

DOI: 10.16781/j.CN31-2187/R.20230240

• 专题报道 •

## 单细胞测序在干细胞生物学中的应用研究进展

肖邦<sup>1△</sup>, 朱静<sup>2△</sup>, 徐莎<sup>3</sup>, 陈国栋<sup>3</sup>, 黄超<sup>4\*</sup>

1. 海军军医大学(第二军医大学)基础医学院医学遗传学教研室, 上海 200433
2. 海军军医大学(第二军医大学)中医系中医妇科学教研室, 上海 200433
3. 海军军医大学(第二军医大学)转化医学研究中心干细胞与再生医学研究室, 上海 200433
4. 海军军医大学(第二军医大学)基础医学院人体解剖学教研室, 上海 200433

**[摘要]** 细胞间的变异或异质性是干细胞群的基本和内在特征。传统的组学分析在混合细胞群体上进行, 导致这些差异被掩盖。近年来, 单细胞基因组、表观基因组、转录组测序等技术发展迅速, 为全面解析细胞异质性和识别不同表型细胞类型提供了强有力的工具。这些方法应用于不同类型的干细胞, 包括多能干细胞和组织特异性干细胞, 取得了令人兴奋的新发现。本文综述了单细胞组学测序技术的最新进展, 并对单细胞组学测序的方法和应用前景进行了展望。

**[关键词]** 单细胞组学; RNA 测序; 表观基因组测序; 干细胞; 细胞异质性

**[引用本文]** 肖邦, 朱静, 徐莎, 等. 单细胞测序在干细胞生物学中的应用研究进展[J]. 海军军医大学学报, 2023, 44(7): 792-799. DOI: 10.16781/j.CN31-2187/R.20230240.

### Application of single cell sequencing in stem cell biology: research progress

XIAO Bang<sup>1△</sup>, ZHU Jing<sup>2△</sup>, XU Sha<sup>3</sup>, CHEN Guodong<sup>3</sup>, HUANG Chao<sup>4\*</sup>

1. Department of Medical Genetics, College of Basic Medical Sciences, Naval Medical University (Second Military Medical University), Shanghai 200433, China
2. Department of Gynecology of Traditional Chinese Medicine, School of Traditional Chinese Medicine, Naval Medical University (Second Military Medical University), Shanghai 200433, China
3. Laboratory of Stem Cell and Regenerative Medicine, Center of Translational Medicine, Naval Medical University (Second Military Medical University), Shanghai 200433, China
4. Department of Anatomy, College of Basic Medical Sciences, Naval Medical University (Second Military Medical University), Shanghai 200433, China

**[Abstract]** Cell-to-cell variation and heterogeneity are fundamental and intrinsic characteristics of stem cell populations, but these differences are masked when bulk cells are used for traditional omic analysis. Technologies such as single-cell genome, epigenome and transcriptome sequencing have been developing rapidly in recent years. Single cell sequencing technologies serve as powerful tools to comprehensively dissect cellular heterogeneity and to identify distinct phenotypic cell types. The application of these methods to different types of stem cells, including pluripotent stem cells and tissue-specific stem cells, has led to exciting new findings in the stem cell field. This paper reviews the recent progress as well as future perspectives in the methodology and application of single-cell omic sequencing technologies.

**[Key words]** single-cell omics; RNA sequencing; epigenome sequencing; stem cells; cellular heterogeneity

**[Citation]** XIAO B, ZHU J, XU S, et al. Application of single cell sequencing in stem cell biology: research progress [J]. Acad J Naval Med Univ, 2023, 44(7): 792-799. DOI: 10.16781/j.CN31-2187/R.20230240.

[收稿日期] 2021-02-22 [接受日期] 2022-09-02

[基金项目] 国家自然科学基金(82201789), 上海市自然科学基金(22ZR1477000)。Supported by National Natural Science Foundation of China (82201789) and Natural Science Foundation of Shanghai (22ZR1477000)。

[作者简介] 肖邦, 博士, 讲师. E-mail: 15821804254@163.com; 朱静, 硕士, 主治医师. E-mail: 88561632@qq.com

△共同第一作者(Co-first authors)

\*通信作者(Corresponding author). Tel: 021-81870952, E-mail: 18583626995@163.com

细胞是生物体中最小的功能单位,基因表达的调节是在单个细胞内或细胞间进行的。因此,在理想情况下分析基因表达应在单细胞中进行。然而,由于单个细胞很小以及技术的局限性,前期几乎所有文献中描述的基因表达研究(特别是全基因组水平的研究)都是使用数千个甚至数百万个细胞的大量样本进行的。这样基于集成分析的数据是有效的,然而细胞之间的基因表达异质性却未得到很好的解析。

单个细胞处在影响基因表达的一系列生理或病理条件形成的微环境中,所以细胞异质性是组织的一个普遍特征。另外,在整个细胞周期中基因表达可能不同并存在内在的随机性,即使是“纯”细胞类型也存在基因表达的异质性<sup>[1]</sup>。干细胞既具有无限的自我更新能力,又具有分化成特殊类型细胞的潜力,其一般分化为多能干细胞和组织特异性干细胞,前者可以产生所有3个胚层(外胚层、中胚层和内胚层)的细胞,后者在胚胎组织的发育和成体组织的稳态中起着至关重要的作用。哺乳动物早期胚胎中的多能干细胞数量很少,组织特异性干细胞通常也只占特定组织或器官细胞群的一小部分。这些小细胞群与胚胎或成人组织中的各种分化和中间细胞类型混合在一起形成异质群体。单细胞RNA测序(RNA sequencing, RNA-seq)技术为描述异质细胞群(包括干细胞)的组学尺度特征提供了强有力的工具。

