产物)占比仅为3.01%(补充表1),表明文库片段长度分布集中,无明显降解,且胆固醇标签的引入未造成显著的背景噪音。与此同时,专门用于样本身份识别的样本标签文库表现更为出色,其有效 Reads 占比达98.55%,非目标片段仅占3.10%(补充图1A)。极高的信噪比证实了胆固醇修饰标签在复杂混合体系中具有优异的检出特异性和稳定性,满足多样本混合测序后的高精度拆分需求。

为了深入评估单细胞水平的基因检出能力与数据分布特征,本研究利用累计曲线法筛选有效细胞。如补充图1B所示,将细胞标签按 Reads 数降序排列后,其累计 Reads 占比曲线呈现典型的“快速上升后逐步饱和”特征。曲线初期的陡峭上升表明绝大多数测序数据集中于高质量的有效细胞中,而随后的平缓段则意味着低质量细胞或背景噪音的干扰信号占比极低。通过在曲线拐点处进行线性拟合,我们最终将有效细胞的阈值界定为2 750条 reads(补充图1B红点处)。在此基础上,进一步表征了每个单细胞检测到的基因数随测序深度的变化关系(补充图1C-D)。实验数据显示,在平均测序深度约为20 000 Reads/cell的条件下,该方法在每个细胞中平均检出4 145.6个基因和7 189.2个转录本,证明了该技术具有较好的基因检出能力,并未因多步化学标记操作而损失转录组信息。

综上所述,胆固醇修饰的核酸标签的引入未对 Well-Multi-seq 平台的转录组捕获效率产生负面影响,也未破坏数据的完整性。整体测序数据质量稳定可靠,为后续复杂的生物学挖掘奠定了坚实的数据基础。

2.5 基于PDF拟合的胆固醇标签阈值优化与细胞分型

在单细胞测序数据分析中,双细胞(即两个细胞被错误地分配了同一个标签)的存在会严重干扰细胞聚类分析与样本归属判断^[18]。现有的纯计算算法主要依赖转录组特征的异常共表达来识别双细胞,但这类方法易受细胞异质性、测序深度波动等影响,往往存在假阳性率较高的问题^[19-21]。针对这一挑战,本研究利用胆固醇修饰寡核苷酸标签的物理标记特性,提出了一种不依赖转录组特征的直接判定

策略。其原理在于样本特异性标签的共现模式,若一个细胞微球单元同时检出两个不同样本的高强度特异性标签信号(例如同时出现高丰度的人源标签和鼠源标签),则可直接判定该单元为异源双细胞。

为了进一步提升该识别策略的客观性及稳健性,我们将上述标记体系与基于PDF拟合的阈值筛选策略相结合^[22]。通过对样本标签的UMI计数信号进行标准化处理与分布拟合,实现背景噪声与真实信号的自动分离,从而精准识别双细胞并高效拆分混合样本。为验证样本拆分的准确性,本研究为人源和鼠源样本分别设计了两套正交的胆固醇修饰寡核苷酸样本标签:人源细胞特异性标签(HC-bar)和鼠源细胞特异性标签(MC-bar)。基于PDF拟合策略,分别对混合样本中每个细胞的HC-bar和MC-bar信号强度分布进行概率密度拟合与阈值优化。原理如补充图2A所示,对样本标签UMI计数经标准化预处理后,通过核密度估计拟合信号分布并识别有效信号与噪声峰值区间,以单标签细胞占比最大化为目标确定阈值,最终将细胞划分为单细胞、双细胞及阴性细胞。该方法不依赖转录组特征,仅依靠标签信号即可实现高效、可靠的双细胞判定。

分析结果如补充图2B-C所示,在人源与鼠源混合样本中,HC-bar的UMI计数均呈现清晰的双峰分布(Bimodal distribution)。左侧低值峰对应于背景噪声或非人源细胞(阴性群体),右侧高值峰对应于携带有效人源标签的细胞(阳性群体)。两峰之间界限分明,通过定位双峰间的谷底(Valley),即可直接确定最优分类阈值,实现人源细胞的精准提取。类似地,经过标准化预处理后,MC-bar的信号分布也恢复了清晰的双峰形态,有效区分了鼠源细胞与背景/非鼠源细胞。这种基于物理标签的双峰分布特征,避免了传统算法对转录组数据质量的依赖。通过设定基于PDF拟合的动态阈值,本研究建立的分类体系能够客观、无偏地将细胞划分为3类:单标签细胞(仅含HC-bar或MC-bar,判定为真实单细胞),多标签双细胞(同时含有高强度的HC-bar和MC-bar,判定为异源双细胞)以及阴性捕获单元(两者均缺失,判定为空微球或背景)。

