单细胞多组学技术新进展及其在发育生物学研究中的应用

程馨, 燕蕊, 郭帆*

四川大学华西医学中心&华西第二医院, 出生缺陷与妇女儿童疾病教育部重点实验室, 成都 610041

* 联系人, E-mail: guofan@scu.edu.cn

收稿日期: 2020-07-30; 接受日期: 2020-09-28; 网络版发表日期: 2020-12-16

摘要 单细胞多组学技术基于单细胞测序技术和高通量组学方法发展而来, 能在单细胞分辨率下同时整合转录组、基因组、表观遗传组、蛋白质组或空间状态等两种或多种组学信息. 相比单细胞单组学技术, 单细胞多组学技术能多层次、多角度地描绘细胞间的异质性, 挖掘各层组学之间的直接和潜在联系, 从而更加全面和系统地描绘细胞的状态和命运, 在生命科学和医学多领域有广泛的应用前景. 本文总结了现有的单细胞多组学技术, 预测了其发展趋势, 讨论了其在发育生物学中的应用实例和潜力.

关键词 单细胞测序, 单细胞多组学, 发育生物学

单细胞测序技术的不断发展解决了以往群体细胞组学分析面临的两个瓶颈: 一是大量细胞的混合测序会掩盖细胞间的异质性; 二是传统的测序技术需要大量的细胞作为起始, 细胞数目稀有的样品则难以研究. 利用单细胞测序方法不仅能够揭示组织器官中的细胞类型和稀有亚群、细胞状态和命运的变化, 还可以研究细胞数稀有的珍贵样品, 从而帮助人们理解发育、衰老和疾病发生发展的机制. 2013年, 单细胞测序技术被评选为*Nature Methods*的年度方法^[1]. 自此, 随着单细胞转录组、基因组、表观遗传组、蛋白组等各组学测序技术的快速发展, 研究者开始将它们相互组合, 使得单个细胞或同类细胞的多个组学数据能够联合分析. 因此, 利用单细胞多组学技术能够更有效地研究细胞在特定时空状态下各组学间的复杂联系, 从而揭示

细胞状态和命运调控的分子机制. 2019年, 单细胞多组学技术凭借其潜在的发展和应用空间被*Nature Methods*评选为年度方法^[2].

哺乳动物的胚胎由单个受精卵发育而来, 经过一系列连续、复杂且精准的调控, 形成包含所有细胞谱系和类型的完整个体^[3]. 尽管围绕胚胎发育的研究已开展多年, 但关于哺乳动物的发育机制仍有许多重要问题有待解决. 例如, 胚胎发育过程中的多能性如何建立? 细胞谱系和命运决定如何发生? 而这些重要科学问题的核心是: 遗传信息如何按照一定的时间和空间顺序来调控个体发育. 单细胞多组学技术不仅克服了早期胚胎细胞数目少的限制, 而且使人们能够更加全面和详细地对细胞命运调控进行研究^[3], 为发育生物学中重大问题的解决提供了有力帮助. 本文分析了

引用格式: 程馨, 燕蕊, 郭帆. 单细胞多组学技术新进展及其在发育生物学研究中的应用. 中国科学: 生命科学, 2021, 51: 496–506
Cheng X, Yan R, Guo F. Advances in single-cell multi-omics methods and their applications in developmental biology (in Chinese). *Sci Sin Vitae*, 2021, 51: 496–506, doi: [10.1360/SSV-2020-0244](https://doi.org/10.1360/SSV-2020-0244)

单细胞多组学技术的发展现状和未来趋势,总结了其在发育生物学研究中的重要应用和巨大潜力。

1 单细胞多组学技术的发展概况

1.1 单细胞多组学技术的概念

单细胞转录组测序技术因其能直观反映细胞基因表达水平的异质性,迅速成为发现复杂组织中新细胞类型、鉴别发育和疾病中细胞状态的热门技术^[4]。从2009年发展至今,其单次测序通量已达数十万个细胞,测序成本不断降低,建库技术更加自动化,数据分析方法也更成熟^[5]。同时,细胞内其他分子层面的高通量测序和检测方法也不断发展,出现了能在单细胞分辨率上对基因组^[6,7]、染色质可及性^[8~10]、组蛋白修饰^[11,12]、核小体定位^[9,13]和染色质3D构象^[14]等进行测量的多种技术方法(图1)。单细胞组学技术方法的不断发展和改进,为单细胞层级的多组学整合分析奠定了基础。

从单个细胞到复杂组织再到生物个体,生命活动的进行始终伴随着复杂的分子网络调控,遵从特定的生物学规律。想要解析细胞内最基本的调控网络,仅单方面测量mRNA或其他任何一类大分子都是不够的。为了能深入分析和理解细胞状态和命运决定的机

制,将单细胞各种组学技术方法有机结合在一起的单细胞多组学技术应运而生。

2016年发表的DR-seq^[15]和G&T-seq^[16]技术,首次实现在单细胞内同时进行基因组和转录组的测序分析,使得研究人员能够在分析细胞间基因组差异和基因表达异质性的同时,进一步理解基因组序列差异、拷贝数变异与转录组异质性之间的关系。随后在同一个单细胞内,基于转录组、染色质可及性^[17~20]、染色质构象^[21,22]、DNA拷贝数^[15,16,23,24]、DNA甲基化^[25~27]、蛋白质丰度^[28,29]或者空间状态^[30~34]等整合分析的多组学方法也相继出现。