本文讨论了最近开发的单细胞组学测序方法(包括单细胞转录组、表观基因组测序技术),重点介绍了它们在干细胞(包括多能干细胞和组织特异性干细胞)中的应用,最后展望了单细胞RNA-seq技术在干细胞领域的研究方向和应用前景。

## 1 单细胞RNA-seq技术

RNA-seq技术可以在单碱基分辨率下提供一个无偏倚的细胞转录组概貌。研究表明,干细胞转录组可以准确地反映其多能性或分化状态<sup>[2]</sup>。在单细胞分辨率下探索干细胞的自我更新和分化的多样性及动态变化具有重要意义。第1种单细胞RNA-seq方法于2009年被报道<sup>[3]</sup>。随后,许多基于不同细胞捕获、cDNA扩增和文库建立策略的单细胞RNA-seq方法也相继被建立起来,包括多次退火环状循环扩增、基于sem随机引物PCR的mRNA扩

增、体内转录组分析、微流控单细胞RNA-seq和大规模并行单细胞RNA-seq<sup>[4-8]</sup>。最近,科学家还开发了原位单细胞RNA-seq<sup>[9]</sup>、高度多重分析<sup>[10]</sup>、单细胞分辨率三维重建<sup>[11]</sup>等方法。然而,上述方法都存在一个共同的缺点:只检测来自单个细胞的poly(A)<sup>+</sup>RNA,而忽略了重要的poly(A)<sup>-</sup>RNA。为了克服上述缺点,Martin等<sup>[12]</sup>开发了单细胞通用的、poly(A)-独立的RNA-seq技术,实现从单个细胞中同时检测poly(A)<sup>+</sup>RNA和poly(A)<sup>-</sup>RNA,并且在小鼠植入前胚胎中发现了数千个环状poly(A)<sup>-</sup>RNA和数百个线性poly(A)<sup>-</sup>RNA。

为了全面了解复杂细胞群的异质性,需要对大量的单个细胞进行测序,这对RNA-seq技术的通量提出了更高的要求。近年来,随着测序技术的发展和进步,单细胞RNA-seq的通量有了很大的提高。基于微流体和机器人系统的高通量策略可以处理数百个单个细胞<sup>[13]</sup>。液滴单细胞RNA测序(droplets-based single-cell RNA sequencing, Drop-seq)和索引液滴RNA测序(indexing droplets RNA sequencing, inDrop-seq)通过联合一珠一胞液滴和独特的条形码将每次测序的通量提高到数千甚至数万个单个细胞<sup>[14]</sup>。另外,对于测序的深度,一般情况下每个细胞只需测序5万个读长就可以区分出不同的细胞类型。然而,如果要区分具有相对细微差异的细胞类型,则需要更高的测序深度。

此外,在测序数据分析环节,应用于混合细胞群RNA-seq分析的大部分生物信息学方法同样适用于单细胞RNA-seq数据。特别地,在干细胞研究中可以通过无偏聚类和差异基因表达分析的方法从数据集中识别不同的细胞类型和亚群以及它们的标记基因。Qian等<sup>[15]</sup>开发了一种基于双聚类的算法BackSPIN,提高了从单细胞RNA-seq数据识别不同细胞类型的准确性。Monocle和Waterfall算法通过计算最小生成树,降维数据空间产生“拟时”轨迹,以重建干细胞发育或分化过程的时间序列<sup>[16]</sup>。

目前,单细胞RNA-seq方法已被应用于多种系统,包括早期哺乳动物胚胎<sup>[17]</sup>、发育中的组织<sup>[18]</sup>、成人组织<sup>[19]</sup>、免疫细胞<sup>[20]</sup>、癌细胞<sup>[21]</sup>以及体内分离或体外培养的干细胞<sup>[22]</sup>。Kozareva等<sup>[23]</sup>对从小鼠初级体感觉皮质(S1)和海马CA1区分离的3005个单细胞的转录组进行了无偏性测序,鉴定出47种分子不同的细胞亚类,包括9种主要细胞类

型: S1 和 CA1 锥体神经元、中间神经元、少突胶质细胞、星形胶质细胞、小胶质细胞、血管内皮细胞、壁细胞和室管膜细胞。因此,单细胞 RNA-seq 技术已经成为一个成熟的、强大的工具,将对干细胞生物学的研究产生强有力的推动作用。

## 2 单细胞 RNA-seq 技术在干细胞生物学研究中的应用

2.1 植入前发育 哺乳动物胚胎着床前发育代表着新生命的开始,整个过程中的基因表达发生着变化。由于在这一发育过程中的细胞数量非常有限,单细胞 RNA-seq 技术提供了一个前所未有的机会来破译这一过程中的基因表达动态。目前,科学家已经获得了来自着床前发育的人和小鼠细胞的全面转录组谱,准确捕捉了母体-合子转化的基因表达特征<sup>[24]</sup>。虽然同一阶段的细胞相对相似,但有证据表明早在小鼠胚胎的四细胞阶段卵裂球间的差异就已经出现。这些差异可能在功能上与着床前胚胎的第一个细胞命运决定事件有关,即营养外胚层(trophectoderm, TE)和内细胞群(inner cell mass, ICM)之间的分离,后者进一步分离为原始内胚层(primitive endoderm, PE)和多能外胚层(pluripotent epiblast, EPI),产生胚胎本身的所有细胞谱系<sup>[25]</sup>。研究表明,DNA 结合抑制物 2(inhibitor of DNA binding 2, ID2)和性别决定区(sex determining region, SOX) 2 分别是 TE 和 ICM 细胞中的早期标志物<sup>[26]</sup>。此外,在 PE 和 EPI 分离之前前体细胞同时表达 PE 和 EPI 标志物,即随机发生细胞间基因异质性表达,然后是这种基因异质性表达信号的强化,从而决定了不同的细胞命运<sup>[27]</sup>。

