



## 单细胞多组学在灵长类早期胚胎发育中的研究应用\*

潘志辉<sup>1,2)</sup> 郭荣荣<sup>1)</sup> 张又月<sup>1)</sup> 谭 韬<sup>1,2)\*\*</sup>

<sup>1)</sup> 昆明理工大学灵长类转化医学研究院, 省部共建非人灵长类生物医学国家重点实验室, 昆明 650500;

<sup>2)</sup> 西南联合研究生院, 昆明 650032)

**摘要** 啮齿类是研究胚胎发育的重要模型, 然而啮齿类与灵长类在发育过程中存在巨大的种属差异, 啮齿类中的研究并不能简单推广到灵长类中, 因此灵长类生殖发育研究亟待开展。早期胚胎细胞数量稀少且复杂多样, 通常采用单细胞测序技术来揭示其发育过程及调控机制。然而仅依靠单细胞单组学分析并不能有效全面地解析细胞间编码的复杂网络信息。单细胞多组学有效地联合转录组、表观遗传组、蛋白质组、代谢组等进行分析, 使研究者能够在同一细胞水平上系统地解码不同细胞类型的异质性及发育轨迹, 从多维度去理解灵长类早期胚胎发育的关键事件。本文总结了单细胞RNA 治疗发展到多组学过程中关键技术的发展历程, 并概述了多组学技术在深入解读灵长类早期胚胎发育研究中的应用及未来潜在的发展方向和挑战。

**关键词** 单细胞测序, 多组学技术, 灵长类, 早期胚胎发育  
**中图分类号** R321, Q811.4

**DOI:** 10.16476/j.pibb.2024.0036

人体由在空间和时间上受到精确调控的200多种错综复杂的细胞类型形成, 整个发育分为胚前发育、胚胎发育、胎儿发育和胎后发育。简言之, 胚胎发育包括受精、卵裂、囊胚形成、原肠运动、胚层分化及器官原基发育等阶段。精子和卵细胞结合形成受精卵, 通过卵裂, 经历了2-细胞、4-细胞、8-细胞、桑葚胚 (morula) 之后发育至囊胚 (blastocyst), 该阶段称之为着床前胚胎 (pre-implantation) 发育。从桑葚胚转变为中空泡状的早期囊胚阶段过程中, 细胞的形态和分布发生显著的变化, 并且发生第一次细胞命运决定, 形成了内细胞团 (inner cell mass, ICM) 和滋养层细胞 (trophoblast, TE)。

随后囊胚从透明带中孵化出来, 在第7天左右植入子宫内膜着床 (implantation), ICM增殖分化为上胚层 (epiblast, EPI)、下胚层 (hypoblast, primitive endoderm, PrE) 完成第二次细胞命运决定, 上胚层构成羊膜腔的底, 而下胚层细胞增生构成卵黄囊的顶, 此时上下胚层形成圆盘状的双胚盘

结构。第14天双胚盘中的EPI后端开始产生原条 (primitive streak) 并通过原肠运动 (gastrulation) 形成内、中、外三个更为复杂的胚层, 该阶段称为围床期。第28天三胚层细胞开始形成各种组织器官原基并最终发育成整个胎儿 (fetus)<sup>[1]</sup>, 而TE发育成胎盘连接母体为胚胎发育提供支持 (图1)。早期胚胎发育的关键时期与许多先天性疾病、着床失败和孕龄妇女流产等密切相关, 因此, 研究早期胚胎发育对于预防人类出生缺陷、先天畸形等和揭示多种发育源性疾病机制具有重要意义。

灵长类胚胎体外培养技术为研究胚胎早期发育提供有益的参考。2016年Shahbazi等<sup>[2]</sup>和Deglincerti等<sup>[3]</sup>首次在体外成功培养人胚胎长达13 d, 2020年Xiang等<sup>[4]</sup>揭示三维培养人类原肠运动前胚胎发育。但是由于国际社会对于人胚胎研究