相较于利用不同的单细胞单组学测序方法从一类细胞中分别获取转录组、基因组、表观遗传组等进行分析,单细胞多组学测序技术可以从同一个单细胞中同时获取多种组学数据,可以更好地反映细胞在特定状态下各组学间的关联以及复杂的分子调控网络。例如,将转录组与细胞表面蛋白分析结合的CITE-seq^[28]和REAP-seq^[29],能更好地分辨转录组层面差异微小而功能蛋白分子存在差别的细胞亚群,非常适用于鉴定和研究免疫细胞及其功能。另外,机体内每个细胞发挥功能都是在其特定的细胞微环境中,在进行单细胞测序的同时,记录细胞在组织内的空间状态信息,可以帮助研究者更好地分析组织或器官内特定细胞的功能。

图1 目前单细胞测序技术可以检测的细胞内代表性分子层面

Figure 1 Representative molecular layers in a cell that can be detected by single-cell sequencing technologies

能和不同细胞间的关系。

1.2 目前代表性单细胞多组学技术

目前, 文献报道的单细胞多组学测序技术已超二十余种, 涵盖多种组学的整合方式(图2)。在此, 本文将现有的代表性单细胞多组学测序技术归为两大类: 一类是分子检测型, 指检测细胞在特定状态下的多种分子层面特征的方法; 另一类是功能研究型, 这类方法集分子检测与功能研究为一体, 探索细胞在人为扰动下基因调控网络的变化。例如, 基因编辑技术CRISPR (clustered regularly interspaced short palindromic repeats)与单细胞测序技术的联用。

分子检测型单细胞多组学测序技术大多数以转录组测序为核心, 这是因为在中心法则中信使RNA(messenger RNA, mRNA)位于中间位置, 与其他重要分子层面有密切关联, 能很好地体现细胞间功能差异, 并且转录组测序也是目前商业化最为成熟的技术^[35]。首先

出现的单细胞多组学方法就是将转录组与其他组学技术结合。例如, TARGET-seq^[24], DR-seq^[15], G&T-seq^[16], SCGT^[36]和SIDR^[23]等技术能同时进行基因组和转录组测序, 不仅能探究基因表达过程中DNA拷贝数变异与基因表达变化的关系, 还可以利用这两个层面的信息构建更完整、更准确的细胞发育轨迹^[37]。

转录组与表观遗传组联合分析, 能够深入研究基因表达的表观遗传调控规律。scMT-seq^[25], scM&T-seq^[27]和sc-GEM^[26]等测序方法整合了单细胞转录组和DNA甲基化组测序技术, 使研究者能够分析细胞中mRNA的可变剪接与DNA甲基化间的关系; 而sci-CAR-seq^[38], SNARE-seq^[17], scCAT-seq^[39], ATAC-RNA-seq^[19]和Paired-seq^[20]等技术能在单细胞中同时获得染色质可及性和转录组数据, 使研究者可以在单细胞分辨率下研究基因组顺式调控元件和反式调节因子在细胞命运决定中的作用; CITE-seq^[28], REAP-seq^[29], PLAYR^[40]和PEA/STA^[41]等技术将单细胞转录

技术/方法名称	转录组	染色质可及性	DNA甲基化	DNA拷贝数	空间信息	蛋白质组	染色质构象	TCR	CRISPR扰动	参考文献
sciCAR-seq	●	●								[38]
Paired-seq	●	●								[20]
SNARE-seq	●	●								[17]
scCAT-seq	●	●								[39]
ATAC-RNA-seq	●	●								[19]
scMT-seq	●		●							[25]
scM&T-seq	●		●							[27]
sc-GEM	●		●							[26]
TARGET-seq	●			●						[24]
DR-seq	●			●						[15]
G&T-seq	●			●						[16]
SCGT	●			●						[36]
SIDR	●			●						[23]
Slide-seq	●				●					[33]
MERFISH	●				●					[30]
ST-RNA-seq	●				●					[46]
STAR-map	●				●					[34]
seqFISH+	●				●					[32]
CITE-seq	●					●				[28]
REAP-seq	●					●				[29]
PLAYR	●					●				[40]
PEA/STA	●					●				[41]
scNOMe-seq		●	●							[49]
Methyl-HiC			●				●			[22]
scCOOL-seq		●	●	●						[48]
scCOOL-seq2		●	●	●						[73]
scNMT-seq	●	●	●							[18]
scTrio-seq	●		●	●						[44]
scTrio-seq2	●		●	●						[45]
ECCITE-seq	●					●		●	●	[42]
CRISP-seq	●								●	[51]
Perturb-seq	●								●	[52]
CROP-seq	●								●	[53]
Mosaic-seq	●								●	[54]
Perturb-ATAC		●							●	[55]