基于单细胞 RNA-seq 数据,人们已经明确人类和小鼠之间植入前发育基因表达的异同。人类和小鼠发育的一个巨大差异是合子基因组激活的时间不同。在混合遗传背景小鼠(CAST/EiJ× C57BL/6J)中,胚胎在双细胞阶段表现出快速的母体转录物清除和合子基因组激活以及显著的等位基因特异性表达<sup>[28]</sup>。在人类中,主成分分析和差异表达分析证实合子基因组激活发生在四细胞和八细胞阶段之间<sup>[29]</sup>。人类和小鼠在发育方面还存在许多其他重要差异,例如,转录因子 KLF17 仅在人类 EPI 中表达; TGFβ 信号通路的关键成分在人类胚胎中高度富集,而在小鼠胚胎中则没有<sup>[30]</sup>。此外,关键因

子 ID2、E74 样因子 5 和脱中胚蛋白同源物仅在小鼠 TE 细胞中表达,而在人类中不表达<sup>[30]</sup>。

### 2.2 多能干细胞

2.2.1 胚胎干细胞(embryonic stem cell, ESC) 小鼠和人 ESC 是研究多能干细胞自我更新能力和分化潜力的理想体外模型。当在适当的多能性维持条件下培养时,囊胚的 ICM 可以形成 ESC,并且已经使用单细胞 RNA-seq 方法追踪了人类和小鼠 ESC 的来源<sup>[31]</sup>。人类 ESC 和 EPI 的比较表明,参与多能性的基因是保守的,但在不同的途径上富集<sup>[32]</sup>。人 EPI 在氧化磷酸化信号通路中富集且优先转向糖酵解代谢,这反映了 EPI 与 ESC 不同的生长环境(ESC 在体外富氧条件培养,而 EPI 的体内生存环境氧浓度相对较低)。ESC 基因在调节细胞增殖、MAPK 和 Wnt 信号通路中富集,这表明 EPI 和 ESC 在维持多能状态方面存在不同的机制。

虽然 ESC 是相对同质的,但包含不同的亚种群。单细胞 RNA-seq 分析显示,许多基因在单个小鼠胚胎干细胞(mouse embryonic stem cell, mESC)中具有不同的表达,而且已经确定了具有不同转录组的亚群<sup>[33]</sup>。García-Castro 等<sup>[34]</sup>通过使用液滴条形码方法对近 1 000 个 mESC 进行测序,描述了几个较小的亚群,包括一个外母细胞样亚群、一个高 PR 结构域蛋白 1 表达亚群和一个高热休克蛋白 90 表达亚群。该研究还对数千个细胞进行了测序,以检测白血病抑制因子停用后 mESC 的分化,并描述了几个亚群在分化过程中的动态变化,这些亚群与任何已知的细胞类型都不同。

2.2.2 诱导性多能干细胞(induced pluripotent stem cell, iPSC) 多能干细胞包括 ESC 和 iPSC。人类 iPSC 是通过将分化的体细胞进行重编程而产生的,避免了使用人类胚胎引起的伦理问题,具有独特的优势。另一方面,人类 iPSC 的分化潜能表现出异质性,是阻碍其临床应用(细胞治疗)的一大因素<sup>[35]</sup>。例如,人类 iPSC 用于软骨再生的治疗在很大程度上受到软骨细胞产量低及软骨形成过程中细胞不可预测和异质脱靶分化的阻碍。单细胞 RNA-seq 技术的应用有可能为其提供更优的解决方案。Wu 等<sup>[36]</sup>结合了混合细胞 RNA-seq、单细胞 RNA-seq 和生物信息学分析(包括加权基因共表达分析),研究在诱导形成软骨条件下调节人类 iPSC 分化的基因调控网络。他们以特异性 Wnt 和

小眼畸形相關轉錄因子 (microphthalmia-associated transcription factor, *MITF*) 作為關鍵基因, 在軟骨形成過程中調控人類 iPSC 脫靶分化為神經細胞和黑色素細胞。通過靶向 *Wnt* 和 *MITF* 可以消除這些細胞系的形成, 顯著提高人類 iPSC 衍生軟骨細胞的產量和均一性。Kamatani 等<sup>[37]</sup> 基於單細胞 RNA-seq 發現, 人類 iPSC 衍生軟骨組織植入髓核丟失的大鼠模型後分化為軟骨細胞樣髓核細胞而非脊索髓核細胞, 從而實現了髓核的功能性再生。人類 iPSC 衍生的人類腸道類器官 (human intestinal organoid, HIO) 缺乏一些在天然器官中存在的細胞群 (如脈管系統), 從而極大地影響了其臨床應用。Holloway 等<sup>[38]</sup> 首先應用單細胞 RNA-seq 發現, 分化早期的 HIO 中存在的內皮細胞群體在標準培養條件下不能隨著培養時間的延長而很好地維持; 通過優化方案使 HIO 培養條件始終處於起始培養的狀態, 可顯著提高 HIO 中內皮細胞的比例, 這為推動 HIOs 的臨床應用打下堅實的基础。