综上所述,本文提出的基于PDF拟合的胆固醇修饰寡核苷酸标签分类体系,提供了一种无需主观经验参数的自动化分型方法。该方法显著提升了单细胞数据的纯度,为后续高精度的生物学分析提供了可靠保障。

2.6 多样品混合拆分的可行性验证

为了考察该平台对多样品的混合拆分情况,首先开展了细胞系物种混合实验。将标记HC-bar的人K562细胞与标记MC-bar的小鼠NIH/3T3细胞按照数量比1:1进行混合,然后通入芯片中进行细胞捕获以及文库构建。将以上细胞编码测序reads的cDNA信息与标准的人与小鼠数据库进行比对,确认物种来源,并统计各细胞编码对应的UMI数量。规定当某一细胞编码中大于90%的UMI比对到人类基因组时,其属于人源细胞;同样,当细胞编码中大于90%的UMI比对到小鼠基因组时,其属于鼠源细胞;其余均视为混合细胞,该情况可能是由于单个配对单元中捕获多个细胞,导致编码微球同时捕获了两个物种的mRNA信息。在上述2 952个高质量的细胞编码中,1 322个属于人细胞,1 344个属于小鼠细胞,239个混合细胞,两物种来源的细胞比例约为1:1,细胞混合率约为5.49%。

为了验证该平台在样本标签拆分上的有效性,我们应用前文建立的基于PDF拟合的阈值优化策略,对每个细胞标签的HC-bar与MC-bar信号进行定量判定。分析结果显示,鼠源细胞在HC-bar维度仅呈现极低的背景信号(平均UMI计数小于100),而在MC-bar维度则表现出显著的高丰度信号(平均UMI计数大于700);人源细胞则呈现完全相反的信号模式,其MC-bar信号普遍维持在低水平(小于100),而HC-bar信号显著升高(2 200);表明两套胆固醇标签的锚定具有高度特异性,即对同源标签捕获高效而对异源标签几乎无结合。在混合双胞中,这两种样本标签均表现出高UMI信号,人源标签UMI均值可达2 200以上,鼠源标签UMI均值可达700以上,这与双胞同时携带人、鼠两套标签的物理特征完全吻合(补充图2D-E)。这种双胞在标签信号上的共现现象,证实了该标记体系能够凭借标签信号实现样本的精准拆分与双胞的有效识别。

2.7 细胞聚类与双细胞鉴定

在完成多样品拆分可行性验证后,本文进一步通过转录组层面的物种注释与聚类分析,以评估Well-Multi-seq样本拆分策略的可靠性及双胞鉴定能力。首先对人鼠混合质控验证样本的基因表达矩阵进行UMAP非线性降维,将高维转录组数据投影至二维空间。如图4A所示,将转录组物种注释结果

映射至 UMAP 空间后, 人源(Human)与鼠源(Mouse)细胞各自聚集成边界清晰的独立细胞簇, 少量疑似混合双胞(Mix)分布于两簇边界区域。在此基础上, 将基于样本标签(HC-bar/MC-bar)的结果映射至同一 UMAP 分布中进行直观展示(图 4B)。依据 HC-bar 与 MC-bar 的 UMI 信号强度与相对比例, 可将细胞明确划分为人源、鼠源、混合双胞及阴性细胞(Negative)四类。

为定量评估两种标注体系的一致性, 构建了分类占比堆叠柱状图(图 4C)与混淆矩阵(图 4D)。图 4C 清晰地展示了在样本标签与 RNA 两种标注方式下的细胞构成比例高度重叠: 人源与鼠源细胞均占据绝对主导, 混合双胞与阴性细胞占比极低, 证明了拆分结果的准确性。图 4D 的结果显示, 在样本标签判定的人源细胞中, 92.60% 在转录组层面同样被注释为人源; 相似地, 鼠源细胞的匹配度准确率高达 95.28%; 少量交叉仅源于混合双细胞群体, 这与混合双细胞同时携带双重物种信号的物理特性完全吻合。

