



10x 单细胞建库 & 胶珠分选

核心原理详解 · 面向零基础读者 | 2026年7月

💡 **阅读前提：** 本文不需要任何分子生物学背景。你需要一个直觉类比——**快递分拣中心**。10x 做的事情，本质上就是给全城每个人一封信，贴上独一无二的房间号，然后混在一起也能分清谁是谁。

一、核心类比：快递分拣中心

你要给全城 **10000 个人** 发问卷，但有三个变态要求：

- ① 每个人的问卷必须有**独一无二的编号**
- ② 你不知道谁是谁——问卷回收后要能分清"这一份来自 A 先生还是 B 女士"
- ③ 所有人填完的问卷混在一个大箱子里，你仍然能知道**每个答案来自哪个人**

10x 的解决方案：把每个人关进独立小房间 → 在每个房间里放一张带房间号的贴纸 → 让他在房间里填完问卷（贴纸贴在问卷上） → 把所有问卷混在一起 → 靠房间号分清谁是谁。

二、三步走完整个流程

第一步：把每个细胞关进独立"小房间"（GEM 液滴）

细胞悬液
人山人海

+

胶珠悬液
带编号贴纸

+

油相
包裹液

→

微流控芯片
分流闸机

→

上万个 GEM 液滴
独立小房间

🔑 **GEM = Gel Bead-in-Emulsion**：液滴里包着一颗胶珠 + 一个细胞 + 反应液。理想情况每个液滴最多一个细胞、最多一颗胶珠。

怎么保证"刚好一个细胞"? 纯靠概率。细胞浓度稀释到极稀，大部分液滴是空的或单细胞，双胞胎率

控制在 <5%。空液滴在数据分析时自动排除。

第二步：每颗胶珠上都写了什么？

● 胶珠（Gel Bead）上密密麻麻挂着几十万条 DNA 短链，每条链有三个功能区：



组成部分	快递分拣比喻	实际作用
Cell Barcode	 房间号	同一颗胶珠上所有链共享同一个 Barcode（共 ~75 万种），液滴内所有 mRNA 都被贴上同一个"房间号"——测序后就知道这个基因是从哪个细胞来的
UMI	 快递单号	同一颗珠子上每条链的 UMI 不同（几百万种组合）。每条 mRNA 分子都获得独一无二的"单号"——能区分"真的高表达"和"PCR 扩增出来的假重复"
Poly(dT)	 磁铁	mRNA 尾巴是 Poly(A)（一串 A），Poly(dT)（一串 T）能像磁铁一样精准抓住它，把 mRNA 从裂解液里"钓"出来

第三步：液滴内发生的一切（"包裹里完成"）

 单个 GEM 液滴内部流程



破乳 → 所有人 cDNA 混在一管 → 建库 + 测序 → Barcode → 哪个细胞
打碎所有液滴 UMI → 真实表达量

三、胶珠分选：Barcode 是怎么做到每颗珠子不一样的？

核心技术：分割-混合法 (Split-Pool Synthesis)

想象你要给 **43 亿** 个人每人发一个不重号的 16 位数字身份证。你可以这么做：



数学结果： $4^{16} = 4,294,967,296$ 种可能组合。实际 10x 基因组学使用 ~75 万种 Barcode——取 43 亿种可能性的一小部分，足够覆盖每个样本中的细胞数（通常几千到几万个）。

胶珠悬浮在反应液中，每轮经过一个含特定碱基的溶液池，珠子表面会加长一个碱基。全部珠子混匀后再随机分入下一轮——就像洗扑克牌一样，每轮洗一次，16 轮后顺序完全不同。

两颗粒子 16 轮走完全一样的路径的概率 = $(1/4)^{16} \approx 2.3 \times 10^{-10}$ ——也就是 43 亿分之一。在实际细胞数（几千到几万）中，撞号概率可以忽略不计。

四、为什么需要 UMI? (分子防伪码的核心价值)

✗ 没有 UMI 时

一条 mRNA 被 PCR 扩增了 1000 倍 → 测序读出 1000 条相同的序列
→ 你以为这个基因表达了 1000 次
→ 其实它只表达了 1 次

✓ 有 UMI 时

每条 mRNA 在反转录时被贴上不同的 UMI
→ PCR 扩增 1000 倍后，所有拷贝共享同一个 UMI
→ 同一个 UMI 的所有 reads 只算作 1 个分子
→ 真实表达量被还原

简单记：Cell Barcode 回答"这个基因来自哪个细胞"；UMI 回答"这个基因在这个细胞里真的表达了几个分子"。

五、常见新手问题

? 10x 和传统 RNA-seq 有什么区别?