图2 目前代表性单细胞多组学技术总结

组测序和细胞表面蛋白检测相结合, 能够更有效地鉴定细胞类型和研究细胞功能。

随着研究的需求和技术的进步, 单细胞转录组结合其他多种组学的方法也逐渐丰富起来。ECCITE-seq^[42]技术整合了单细胞转录组、表面蛋白、CRISPR 扰动和TCR(T cell receptor)克隆检测, 能够用于精确定义免疫细胞类型。最近, Overall等人^[43]对ECCITE-seq技术进行了优化, 借助新研发的pMHC-I(peptide-MHC, 抗原肽-MHC分子复合物)四聚体探针文库, ECCITE-seq能更准确地进行TCR克隆检测。scTrio-seq^[44]和scTrio-seq2^[45]技术能同时检测单细胞转录组、DNA 拷贝数和DNA甲基化, 已被应用于研究肝癌和结肠癌发生发展的机制。scNMT-seq^[18]技术整合了单细胞转录组、染色质可及性和DNA甲基化三个重要组学层面, 研究者利用该方法探索小鼠胚胎干细胞的多能性维持与分化, 解析了该过程中三种组学间的互动关系。

在复杂组织中, 细胞的空间位置和微环境对于细胞功能有重要的影响。Slide-seq^[33], MERFISH^[30], seq-FISH+^[32], ST-RNA-seq^[46]和STAR-map^[34]等技术能在保留细胞空间信息的同时进行单细胞转录组测序, 将细胞的功能与微环境联系起来, 这将有利于研究者深入研究和分析生理、疾病状态下各种细胞的功能以及细胞间的关系。

分子检测型单细胞多组学测序技术中也有一些集中于探究单细胞内表观遗传组各层面的变化及关联。同一生物个体的所有体细胞拥有几乎相同的基因组, 基因表达差异、特定细胞功能和表型的产生受DNA修饰、组蛋白修饰和染色质状态及结构的影响^[47]。因此, 在单细胞中分析表观遗传组的多个不同分子层面, 能够研究细胞表观遗传差异与细胞表型间的关系。scCOOL-seq^[48]和scNOMe-seq^[49]技术可以在单细胞中同时检测染色质可及性和DNA甲基化, 可用于研究发育过程中细胞内DNA甲基化的变化对染色质状态的影响。Methyl-HiC^[22]技术则可以同时分析DNA甲基化和染色质构象, Li等人利用此技术发现细胞中三维空间上临近的DNA分子具有相似的DNA甲基化水平, 并解析了复杂神经组织中染色质结构和DNA甲基化在不同细胞间的异质性。

功能研究型单细胞多组学测序技术, 结合了强大的CRISPR基因编辑技术, 从而可以深度挖掘基因表达调控与细胞功能以及命运决定之间的因果关系^[50]。同

时, 结合CRISPR文库筛选和单细胞转录组测序的单细胞多组学技术解决了功能基因组学的技术瓶颈, 即在进行高通量基因编辑的同时进行细胞表型测定, 极大地提高了基因功能鉴定的效率。

CRISP-seq^[51]技术是最早的功能研究型单细胞多组学测序技术, Jaitin等人利用CRISP-seq技术重建了血液髓系细胞在分化过程中的核心分子调控网络以及在病原体侵染中的应激反应。随后出现的Perturb-seq^[52], CROP-seq^[53]和Mosaic-seq^[54]等技术的策略也都利用了CRISPR编辑技术结合单细胞转录组测序技术。而Perturb-ATAC^[55]技术则整合了CRISPR文库筛选和单细胞染色质可及性测序技术, 利用CRISPR技术高通量地剔除反式调控因子, 通过检测染色质可及性变化研究反式调控因子在不同类型细胞调控网络中的作用。Perturb-ATAC技术可用于鉴定关键转录因子(transcription factors, TFs)、染色质修饰因子(chromatin-modifying factors)和非编码RNA(noncoding RNAs, ncRNAs)在细胞功能维持和命运决定中的作用。Rubin等人^[55]利用Perturb-ATAC技术解析了在淋巴B细胞中起重要作用的顺式调控元件和转录因子, 揭示了人角质形成细胞命运转变过程中的三个关键转录因子模块。

在未来, 为了更全面系统地研究多层组学之间的互动关系对细胞命运的影响, 一方面, 单细胞多组学测序技术的策略会继续向转录组结合其他多个组学的三组学甚至四组学方向发展。另一方面, 单细胞多组学测序技术会转向研发多组学测序分析和功能验证同时进行的方法, 这种探究因果关系的策略对于揭示发育和疾病发生发展中的关键因子和分子网络模块十分重要。

2 单细胞多组学测序技术的新趋势

2.1 改进不同分子层面的建库效率

单细胞多组学测序技术的发展依赖于各种单一组学检测技术的完善, 但即使在单细胞测序技术迅速发展的情况下, 在单细胞水平协同检测多个不同的分子层面仍然非常具有挑战性。每一种组学都有不同的特性和测量方法, 这些方法在敏感性、特异性、通量和适用性上都有所不同^[47]。因此, 在单细胞水平对多种组学方法进行整合和应用时, 需要在策略上做出改进,

以保证各组学间的上游建库以及下游生物信息分析均能良好地兼容, 同时也要明确技术的局限性及其对研究结果的潜在影响^[47].