整体而言, 样本标签拆分结果与转录组物种注释表现出极高的一致性, 充分验证了 Well-Multi-seq 标签混样拆分技术的可靠性与精准性。

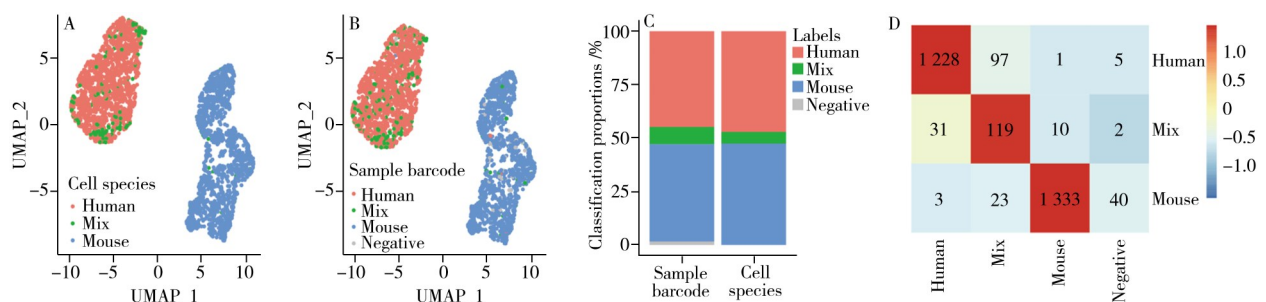


图4 样本细胞降维聚类与物种标签一致性验证

Fig. 4 Dimensionality reduction clustering of sample cells and validation of species label consistency

A: UMAP dimensionality reduction plot of cells, colored by species annotation at the transcriptome level (Human, Mouse and Mix) (细胞 UMAP 降维图, 按转录组水平的物种注释 (Human、Mouse 和 Mix) 着色); B: UMAP dimensionality reduction plot of cells, colored by species labels obtained from sample barcode demultiplexing (Human, Mouse, Mix and Negative) (细胞 UMAP 降维图, 按样本标签拆分得到的物种标签 (Human、Mouse、Mix 和 Negative) 着色); C: stacked bar plot of classification proportions (分类占比堆叠柱状图); D: confusion matrix between sample barcode demultiplexing and RNA species annotation (样本标签拆分与 RNA 物种注释的混淆矩阵)

2.8 成本效益分析与综合性能评估

在验证 Well-Multi-seq 的细胞捕获性能与样品拆分准确性的基础上, 本文进一步对其成本效益与综合性能进行系统评估。对比常规单样本流程与混样策略的试剂成本可见, 常规流程单次实验仅处理 1 个样本。混样策略在试剂与实验流程保持不变的基础上, 仅新增低成本样本标签, 即可将固定试剂成本在多个样本间均摊。单次实验混样数为 2 个时, 较常规流程成本降低 47.20%; 当混样数提升至 8 个时, 成本降幅达 84.70%, 显著降低了单细胞测序样本处理成本(补充表 2)。综合对比 Drop-seq、10x Genomics、SMART-seq2 及 MULTI-seq 等主流单细胞转录组技术的多项核心指标(补充表 3), 结果表明 Well-Multi-seq 具有优异的综合性能与应用潜力。尽管上述结果表明 Well-Multi-seq 在细胞捕获效率、样品拆分准确性及成本控制等方面均展现出显著优势, 但其在复杂样本体系中的适用性仍需进一步验证。不同类型细胞在膜组成及物理性质上的差异, 可能影响胆固醇修饰标签的插入效率, 从而影响标记均一性。此外, 该方法在低丰度转录本检测及整体灵敏度方面仍有优化空间。未来工作可围绕提升标记效率、增强检测灵敏度以及拓展至原代细胞和临床样本等方向展开。

3 结论

本研究成功开发了 Well-Multi-seq 多样本混合单细胞转录组测序新方法, 通过创新性整合双孔嵌套式微流控芯片与胆固醇修饰寡核苷酸标签标记体系, 系统性解决了传统单细胞转录组测序技术中细胞损失率高、双细胞污染严重及多样本分析成本高昂等核心瓶颈。在细胞捕获方面, Well-Multi-seq 芯片结合尺寸排阻与准静态流体动力学原理, 突破泊松分布限制, 实现了高达 94% 的微球占有率、79% 的细胞占有率以及 87% 的细胞捕获率。在多样本标记与拆分方面, 双胆固醇修饰标签可稳定锚定细胞膜, 实现多样本一管混合测序; 依托标签共现模式结合 PDF 拟合阈值策略, 实现精准双胞识别,