\* 国家自然科学基金 (GZ2021GB001A) 资助项目。

\*\* 通讯联系人。

Tel: 0871-65946625, E-mail: tant@lpbr.cn

收稿日期: 2024-01-26, 接受日期: 2024-04-01

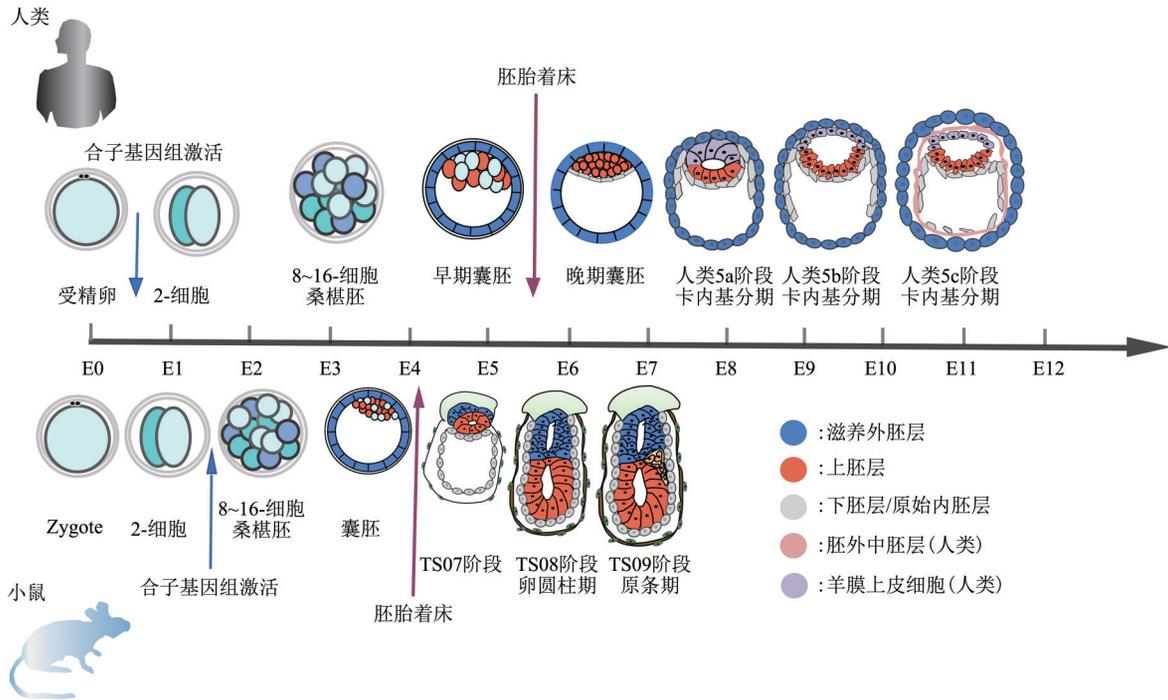


Fig. 1 Differently schematic diagram of early mammalian embryos in the developmental patterns

图1 哺乳动物早期胚胎发育模式的差异

存在“14天规则”伦理限制且本身的稀缺性,使得E14之后的人类早期胚胎发生原肠运动等关键事件尚未得到全面解析。非人灵长类胚胎在早期胚胎发育方面与人类高度相似,是研究人类早期胚胎发育的理想模型。目前研究者通过不断优化猴胚培养体系,在体外培养猴胚胎至E20的基础上<sup>[5-6]</sup>,实现食蟹猴胚胎体外25 d的长时程培养,解析了灵长类胚胎中晚期原肠运动和早期器官的发生<sup>[7-8]</sup>。

然而,由于早期胚胎具有总体细胞数稀少且细胞间异质性高的特点,如果上述问题采用常规的群体性基因表达分析,不仅依赖数量巨大的组织或细胞,而且所分析的结果是群体平均后的状态,掩盖了真实细胞间的差异,无法揭示其中的复杂生物学问题。随着科技的进步,2009年首次出现的单细胞测序(single-cell sequencing)技术在后来的十几年迅速发展。单细胞测序是指在单细胞分辨率上对细胞的转录组、蛋白质组、代谢组、表观遗传修饰等进行测序和解析。它通过绘制细胞发育图谱,锁定目标细胞以进行深入分析,明确细胞的分子调控模式和状态变化,是解析胚胎发育中不同细胞间互动网络的强有力工具<sup>[9]</sup>,在胚胎发育、肿瘤、微生物、脑科学、免疫学等领域广泛应用。