(1) 提高单次建库中单细胞的分离效率. 发明和利用高通量的单细胞分离工具能显著缩短实验时间, 借助微流控设备(microfluidic devices)能将单细胞分离通量提高到上万个单细胞, 同时反应体积的缩小也显著降低了每个细胞的检测成本^[47], 适合分析多细胞的复杂组织. 低通量的单细胞分离技术能保证珍稀样本的采集效率和测序深度, 一些技术还能保留细胞的空间信息, 如激光捕获显微切割技术(laser capture microdissection)^[56], 但相应的测序成本也随之增高. 发明和利用新的单细胞分离工具, 在提高单细胞分离通量的同时保证采集效率和测序深度将是未来的发展趋势.

(2) 提高单次建库中单细胞的检测通量. 高通量的单细胞转录组测序技术通常是在分离或捕捉到单细胞后, 对其mRNA分子添加特异的细胞标签(cell barcode), 即可在测序数据中区分不同单细胞来源的mRNA. 根据添加细胞标签方式的不同, 可将单细胞转录组测序技术分为三大类: 一是基于微流控系统的Drop-seq^[57], inDrops^[58]和10x Chromium Single-cell RNA-seq等方法; 二是基于微孔板或微孔芯片的CEL-seq^[59], seq-Well^[60]和Microwell-seq^[61]等方法; 三是基于拆分-合并(split and pool)的随机组合标记方法, 如sci-RNA-seq^[62]和SPLiT-seq^[63]技术. 但目前, 仅有个别单细胞多组学测序技术, 如CITE-seq^[28]能与10x Genomics高通量测序系统兼容^[50], 大部分单细胞多组学技术仍受到细胞检测通量的限制.

(3) 提高单细胞内各组分的提取效率. 在单细胞多组学测序中, 除了选择合适的单细胞分离技术, 还需要考虑以下问题: 如何更完整地提取每一种大分子; 如何更好地识别和扩增其代表的信息; 如何给不同的细胞、不同的分子添加合适的标记; 如何减少建库、测序、算法识别等操作过程中引起的人工误差. 在这些问题的基础上, 还需要考虑如何裂解细胞保证同时获取多种大分子且保持各自完整性; 依据不同大分子的特性, 在什么节点从裂解液中分离各组分; 如何平衡降低各组分的损失与保证测序质量的关系.

例如, 在单细胞中同时进行转录组与基因组测序时, DNA与mRNA的分离策略有不同的方式. 在DR-seq^[15]中细胞被完全裂解, DNA与mRNA逆转录得到

的cDNA在同一体系中同时扩增, 然后一分为二分别使用CEL-seq^[59]方法测量转录组和MALBAC^[64]方法测量基因组. 而G&T-seq^[16]选择使用磁珠将具有poly-A尾巴的mRNA从DNA中先分离出来, 然后通过Smart-seq2^[65]方法和WGA^[66](whole genome amplification)方法分别进行单细胞转录组与基因组文库的构建. 此外, SIDR^[23]技术则使用低渗裂解液分离细胞核和胞质裂解液, 然后分别对细胞质mRNA和细胞核DNA进行单细胞测序分析.

2.2 创新实验技术和组合策略

目前, 单细胞多组学技术在各组学大分子组分的有效分离、低损失建库、降低成本、提高检测灵敏度和准确性、优化生物信息算法等方面都面临众多挑战.

(1) 增加单细胞各组学的覆盖度. 在进行单细胞多组学测序分析时, 首先需要从样本中获取和扩增信号, 但在提取和信号扩增的过程中不可避免会丢失信息和引入偏差. 若想减少信息丢失并提高覆盖度, 一方面可以通过计算方法进行多组学间信息的互补分析, 另一方面可通过增加自动化操作, 减少转移、分离等步骤尽量减少信息丢失. 目前针对最成熟的转录组测序, 研究者也正在开发新一代能直接对RNA进行测序的纳米孔测序系统, 该系统有望消除逆转录和非特异性扩增引入的偏差^[3].

(2) 引入蛋白组测量技术. 目前, 研究细胞表面蛋白最常用的方法是利用耦联发光分子(荧光染料或稳定金属同位素)的抗体来标记目标蛋白, 然后进行FACS(fluorescence activating cell sorting)检测, 但其能同时检测的蛋白数有限, 最多仅能达到40多种^[47]. 而以CITE-seq^[28], REAP-seq^[29], ECCITE-seq^[42]为代表的单细胞多组学测序技术将抗体耦联的发光分子替换为带有DNA标签(DNA barcode)的寡核苷酸序列, 将蛋白质的丰度测量转变为特定DNA序列的定量扩增上, 使细胞表面蛋白能够与DNA和RNA兼容地进行测序和分析. 然而, 由于抗体的灵敏度和现有抗体种类的限制, 包括上述策略在内的所有依赖于抗体捕捉蛋白的方法难以达到更高的细胞通量^[47]. 值得一提的是, 单细胞质谱分析蛋白质组学方法(single-cell proteomics by mass spectrometry, SCoPE-MS)^[67]利用串联质谱标签系统(tandem mass tags, TMT)目前可以测量多达583

种蛋白质.若能继续提高质谱灵敏度和检测速度,将会极大地促进单细胞蛋白质组检测技术的发展^[47].