**2.3 組織特异性干细胞** 組織特异性干细胞存在于发育或分化的组织中, 它们也进行自我更新, 并有潜力分化成各种特定的细胞类型。单细胞 RNA-seq 方法已经应用于组织特异性干细胞的研究, 这些研究确定了新的干细胞类型, 并在“同质”干细胞群中解析了细胞异质性。

**2.3.1 造血干细胞 (hematopoietic stem cell, HSC)** HSC 可以分化成所有的血细胞谱系, 分为长期 HSC 和短期 HSC。长期 HSC 位于造血系统的顶层, 可以进行自我更新和分裂以补充短期 HSC。细胞周期差异主导着每种 HSC 类型的细胞异质性。利用单细胞转录组数据可以重建 HSC 的细胞周期进程, 为研究静止和增殖干细胞的特性提供了一种新方法。对非循环细胞的分析显示, 长期 HSC 和短期 HSC 之间有明显差异。通过分析造血基因发现, 在长期 HSC 中与特定谱系标记相关的细胞亚群也存在异质性<sup>[39]</sup>。Kowalczyk 等<sup>[40]</sup> 比较了年轻和老年小鼠单细胞 RNA-seq 数据发现, 衰老与长期 HSC G1 期长度的减少有关, 这可能与老年小鼠中长期 HSC 的积累有关; 老化的 HSC 的转录组状态与其分化状态呈负相关。

**2.3.2 神经干细胞 (neural stem cell, NSC)** 在成年哺乳动物大脑中, 齿状回脑室下区和粒下区 NSC 不断生成新的神经元和胶质细胞。神经发生过程

是从静止的 NSC 开始, 成为激活的 NSC, 随后是早期中间祖细胞。利用拟时序分析单细胞转录组数据, Zhou 等<sup>[41]</sup> 绘制了早期神经发生过程的持续发展轨迹, 证明如果在给定时间点对种群中合理数量的单个细胞进行测序, 单细胞 RNA-seq 可以提供发育过程转录组动态的快照。同时, 科学家们也对非生理条件对干细胞群的扰动进行了研究。Bottes 等<sup>[42]</sup> 分析缺血性脑损伤中的 NSC 发现, 在生理条件下 NSC 经历从休眠 NSC 到启动静止的 NSC 再到激活的 NSC 的转变; 在损伤的 NSC 中, 休眠 NSC 的比例降低, 而启动静止的 NSC 和激活的 NSC 的比例大大增加。

**2.3.3 肠干细胞** 持续自我更新的肠上皮细胞是另一个成熟的研究成体干细胞的模型, 位于隐窝底部的富含亮氨酸重复序列 G 蛋白偶联受体 5 (leucine rich repeat containing G protein coupled receptor 5, *lgr5*) 阳性细胞被认为是干细胞, 参与肠上皮自我更新过程。Gehart 等<sup>[43]</sup> 对近 200 个 GFP 标记的 *lgr5* 阳性细胞进行测序发现, 这些细胞形成了一个单一的大同质群体, 表明其群体结构明显不同于 HSC 和 NSC。这些研究表明, 单细胞 RNA-seq 可以提供关于干细胞群结构及其在不同条件下行为的丰富信息, 为组织特异性干细胞的功能提供重要的见解。

**2.4 原始生殖细胞 (primordial germ cell, PGC)** PGC 是成熟生殖细胞 (卵细胞和精子) 的前体。科学家已经建立了人类 PGC 从迁移阶段到性腺阶段的单细胞 RNA-seq 数据集, 揭示了 PGC 发育过程中多能性基因和种系特异性基因的动态平衡表达<sup>[44]</sup>。早期 PGC 的细胞群在有丝分裂中是相对同质的, 而后期女性 PGC 在减数分裂阻滞中是高度异质的, 这一发现表明女性 PGC 进入减数分裂阻滞是不同步的。研究人员还系统地探索了区分人类 PGC 与小鼠 PGC 的独特特征, 人类早期 PGC 高表达 *SOX15* 和 *SOX17*, 而小鼠早期 PGC 则表达 *SOX2*。

**2.5 肿瘤干细胞 (cancer stem cell, CSC)** 癌症组织通常包含具有表型和功能异质性的细胞亚群。CSC 理论认为在肿瘤细胞分级的顶端有一个高度恶性的干细胞亚群。然而, 在许多癌症类型中这些 CSC 的存在仍然存在争议。每种肿瘤均在许多方面表现出高度的肿瘤细胞异质性, 单细胞 RNA-seq 有可能帮助识别这些细胞, 为复杂的肿瘤内异质性提供新的见解。Nefel 等<sup>[45]</sup> 对 5 个胶质母细胞瘤

样本中的 672 个单细胞进行了测序,通过检测一组“干性”基因在 5 个肿瘤样本的单个细胞表达,揭示了其连续而非离散的干性相关表达状态,反映了原发肿瘤内的复杂性。目前只有少数研究在单细胞分辨率水平上揭示了肿瘤转录组的异质性,将来随着单细胞测序技术的不断发展,人们对各种癌症类型异质性(包括 CSC 的特征)的了解将更全面和准确。