有效规避了纯计算算法的假阳性风险。人鼠混合样本实验验证表明, 样本标记准确率分别为92.60%(人源)与95.28%(鼠源), 剔除阴性细胞后样本拆分总准确率达95.48%。在成本效益方面, 混样数提升至8个时试剂成本较常规流程降低84.70%, 综合性能评估也表明 Well-Multi-seq 在多项核心指标上优于或持平于 Drop-seq、10x Genomics、SMART-seq2 及 MULTI-seq 等主流技术。整体而言, Well-Multi-seq 兼具高细胞利用率、低污染、低成本与高标记特异性等多重优势, 为循环肿瘤细胞、微针活检样本等稀有细胞类型的研究, 以及药物筛选、胚胎发育等动态生物学过程的大规模平行分析提供了切实可行的技术方案。未来, 随着标记效率的进一步优化、检测灵敏度的持续提升以及对原代细胞和临床样本适用性的系统验证, Well-Multi-seq 有望在基础生物学探索与临床精准诊断领域发挥更广泛的应用价值。

参考文献:

- [1] Svensson V, Vento-Tormo R, Teichmann S A. *Nat. Protoc.*, **2018**, 13(4): 599–604.
- [2] Tang F, Barbacioru C, Wang Y. *Nat. Methods*, **2009**, 6(5): 377–382.
- [3] Doufène K, Tourné-Péteilh C, Etienne P, Aubert-Pouessel A. *Langmuir*, **2019**, 35(39): 12597–12612.
- [4] Qu H Q, Kao C, Hakonarson H. *Stem Cells*, **2024**, 42(1): 1–12.
- [5] Zhou W M, Yan Y Y, Guo Q R. *J. Nanobiotechnol.*, **2021**, 19(1): 312.
- [6] Zhang Y, Xu S, Wen Z. *Cell. Mol. Life Sci.*, **2022**, 79(10): 500.
- [7] Macosko E Z, Basu A, Satija R. *Cell*, **2015**, 161(5): 1202–1214.
- [8] Ziegenhain C, Vieth B, Parekh S. *Mol. Cell*, **2017**, 65(4): 631–643.
- [9] Haghverdi L, Lun A T L, Morgan M D, Marioni J C. *Nat. Biotechnol.*, **2018**, 36(5): 421–427.
- [10] Fang L, Li G, Sun Z. *Mol. Syst. Biol.*, **2021**, 17(4): e10060.
- [11] Stoeckius M, Zheng S W, Houck-Loomis B, Hao S, Yeung B Z, Mauck III W M, Smibert P, Satija R. *Genome Biol.*, **2018**, 19: 224.
- [12] Cusanovich D A, Daza R, Adey A. *Science*, **2015**, 348(6237): 910–914.
- [13] Stoeckius M, Hafemeister C, Stephenson W. *Nat. Methods*, **2017**, 14(9): 865–868.
- [14] Mylka V, Matetovici I, Poovathingal S, Aerts J, Vandamme N, Seurinck R, Verstaen K, Hulselmans G, Van den Hoecke S, Scheyltjens L, Movahedi K, Wils M, Reumers J, Van Houdt J, Aerts S, Saeys Y. *Genome Biol.*, **2022**, 23: 55.
- [15] Kodr D, Kužmová E, Pohl R, Kraus T, Hocek M. *Chem. Sci.*, **2023**, 14(15): 4059–4069.
- [16] Hu C, Liu S, Huang G. *ChemBioChem*, **2024**, 25(24): e202400748.
- [17] Feng Y J, Chen D L, Aplegate C, Gonzalez Medina N, Kuo C W, Arogundade O H, Wrigh C L, Xu F X, Drnevich J, Smith A M. *ACS Nano*, **2025**, 19(24): 22079–22092.
- [18] Wolock S L, Lopez R, Klein A M. *Cell Syst.*, **2019**, 8(4): 281–291.
- [19] McGinnis C S, Murrow L M, Gartner Z J. *Cell Syst.*, **2019**, 8(4): 329–337.
- [20] Xi N M, Li J J. *Cell Syst.*, **2021**, 12(2): 176–194.
- [21] Zhu Q, Conrad D N, Gartner Z J. *Genome Biol.*, **2024**, 25: 37.
- [22] Oleiwi Z C, Al-Mayahi N F, Jasim Sabib D, Kamil Hadi N. *J. Phys.: Conf. Ser.*, **2019**, 1294(3): 032028.

(责任编辑: 龙秀芬)