长久以来,受到起始样品量少以及不能直接放大蛋白质信号的限制,单细胞蛋白质组学研究方法的开发相比其他单细胞组学方法更加困难.在未来,蛋白质组学技术研究会朝着利用物理、化学和生物等多学科交叉的创新方法前进,提高蛋白质组学检测通量,解决单细胞内蛋白质组学与其他组学信息的同时获取等难题.

(3) 引入空间组学技术.随着单细胞各组学研究方法的发展,以神经科学为代表的生命科学研究对能够保留细胞固有空间信息的空间组学的需求也与日俱增.经过研究者的不断尝试,单细胞空间转录组学有了重大进展.最初研究者尝试利用RNA荧光杂交技术对细胞或组织进行原位测序(single molecule fluorescent in situ hybridization, smFISH)^[68],但该方法会受到荧光信号密度和转录本长度等限制.为了克服这些问题,osmFISH^[31]和seqFISH^[69]方法使用带有荧光团的探针进行多轮顺序杂交,识别特定的转录本并采集成像,然后计算每个位置的基因转录水平.目前,改进后的seqFISH+^[32]方法可以达到1万个基因转录本的测量和定位.但以上两种方法都需要特殊设备,以及预先选择用于测量的基因探针文库.2019年报道的Slide-seq^[33]技术则是将组织切片中的mRNA转移到包裹有带位置信息的标签微珠(barcode beads)的液滴中,然后再进行常规的逆转录、扩增和文库构建.

不论是基于显微镜成像的荧光原位测序技术,还是基于带有位置信息的DNA标签的磁珠捕捉测序技术,都可以结合将单细胞映射到空间参考图的计算方法,在单细胞分辨率下构建全转录组的空间图谱^[4].随着单细胞空间转录组测序技术的成熟和商业化,在未来,新的单细胞空间组学采集技术和算法的开发,有望将除转录组以外的其他组学层次绘制成二维和三维图谱.其他组学与空间方法的整合将有助于识别特定微环境对细胞分子调控网络的影响,这将有利于研究者对健康组织和肿瘤组织的细胞谱系、演变轨迹和功能状态等进行深入研究^[4].

(4) 增加单细胞多组学组合策略.原理上单细胞各个组学可以随机组合,用于研究它们之间的互动关系,但单细胞多组学技术的开发常受到可行性和实用性的限制.目前,单细胞多组学技术大多是各实验室

根据研究方向的需求和当前的技术水平进行开发,并且这些技术的发展也各有差异,其中一些趋于成熟且得到应用,一些不够稳定应用较少^[47].目前,单细胞中能够检测的组学层次主要有7种(图1),已有的技术涵盖了其中部分组合方式(图2),还有很多可能的组合方式没有被开发^[35].限于单细胞内各组学层面的分子起始数量极低,在单细胞分辨率上建立新的多种组学或者全组学的测量技术仍然十分困难.

2.3 优化数据统计和计算方法

单细胞多组学测序主要包含细胞分离、文库构建、高通量测序和生物信息学分析等步骤.因此,单细胞多组学测序技术的发展也为相关算法的开发带来了机遇和挑战^[2].利用最先进的多模态计算方法不仅可以更好地进行细胞分群和谱系追踪,还可以推断基因表达调控网络和解析细胞在组织中的空间位置等信息^[4].多组学的联合分析需要解决的核心计算问题是如何将大量单细胞产生的多种不同的组学数据进行有效地生成、识别、整合与分析,并且降低数据背景噪音和样本间的批次差异.

目前,单细胞多组学的生物信息分析方法仍有待发展,面临着诸多问题与挑战.第一个挑战是对大量不同组学数据的处理能力.单细胞多组学的数据合并后,数据量更加庞大,也带来了更大的计算压力,在有限的计算资源下,合理分配资源和处理可用数据变得尤为重要^[35].第二个挑战则是对多组学数据进行综合分析处理,包容不同组学数据各自的统计学和生物学特征,消除样本间批次效应等混杂因素,所以需要采用复杂的统计和计算策略进行集成计算^[2].第三个挑战是需要建立稳定的、用于对单细胞多组学方法进行评估和检验的基准测试算法(benchmarking pipeline)^[35].第四个挑战则是如何对这些数据进行深入挖掘,不仅要揭示细胞状态和命运转变的机制,还要能解析出基因表达调控的复杂网络^[2].

3 单细胞多组学测序技术在发育生物学中的应用

一直以来,绘制个体发育过程中的细胞图谱和揭示细胞命运决定的内在调控机制是发育生物学的基本目标.利用单细胞转录组测序技术,研究人员绘制了小

鼠胚胎发育多个阶段的细胞转录组图谱。单细胞DNA甲基化和染色质可及性等表观遗传组测序技术开辟了研究细胞基因表达调控机制的途径。整合后的单细胞多组学测序技术通过密集采样、全面记录细胞在不同阶段的多种分子水平的动态变化,可以研究基因表达调控网络与细胞命运决定的关系,有助于研究者理解胚胎发育机制,并为在体外重建细胞分化和器官发生等过程提供重要线索^[70]。