### 3 单细胞表观基因组测序技术

干细胞的发育、维持和分化是由其基因组的表观遗传修饰(包括 DNA 和组蛋白的共价修饰)所调节的。细胞间表观遗传变异是基因表达转录调控细胞异质性的一个重要层面。表观基因组异质性是多能干细胞和成体干细胞等转录组异质性的基础。这些异质性如何与单个细胞中染色体构象的变化相联系尚不清楚。

**3.1 DNA 修饰** DNA 甲基化是哺乳动物基因组中的主要修饰方式,其在许多生理生化过程中起着重要作用。科学家使用单细胞 DNA 甲基化组测序方法绘制了人类和小鼠胚胎植入前发育的 DNA 甲基化图景<sup>[46]</sup>。有研究在基因组尺度和单碱基分辨率水平全面描述了发生在哺乳动物胚胎植入前和 PGC 发育过程中的 2 种总 DNA 去甲基化波<sup>[47]</sup>。研究表明,人 PGC 在妊娠后约 10~11 周的甲基化水平(6%~8%)低于其他类型的细胞。这一系列人 PGC 的低甲基化 DNA 甲基化组数据集可以作为评估从人 ESC 或人 iPSC 分化的 PGC 样细胞质量的参考标准。Smallwood 等<sup>[48]</sup>证明,仅整合 12 个卵母细胞单细胞亚硫酸盐测序数据集就可以在很大程度上重现整个 DNA 甲基化组的主要模式。除了 DNA 甲基化外,科学家最近还发现了基因组 DNA 上的 5-羟甲基胞密啶、5-甲酰胞嘧啶和 5-羧化胞嘧啶修饰<sup>[49]</sup>。目前在混合细胞全基因组尺度上检测这些 DNA 修饰的方法已经建立,但在单细胞水平上进行检测的方法仍有待进一步开发。

**3.2 染色质可及性** 混合细胞染色质可及性评估的方法经过改进优化后可以应用于单细胞水平的染色质可及性检测。O'Connell 等<sup>[50]</sup>基于染色质转座酶可及性测序并依赖原核生物 Tn5-转座酶优先插入基因组中可及染色质转座酶区域的能力开发了单细胞水平上染色质可及性测序。Meuleman 等<sup>[51]</sup>使

用的方法基于更传统的 DNA 酶测序。后者似乎比前者能检测到单个细胞中更多的染色质开放区域。此外,染色体结构捕获技术最近已被应用于单个细胞分析<sup>[52]</sup>。这些方法能够在不同染色质状态上正确区分 ESC 和其他细胞类型。在不久的将来,分析干细胞群体染色质状态的异质性将成为可能。

**3.3 组蛋白修饰** 组蛋白修饰在干细胞基因表达调控中起着至关重要的作用。染色质免疫沉淀测序(chromatin immunoprecipitation sequencing, ChIP-seq)是一种在全基因组尺度上广泛使用的组蛋白修饰测定的方法。Mishra 等<sup>[53]</sup>最近通过 Drop-seq 和条形码策略将 ChIP-seq 应用于单细胞组蛋白修饰的分析,建立了 Drop-ChIP 技术。然而,由于没有同时严格使用非特异性 IgG 抗体作为阴性对照,其单细胞 ChIP-seq 数据集中潜在的非特异性噪声无法排除。Drop-ChIP 在每个细胞只能检测大约 1 000 个组蛋白第三亚基四号赖氨酸的三甲基化峰,对应的峰检测灵敏度约为 5%。尽管如此,该方法能够将小鼠 ESC 分为 3 个亚群,这些亚群在多能性相关转录因子(八聚体结合转录因子 4、SOX2 和同源盒转录因子)和分化相关转录因子(叉头框蛋白 A2)以及表观遗传抑制因子(多梳蛋白和神经元限制性沉默因子辅助抑制因子)结合的位点上具有不同的组蛋白第三亚基四号赖氨酸的二甲基化信号。第 1 个亚群细胞具有这些多能性特征基因的最高信号,第 2 个亚群细胞具有中间信号,第 3 个亚群细胞具有最低信号,而分化和表观遗传抑制因子特征基因的第三亚基四号赖氨酸的二甲基化信号则相反。因此,这些亚群可能具有与多能性和分化启动相关的不同染色质状态。这一发现从新的层面揭示了 ESC 表观基因组中的细胞异质性。

### 4 小结和展望

单细胞 RNA-seq 已被广泛应用于分析干细胞的异质性,但目前所有可用的单细胞组学测序技术都不太理想,存在显著的技术噪声和扩增误差。另外,与混合细胞群 RNA-seq 相比,单细胞 RNA-seq 技术的覆盖率相对较低。因此,单细胞 RNA-seq 技术还有很大的改进空间。

扩增误差是限制单细胞组学测序技术准确性的一个关键参数。单细胞组学测序技术是基于在深度测序之前对单个细胞中的核酸进行预扩增,导

致扩增后被分析的单细胞被“破坏”，因此这些结果无法在同一个细胞中得到验证。研究人员使用 Sanger 测序对来自同一个体细胞的扩增产物进行重新测序以确定突变位点。然而，该策略只能检测到二代测序的错误，而使单细胞扩增错误被隐藏。另一种策略是使用多个细胞来进行相互验证，然后只计算在 3 个或 3 个以上单个细胞中的单核苷酸变异（single nucleotide variation, SNV）<sup>[54]</sup>，缺点是仅仅适用于可以在体外培养克隆和扩增的细胞（这对于大多数类型的原代细胞来说非常困难），如果无法制备这样的原代细胞，这种方法将会导致单个细胞所特有的真正 SNV 被消除，严重限制单细胞组学测序技术的应用。将来更优的单细胞组学测序技术应该能对单个细胞的核酸原始拷贝数进行多次重复测量，可以在同一细胞内准确、直接地评估单个细胞的测序扩增误差，以准确验证单个体细胞中的真实突变。