传统 RNA-seq 把一堆细胞的 mRNA 混在一起测，只能得到"平均表达量"——你只知道这群细胞平均表达了某个基因，但你不知道是"每个细胞都中等表达"还是"少数细胞极高表达、其余不表达"。10x 单细胞测序告诉你：这一个细胞表达了基因 A 100 次，旁边那个细胞表达了 0 次。这就是"单细胞分辨率"。

? 胶珠分选和流式分选 (FACS) 有什么不同?

FACS 是用激光识别细胞表面标志物，然后一个一个挑出来——你需要事先知道要找什么标志物。10x 不挑细胞——它会无差别地把所有细胞包进液滴，然后通过 RNA 表达谱"反向发现"细胞类型。FACS 是"我知道我要谁的 CD4+ T 细胞"；10x 是"我不知道这群细胞里有什么，让我把所有细胞都测一遍，再分类"。

? 空液滴和双胞胎怎么办?

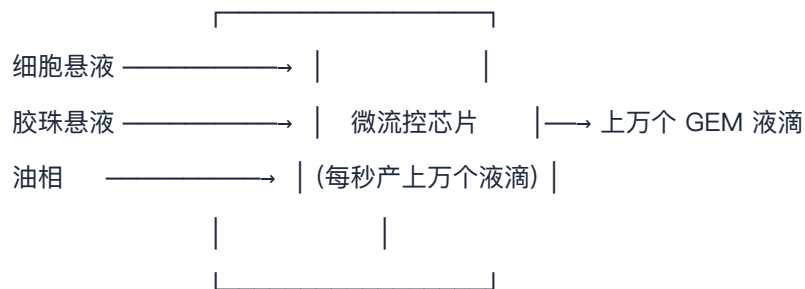
空液滴：里面没有细胞，自然没有 mRNA。测序后这些液滴的 UMI 计数极低或为零，数据分析时自动过滤。

双胞胎：两个细胞进了同一个液滴，会共用同一个 Barcode。如果两个细胞类型不同，它们的混合表达谱像一个“四不像”——分析软件能识别为双胞胎并剔除。如果两个同类型细胞形成双胞胎，确实难以检测，但概率很低。

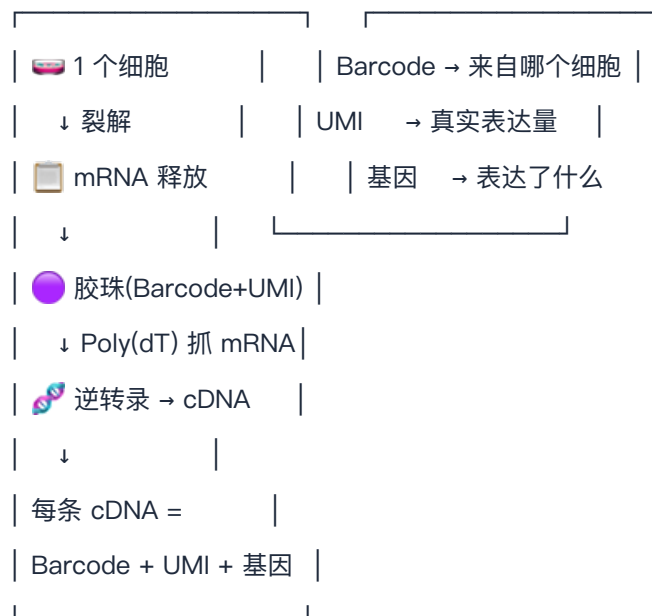
? 10x 能做多少个细胞?

最新 Chromium X 系列可以一次处理 **500–20,000 个细胞**，高通量版本可以到 **100,000+ 个细胞**。对于单样本来说，1000–10000 个细胞是最常见的范围。一次实验的成本大约 ¥5000–20000 / 样本（不包括测序）。

六、一张图总结全流程



单个 GEM 液滴内部： 测序后分析：



 **核心思想一句话：**把每个细胞关进独立小房间（GEM 液滴）→ 在每个房间里贴一张带房间号的标签（胶珠 Barcode）+ 每份问卷贴一张快递单号（UMI）→ 拆掉所有墙把问卷混在一起（破乳）→ 靠房间号分清每个人（生信分析）。

10x 单细胞建库核心原理解析 | 面向零基础读者

关键参考文献：10x Genomics Chromium Single Cell Gene Expression官方技术文档 | Zheng et al.,
Nature Communications 2017