先前,利用单细胞转录组测序结合特定算法,研究者重构了脊椎动物从多能性状态到器官形成的胚胎发育轨迹^[71],但通过单细胞转录组测序进行发育轨迹分析仍有一定的局限性。因为仅基于转录水平变化就推断细胞发育和分化轨迹,会遗漏掉不同的基因表达分子轨迹可能产生相同的细胞类型这一实验观察事实^[71]。相比之下,单细胞多组学测序技术具有从多角度、多层次精确重绘细胞分子轨迹的优势。另外,不论是单细胞单组学还是多组学测序方法,只能对单个细胞的瞬时状态进行一次记录,推断出的分子轨迹模型不一定能真实反应细胞谱系发生^[71]。为此,Wagner和Klein^[70]提出单细胞多组学测序结合谱系追踪(lineage tracing)的新技术,该技术能动态追踪细胞的变化并记录其发育历史,从而绘制细胞谱系图谱。

转录组测序结合表观遗传组测序的策略,能够规避单细胞转录组只能观测高表达基因的局限,是研究细胞内在调控网络的有力手段。2017年发表的scCOOL-seq^[48]技术实现了在单细胞水平同时对转录组、DNA甲基化、染色质可及性、染色质倍性和DNA拷贝数五个层面的分析。Li等人^[72]使用该技术检测了人类胚胎植入前六个重要时间点的DNA甲基化和染色质的动态变化过程,发现在人类胚胎植入前亲本基因组间DNA甲基化和染色质开放程度均具有显著差异。此后,利用优化后的scCOOL-seq2技术,Gu等人^[73]绘制了小鼠卵细胞生长过程中的单细胞多组学图谱,描述了染色质可及性在卵细胞生长过程中的动态变化过程,揭示了染色质可及性与DNA甲基化的关系,并鉴定了与卵母细胞成熟有关的表观遗传因子和转录因子。此外,Liu等人^[39]研发的scCAT-seq技术,综合了ATAC-seq^[8]和Smart-seq2^[65]的建库原理,可以同时单细胞的染色质可及性和全长转录组进行分析,不仅能帮助重建顺式调控元件与靶基因表达之间的调控关系,还具有辅助人类胚胎植入前的胚胎质量诊断,

发现癌症靶标等临床应用潜力。

单细胞多组学测序技术还具有揭示早期胚胎发育中表观遗传调控机制的潜力。Clark等人^[18]开发的scNMT-seq方法可以同时分析单细胞染色质可及性、DNA甲基化和转录组。他们将该技术用于研究小鼠胚胎的原肠发生,发现在三个胚层分化形成前,关键顺式调控元件的表观遗传修饰状态和染色质可及性状态就已经预先完成建立或重构^[74]。

4 单细胞多组学测序技术的前景与展望

单细胞多组学测序技术整合了单细胞各个组学特征,提供了精确揭示细胞内复杂分子调控网络及其互动关系的机会。结合不断进步的生物信息学算法,研究者不仅能更好地揭示复杂组织的细胞间异质性,还能发现更多重要的细胞功能调节机制。

未来,单细胞多组学测序技术的发展,一方面将集中在对现有技术的优化和新方法的开发。现有技术的优化不仅包括建库手段、测序技术和算法工具等方面,还包括优化实验设计、自动化实验流程以及提升实验效率。为了更好地检测到各种大分子的特征,研究者可以考虑从生物、化学和物理等多学科中寻找解决方案。此外,新的计算方法和分析工具能帮助研究者越过一些实验限制,发现更多新奇的生物过程。例如,拟时序分析(pseudotime trajectory)将有助于人们理解早期胚胎发育的细胞分化轨迹。新方法则可以更多地关注目前未能检测到的一些重要的基因表达调控层面,如DNA三维结构、非编码RNA和胞内蛋白等。另一方面,目前的单细胞多组学测序技术发展势头强劲,研究者不断研发并丰富着单细胞测序工具箱,各种组学方法百家争鸣。未来,研究者需要从众多相似的方法中筛选出一些稳定、通用、成本低、可商业化的方法,并进一步扩大其实际应用范围。

哺乳动物的早期胚胎具有细胞数稀少且细胞间异质性高的特点,单细胞多组学技术具有全面捕捉细胞多种组学信息的能力,这对于理解胚胎发育中的一些关键变化有重要意义,如胚胎植入后的细胞分化、细胞状态转变的机制等。综上所述,未来随着单细胞多组学技术的不断进步,研究者将有机会在单细胞分辨率以及多重维度下进一步理解个体发育以及疾病的发生发展机制。