此外，开发一套专门用于单细胞组学数据集分析的新型生物信息学工具也很重要。这些生物信息学工具应同时考虑单细胞组学数据集的缺点（如高技术噪声和假阴性率）以及优点（如高采样数和唯一分子识别符）。目前拟时序分析在解决干细胞分化过程中的一些中间状态方面存在不足，特别是当这些状态与早期干细胞状态和后期定向分化状态有显著不同时，原因是单细胞转录组分析本质上只提供每个细胞基因表达谱的“快照”，而单个细胞的基因表达变化通常被认为是“连续的”，且在短时间段内可追踪。解决“快照”问题的一个可能的方案是对细胞群进行更密集的采样（理想情况是每小时采样 1 次）。另外，在拟时序算法中加入以下假设可能有助于解决干细胞的真正分化途径问题：较晚的分化时间点极有可能包含较早时间点的分化延迟残留干细胞，而较早的时间点不太可能包含完全分化的细胞。例如，在人类 ESC 分化为肝细胞的过程中，仅在 1 d 或 2 d 后就不太可能在人群中发现功能性肝细胞，但在包含功能性肝细胞的分化几周的细胞群中仍有可能发现一些残留的干细胞样细胞。

单细胞多组学测序技术发展很迅速，能够同时从单个细胞获得基因组、转录组和 DNA 甲基化组，这有助于阐明组学在单个细胞中的不同层次之间的关系<sup>[55]</sup>。如果这些多组学技术成为常规方法，那么我们就可以精确地获得单个细胞中基因组、

表观基因组和转录组信息。比较理想的状态是这样的：运用单细胞基因组测序数据进行谱系追踪，以重建体内干细胞分化期间的细胞谱系；然后分析来自这些细胞的转录组数据，并用于识别复杂组织中的不同细胞类型或亚群；随后用来自同一组单细胞的表观基因组信息来研究不同的表观遗传方式；最后利用基因编辑技术在体内敲除干细胞的关键基因，建立基因型和表型之间的因果关系。同时，可以利用干细胞发育分化过程中多个时间点的单细胞多组学测序重建干细胞分化过程中单个细胞内的核心基因调控网络，揭示每个体细胞内或不同体细胞间的基因型-表型关系，深入了解生理和病理条件下基因调控网络的复杂性，为了解人类个体发育和疾病发生的生物学基础提供新的见解。

#### [参 考 文 献]

- [1] HU S, METCALF E, MAHAT D B, et al. Transcription factor antagonism regulates heterogeneity in embryonic stem cell states[J]. *Mol Cell*, 2022, 82(23): 4410-4427. e12. DOI: 10.1016/j.molcel.2022.10.022.
- [2] KOŁODZIEJCZYK A A, KIM J K, TSANG J C, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation[J]. *Cell Stem Cell*, 2015, 17(4): 471-485. DOI: 10.1016/j.stem.2015.09.011.
- [3] TANG F, BARBACIORU C, WANG Y, et al. mRNA-seq whole-transcriptome analysis of a single cell[J]. *Nat Methods*, 2009, 6(5): 377-382. DOI: 10.1038/nmeth.1315.
- [4] ANDREWS N, SERVISS J T, GEYER N, et al. An unsupervised method for physical cell interaction profiling of complex tissues[J]. *Nat Methods*, 2021, 18(8): 912-920. DOI: 10.1038/s41592-021-01196-2.
- [5] SOZEN B, AMADEI G, COX A, et al. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures[J]. *Nat Cell Biol*, 2018, 20(8): 979-989. DOI: 10.1038/s41556-018-0147-7.
- [6] HWANG B, LEE D S, TAMAKI W, et al. SCITO-seq: single-cell combinatorial indexed cytometry sequencing [J]. *Nat Methods*, 2021, 18(8): 903-911. DOI: 10.1038/s41592-021-01222-3.
- [7] CHEN S, LAKE B B, ZHANG K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell[J]. *Nat Biotechnol*, 2019, 37(12): 1452-1457. DOI: 10.1038/s41587-019-0290-0.
- [8] NAZARIE F W, SHIH B, ANGUS T, et al. Visualization and analysis of RNA-seq assembly graphs[J]. *Nucleic Acids Res*, 2019, 47(14): 7262-7275. DOI: 10.1093/nar/gkz599.
- [9] MCKELLAR D W, MANTRI M, HINCHMAN M M, et al.