参考文献

- 1 Method of the Year 2013. *Nat Methods*, 2014, 11: 1
- 2 Method of the Year 2019: Single-cell multimodal omics. *Nat Methods*, 2020, 17: 1
- 3 Pijuan-Sala B, Guibentif C, Göttgens B. Single-cell transcriptional profiling: a window into embryonic cell-type specification. *Nat Rev Mol Cell Biol*, 2018, 19: 399–412
- 4 Efremova M, Teichmann S A. Computational methods for single-cell omics across modalities. *Nat Methods*, 2020, 17: 14–17
- 5 Kolodziejczyk A A, Kim J K, Svensson V, et al. The technology and biology of single-cell RNA sequencing. *Mol Cell*, 2015, 58: 610–620
- 6 Ng P C, Kirkness E F. Whole genome sequencing. In: Barnes M, Breen G, eds. *Genetic Variation. Methods in Molecular Biology (Methods and Protocols)*. Totowa: Humana Press, 2010. 215–226
- 7 Yoshinaga Y, Daum C, He G, et al. Genome sequencing. In: de Vries R, Tsang A, Grigoriev I, eds. *Fungal Genomics. Methods in Molecular Biology*. New York: Humana Press, 2018. 37–52
- 8 Buenrostro J D, Wu B, Chang H Y, et al. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protocols Mol Biol*, 2015, 109: 21–29
- 9 Boyle A P, Guinney J, Crawford G E, et al. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, 2008, 24: 2537–2538
- 10 Kuan P F, Huebert D, Gasch A, et al. A non-homogeneous hidden-state model on first order differences for automatic detection of nucleosome positions. *Stat Appl Genet Mol Biol*, 2009, 8: 1–45
- 11 Kim T H, Dekker J. ChIP-seq. *Cold Spring Harb Protoc*, 2018, 2018(5)
- 12 Lin J S, Lai E M. Protein-protein interactions: co-immunoprecipitation. In: Journet L, Cascales E, eds. *Bacterial Protein Secretion Systems. Methods in Molecular Biology*. New York: Humana Press, 2017. 211–219
- 13 Chereji R V, Bryson T D, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol*, 2019, 20: 198
- 14 Eagen K P. Principles of chromosome architecture revealed by Hi-C. *Trends Biochem Sci*, 2018, 43: 469–478
- 15 Dey S S, Kester L, Spanjaard B, et al. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol*, 2015, 33: 285–289
- 16 Macaulay I C, Teng M J, Haerty W, et al. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc*, 2016, 11: 2081–2103
- 17 Chen S, Lake B B, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*, 2019, 37: 1452–1457
- 18 Clark S J, Argelaguet R, Kapourani C A, et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun*, 2018, 9: 781
- 19 Hendrickson D G, Soifer I, Wranik B J, et al. Simultaneous profiling of DNA accessibility and gene expression dynamics with ATAC-seq and RNA-seq. In: von Stechow L, Santos Delgado A, eds. *Computational Cell Biology. Methods in Molecular Biology*. New York: Humana Press, 2018. 317–333
- 20 Zhu C, Yu M, Huang H, et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat Struct Mol Biol*, 2019, 26: 1063–1070
- 21 Mateo L J, Murphy S E, Hafner A, et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 2019, 568: 49–54
- 22 Li G, Liu Y, Zhang Y, et al. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat Methods*, 2019, 16: 991–993
- 23 Han K Y, Kim K T, Joung J G, et al. SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. *Genome Res*, 2018, 28: 75–87
- 24 Rodriguez-Meira A, Buck G, Clark S A, et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol Cell*, 2019, 73: 1292–1305.e8
- 25 Hu Y, An Q, Guo Y, et al. Simultaneous profiling of mRNA transcriptome and DNA methylome from a single cell. In: Proserpio V, ed. *Single Cell Methods. Methods in Molecular Biology*. New York: Humana, 2019. 363–377
- 26 Cheow L F, Courtis E T, Tan Y, et al. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat Methods*, 2016, 13: 833–836
- 27 Angermueller C, Clark S J, Lee H J, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods*, 2016,

- 13: 229–232
- 28 Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*, 2017, 14: 865–868
- 29 Peterson V M, Zhang K X, Kumar N, et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol*, 2017, 35: 936–939
- 30 Moffitt J R, Zhuang X. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods Enzymol*, 2016, 572: 1–49
- 31 Codeluppi S, Borm L E, Zeisel A, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods*, 2018, 15: 932–935
- 32 Eng C H L, Lawson M, Zhu Q, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 2019, 568: 235–239
- 33 Rodriques S G, Stickels R R, Goeva A, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 2019, 363: 1463–1467
- 34 Wang X, Allen W E, Wright M A, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, 2018, 361: eaat5691
- 35 Ma A, McDermaid A, Xu J, et al. Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotech*, 2020, 38: 1007–1022
- 36 Li W, Calder R B, Mar J C, et al. Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat Res*, 2015, 772: 55–62
- 37 Hu Y, An Q, Sheu K, et al. Single cell multi-omics technology: methodology and application. *Front Cell Dev Biol*, 2018, 6: 28
- 38 Cao J, Cusanovich D A, Ramani V, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, 2018, 361: 1380–1385
- 39 Liu L, Liu C, Quintero A, et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat Commun*, 2019, 10: 470
- 40 Frei A P, Bava F A, Zunder E R, et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat Methods*, 2016, 13: 269–275
- 41 Genshaft A S, Li S, Gallant C J, et al. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol*, 2016, 17: 188
- 42 Mimitou E P, Cheng A, Montalbano A, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods*, 2019, 16: 409–412
- 43 Overall S A, Toor J S, Hao S, et al. High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat Commun*, 2020, 11: 1909
- 44 Hou Y, Guo H, Cao C, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res*, 2016, 26: 304–319
- 45 Bian S, Hou Y, Zhou X, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, 2018, 362: 1060–1063
- 46 Giacomello S, Salmén F, Terebienieć B K, et al. Spatially resolved transcriptome profiling in model plant species. *Nat Plants*, 2017, 3: 17061
- 47 Chappell L, Russell A J C, Voet T. Single-cell (multi)omics technologies. *Annu Rev Genom Hum Genet*, 2018, 19: 15–41
- 48 Guo F, Li L, Li J, et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res*, 2017, 27: 967–988
- 49 Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife*, 2017, 6
- 50 Eisenstein M. The secret life of cells. *Nat Methods*, 2020, 17: 7–10
- 51 Jaitin D A, Weiner A, Yofe I, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell*, 2016, 167: 1883–1896.e15
- 52 Adamson B, Norman T M, Jost M, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*, 2016, 167: 1867–1882.e21
- 53 Datlinger P, Rendeiro A F, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*, 2017, 14: 297–301
- 54 Xie S, Duan J, Li B, et al. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell*, 2017, 66: 285–299.e5
- 55 Rubin A J, Parker K R, Satpathy A T, et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell*, 2019, 176: 361–376.e17
- 56 Datta S, Malhotra L, Dickerson R, et al. Laser capture microdissection: Big data from small samples. *Histol Histopathol*, 2015, 30: 1255–1269

- 57 Bageritz J, Raddi G. Single-cell RNA sequencing with Drop-seq. In: Proserpio V, ed. *Single Cell Methods. Methods in Molecular Biology*. New York: Humana, 2019. 73–85
- 58 Klein A M, Macosko E. InDrops and Drop-seq technologies for single-cell sequencing. *Lab Chip*, 2017, 17: 2540–2541
- 59 Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, 2012, 2: 666–673
- 60 Gierahn T M, Wadsworth Ii M H, Hughes T K, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods*, 2017, 14: 395–398
- 61 Han X, Wang R, Zhou Y, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*, 2018, 172: 1091–1107.e17
- 62 Cao J, Packer J S, Ramani V, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, 2017, 357: 661–667
- 63 Rosenberg A B, Roco C M, Muscat R A, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, 2018, 360: 176–182
- 64 Zong C, Lu S, Chapman A R, et al. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, 2012, 338: 1622–1626
- 65 Picelli S, Faridani O R, Björklund Å K, et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*, 2014, 9: 171–181
- 66 Huang L, Ma F, Chapman A, et al. Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu Rev Genom Hum Genet*, 2015, 16: 79–102
- 67 Budnik B, Levy E, Harmange G, et al. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol*, 2018, 19: 161
- 68 Raj A, van den Bogaard P, Rifkin S A, et al. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods*, 2008, 5: 877–879
- 69 Shah S, Lubeck E, Zhou W, et al. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 2016, 92: 342–357
- 70 Wagner D E, Klein A M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet*, 2020, 21: 410–427
- 71 Schier A F. Single-cell biology: beyond the sum of its parts. *Nat Methods*, 2020, 17: 17–20
- 72 Li L, Guo F, Gao Y, et al. Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol*, 2018, 20: 847–858
- 73 Gu C, Liu S, Wu Q, et al. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res*, 2019, 29: 110–123
- 74 Argelaguet R, Clark S J, Mohammed H, et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature*, 2019, 576: 487–491

Advances in single-cell multi-omics methods and their applications in developmental biology

CHENG Xin, YAN Rui & GUO Fan

Ministry of Education Key Laboratory of Birth Defects and Related Diseases of Women and Children, West China Second Hospital, West China Medical Center, Sichuan University, Chengdu 610041, China

Single-cell multi-omics technologies are developed based on single-cell sequencing methods and high-throughput omics methods, which can simultaneously integrate two or more kinds of omics information such as transcriptome, genome, epigenome, proteome or spatial states at single-cell resolution. Compared with those single-cell sequencing methods, the single-cell multi-omics technologies can measure the heterogeneities among single cells at different molecular layers and from different approaches. Moreover, the single-cell multi-omics technologies can decipher the direct and indirect relationships between each omics, thus depicting the state and fate of individual cell more comprehensively and systematically. Without any doubt, single-cell multi-omics technologies have broad applications in life science and medicine. Here, we summarized the current single-cell multi-omics technologies, predicted their development trends, and discussed their applications and potential in studying developmental biology.

single-cell sequencing, single-cell multi-omics, developmental biology

doi: [10.1360/SSV-2020-0244](https://doi.org/10.1360/SSV-2020-0244)

郭帆, 博士, 2014年在中国科学院生物化学与细胞生物研究所获得博士学位, 随后在北京大学BIOPIC中心从事博士后工作, 2017年任四川大学华西医学中心、华西第二医院研究员。主要研究哺乳动物早期胚胎发育与生殖细胞发生过程中基因表达的表观遗传调控机制, 解析了哺乳动物原始生殖细胞表观遗传重编程图谱和受精卵DNA去甲基化的发生机制, 并建立和拓展了单细胞表观遗传多组学测序技术。以第一作者或通讯作者(含共同)在*Cell*, *Nature*, *Cell Stem Cell*, *Nature Cell Biology*, *Cell Research*和*Nature Communications*等国际知名学术期刊发表多篇研究论文。