- Spatial mapping of the total transcriptome by *in situ* polyadenylation[J]. *Nat Biotechnol*, 2023, 41(4): 513-520. DOI: 10.1038/s41587-022-01517-6.
- [10] CAO Z, ZUO W, WANG L, et al. Spatial profiling of microbial communities by sequential FISH with error-robust encoding[J]. *Nat Commun*, 2023, 14(1): 1477. DOI: 10.1038/s41467-023-37188-3.
- [11] SCHEDE H H, SCHNEIDER C G, STERGIADOU J, et al. Spatial tissue profiling by imaging-free molecular tomography[J]. *Nat Biotechnol*, 2021, 39(8): 968-977. DOI: 10.1038/s41587-021-00879-7.
- [12] MARTIN B K, QIU C, NICHOLS E, et al. Optimized single-nucleus transcriptional profiling by combinatorial indexing[J]. *Nat Protoc*, 2023, 18(1): 188-207. DOI: 10.1038/s41596-022-00752-0.
- [13] HU K H, EICHORST J P, MCGINNIS C S, et al. ZipSeq: barcoding for real-time mapping of single cell transcriptomes[J]. *Nat Methods*, 2020, 17(8): 833-843. DOI: 10.1038/s41592-020-0880-2.
- [14] LEONAVICIENE G, MAZUTIS L. RNA cytometry of single-cells using semi-permeable microcapsules[J]. *Nucleic Acids Res*, 2023, 51(1): e2. DOI: 10.1093/nar/gkac918.
- [15] QIAN X, HARRIS K D, HAULING T, et al. Probabilistic cell typing enables fine mapping of closely related cell types *in situ*[J]. *Nat Methods*, 2020, 17(1): 101-106. DOI: 10.1038/s41592-019-0631-4.
- [16] KIM J, T JAKOBSEN S, NATARAJAN K N, et al. TENET: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data[J]. *Nucleic Acids Res*, 2021, 49(1): e1. DOI: 10.1093/nar/gkaa1014.
- [17] YU L, WEI Y, DUAN J, et al. Blastocyst-like structures generated from human pluripotent stem cells[J]. *Nature*, 2021, 591(7851): 620-626. DOI: 10.1038/s41586-021-03356-y.
- [18] DURANTE M A, KURTENBACH S, SARGI Z B, et al. Single-cell analysis of olfactory neurogenesis and differentiation in adult humans[J]. *Nat Neurosci*, 2020, 23(3): 323-326. DOI: 10.1038/s41593-020-0587-9.
- [19] JUNG M, DOURADO M, MAKSYMETZ J, et al. Cross-species transcriptomic atlas of dorsal root ganglia reveals species-specific programs for sensory function[J]. *Nat Commun*, 2023, 14(1): 366. DOI: 10.1038/s41467-023-36014-0.
- [20] HOFFMAN D, TEVET Y, TRZEBANSKI S, et al. A non-classical monocyte-derived macrophage subset provides a splenic replication niche for intracellular *Salmonella*[J]. *Immunity*, 2021, 54(12): 2712-2723.e6. DOI: 10.1016/j.immuni.2021.10.015.
- [21] ZHANG Z, ZHOU C, LI X, et al. Loss of CHD1 promotes heterogeneous mechanisms of resistance to AR-targeted therapy via chromatin dysregulation[J]. *Cancer Cell*, 2020, 37(4): 584-598.e11. DOI: 10.1016/j.ccell.2020.03.001.
- [22] CHAKRABORTY M, HU S, VISNESS E, et al. MicroRNAs organize intrinsic variation into stem cell states[J]. *Proc Natl Acad Sci U S A*, 2020, 117(12): 6942-6950. DOI: 10.1073/pnas.1920695117.
- [23] KOZAREVA V, MARTIN C, OSORNO T, et al. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types[J]. *Nature*, 2021, 598(7879): 214-219. DOI: 10.1038/s41586-021-03220-z.
- [24] POSFAI E, SCHELL J P, JANISZEWSKI A, et al. Evaluating totipotency using criteria of increasing stringency[J]. *Nat Cell Biol*, 2021, 23(1): 49-60. DOI: 10.1038/s41556-020-00609-2.
- [25] MITTNENZWEIG M, MAYSHAR Y, CHENG S, et al. A single-embryo, single-cell time-resolved model for mouse gastrulation[J]. *Cell*, 2021, 184(11): 2825-2842.e22. DOI: 10.1016/j.cell.2021.04.004.
- [26] GUO G, HUSS M, TONG G Q, et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst[J]. *Dev Cell*, 2010, 18(4): 675-685. DOI: 10.1016/j.devcel.2010.02.012.
- [27] YANAGIDA A, CORUJO-SIMON E, REVELL C K, et al. Cell surface fluctuations regulate early embryonic lineage sorting[J]. *Cell*, 2022, 185(5): 777-793.e20. DOI: 10.1016/j.cell.2022.01.022.
- [28] BONORA G, RAMANI V, SINGH R, et al. Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation[J]. *Genome Biol*, 2021, 22(1): 279. DOI: 10.1186/s13059-021-02432-w.
- [29] LENG L, SUN J, HUANG J, et al. Single-cell transcriptome analysis of uniparental embryos reveals parent-of-origin effects on human preimplantation development[J]. *Cell Stem Cell*, 2019, 25(5): 697-712.e6. DOI: 10.1016/j.stem.2019.09.004.
- [30] BLAKELEY P, FOGARTY N M, DEL VALLE I, et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq[J]. *Development*, 2015, 142(20): 3613. DOI: 10.1242/dev.131235.
- [31] TANG F, BARBACIORU C, BAO S, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis[J]. *Cell Stem Cell*, 2010, 6(5): 468-478. DOI: 10.1016/j.stem.2010.03.015.
- [32] YAN L, YANG M, GUO H, et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells[J]. *Nat Struct Mol Biol*, 2013, 20(9): 1131-1139. DOI: 10.1038/nsmb.2660.
- [33] PHANSALKAR R, RED-HORSE K. Techniques converge to map the developing human heart at single-cell level[J]. *Nature*, 2020, 577(7792): 629-630. DOI: 10.1038/d41586-020-00151-z.
- [34] GARCÍA-CASTRO H, KENNY N J, IGLESIAS M, et al. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics[J].

- Genome Biol, 2021, 22(1): 89. DOI: 10.1186/s13059-021-02302-5.
- [35] YAMANAKA S. Pluripotent stem cell-based cell therapy-promise and challenges[J]. Cell Stem Cell, 2020, 27(4): 523-531. DOI: 10.1016/j.stem.2020.09.014.
- [36] WU C L, DICKS A, STEWARD N, et al. Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis[J]. Nat Commun, 2021, 12(1): 362. DOI: 10.1038/s41467-020-20598-y.
- [37] KAMATANI T, HAGIZAWA H, YARIMITSU S, et al. Human iPS cell-derived cartilaginous tissue spatially and functionally replaces nucleus pulposus [J]. Biomaterials, 2022, 284: 121491. DOI: 10.1016/j.biomaterials.2022.121491.
- [38] HOLLOWAY E M, WU J H, CZERWINSKI M, et al. Differentiation of human intestinal organoids with endogenous vascular endothelial cells[J]. Dev Cell, 2020, 54(4): 516-528.e7. DOI: 10.1016/j.devcel.2020.07.023.
- [39] TSANG J C, YU Y, BURKE S, et al. Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in *Bcl11a*-deficient hematopoietic stem cells[J]. Genome Biol, 2015, 16: 178. DOI: 10.1186/s13059-015-0739-5.
- [40] KOWALCZYK M S, TIROSH I, HECKL D, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells[J]. Genome Res, 2015, 25(12):1860-1872. DOI: 10.1101/gr.192237.115.
- [41] ZHOU Y, SU Y, LI S, et al. Molecular landscapes of human hippocampal immature neurons across lifespan [J]. Nature, 2022, 607(7919): 527-533. DOI: 10.1038/s41586-022-04912-w.
- [42] BOTTES S, JAEGER B N, PILZ G A, et al. Long-term self-renewing stem cells in the adult mouse hippocampus identified by intravital imaging[J]. Nat Neurosci, 2021, 24(2): 225-233. DOI: 10.1038/s41593-020-00759-4.
- [43] GEHART H, VAN ES J H, HAMER K, et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping[J]. Cell, 2019, 176(5): 1158-1173.e16. DOI: 10.1016/j.cell.2018.12.029.
- [44] KREMSKY I, CORCES V G. Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance[J]. Genome Biol, 2020, 21(1): 118. DOI: 10.1186/s13059-020-02036-w.
- [45] NEFTEL C, LAFFY J, FILBIN M G, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma[J]. Cell, 2019, 178(4): 835-849.e21. DOI: 10.1016/j.cell.2019.06.024.
- [46] MIN B, PARK J S, JEONG Y S, et al. Dnmt1 binds and represses genomic retroelements via DNA methylation in mouse early embryos[J]. Nucleic Acids Res, 2020, 48(15): 8431-8444. DOI: 10.1093/nar/gkaa584.
- [47] MODZELEWSKI A J, SHAO W, CHEN J, et al. A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development[J]. Cell, 2021, 184(22): 5541-5558.e22. DOI: 10.1016/j.cell.2021.09.021.
- [48] SMALLWOOD S A, LEE H J, ANGERMUELLER C, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity[J]. Nat Methods, 2014, 11(8): 817-820. DOI: 10.1038/nmeth.3035.
- [49] YAN R, CHENG X, GU C, et al. Dynamics of DNA hydroxymethylation and methylation during mouse embryonic and germline development[J]. Nat Genet, 2023, 55(1): 130-143. DOI: 10.1038/s41588-022-01258-x.
- [50] O'CONNELL B L, NICHOLS R V, POKHOLOK D, et al. Atlas-scale single-cell chromatin accessibility using nanowell-based combinatorial indexing[J]. Genome Res, 2023, 33(2): 208-217. DOI: 10.1101/gr.276655.122.
- [51] MEULEMAN W, MURATOV A, RYNES E, et al. Index and biological spectrum of human DNase I hypersensitive sites[J]. Nature, 2020, 584(7820): 244-251. DOI: 10.1038/s41586-020-2559-3.
- [52] WINICK-NG W, KUKALEV A, HARABULA I, et al. Cell-type specialization is encoded by specific chromatin topologies[J]. Nature, 2021, 599(7886): 684-691. DOI: 10.1038/s41586-021-04081-2.
- [53] MISHRA S, PANDEY N, CHAWLA S, et al. Matching queried single-cell open-chromatin profiles to large pools of single-cell transcriptomes and epigenomes for reference supported analysis[J]. Genome Res, 2023, 33(2):218-231. DOI: 10.1101/gr.277015.122.
- [54] MINUSSI D C, NICHOLSON M D, YE H, et al. Breast tumours maintain a reservoir of subclonal diversity during expansion[J]. Nature, 2021, 592(7853): 302-308. DOI: 10.1038/s41586-021-03357-x.
- [55] BAKKEN T E, JORSTAD N L, HU Q, et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse[J]. Nature, 2021, 598(7879): 111-119. DOI: 10.1038/s41586-021-03465-8.