2013年《自然·方法》(*Nature Methods*)杂

志将单细胞测序技术评选为年度技术<sup>[10]</sup>。2018年三篇《科学》(*Science*)长文报道了研究者分别运用单细胞转录组技术建立了斑马鱼和蛙从受精后到整个胚胎形成组织和器官系统的构建过程及基因表达动态路线图,破解了脊椎动物早期胚胎发育之谜,是理解发育生物学基本问题——“细胞命运如何决定”的里程碑<sup>[11-13]</sup>。正是由于单细胞测序技术在低等生物的早期胚胎发育领域带来的新发现,也使得该技术的应用从简单的鱼类、蛙等延伸到了更为复杂的哺乳类。哺乳类动物相较于低等脊椎动物更为复杂,即使是同属于哺乳类,小鼠与人在早期胚胎发育中,某些细胞类型的分化也因不同物种而异(图1),因此用单细胞组学技术解析高等灵长类动物胚胎发育是必要的。

但是由于灵长类动物复杂的遗传调控网络,仅依靠单细胞转录组并不足以揭示灵长类动物的遗传发育复杂性,因此科研人员在原有单细胞转录组的基础上,开发了可在同一细胞水平同时开展转录组、表观遗传、蛋白质组等从不同层面对单细胞进行测序分析的单细胞多组学技术,例如单核多组学测序,可在同一个细胞中同时检测RNA表达水平和染色质的开放性,同步刻画一个细胞的多个维度,这为精准医学的发展提供了更多可能。因此,

该技术在生命科学研究中具有广泛的应用前景。本文总结了单细胞转录组学、表观遗传组学等研究技术发展单细胞多组学的主要过程,以及应用多组学技术揭示灵长类动物早期胚胎发育过程中的关键事件。

## 1 单细胞组学到单细胞多组学的飞跃

### 1.1 单细胞组学与前沿技术的结合

#### 1.1.1 单细胞转录组学

单细胞技术发展之初主要是单细胞 RNA 测序 (single-cell RNA-sequencing, scRNA-seq), 因为信使 RNA (messenger ribonucleic acid, mRNA) 是基因表达调控的一个重要媒介, 决定着细胞的表型和命运。该技术能在一次实验中对数千个至数万个单细胞在 mRNA 水平上进行高通量分子检测, 探索单个细胞全基因表达谱<sup>[14]</sup>。该技术不仅能剖析转录组异质性, 揭示未知的细胞类型或状态, 而且能追踪细胞的发育轨迹。

经典的单细胞 RNA 测序是基于 SMART (Switching Mechanism at 5' End of RNA Template) 技术<sup>[15]</sup>, 其对目标 RNA 进行反转录、模板置换、PCR 扩增、建库、测序, 得到全长转录本信息。但该技术不稳定, 低水平表达的转录本可能丢失。在此基础上, Picelli 等<sup>[16]</sup>对其进行改良, 结合使用锁核酸 (LNA)、更高浓度的 MgCl<sub>2</sub> 以及甜菜碱, 解决了某些 RNA 空间位阻等影响, 开发出能大大提高 cDNA 产值的 SMART-Seq2 技术, 该技术是目前获取单细胞全长转录本的主要手段。

上述传统的 scRNA-seq 技术多采用显微操作对单细胞进行人工分选, 操作难度大, 耗时长, 细胞通量低, 限制了 SMART 技术的应用。基于高通量的单细胞分离技术的出现为这一情况带来转机。例如, 液滴分离单个细胞的 Drop-seq<sup>[17]</sup>, 该技术适用于对大批量细胞进行研究, 从而可以在不同的群体中识别更为少有的细胞类型。2017 年 10X Genomics 公司推出的 Chromium™ Single Cell 3' Solution 系统通过微流控芯片的应用, 将带有条形码 (barcode, 表示细胞身份信息的图形标识符)、UMI (Unique Molecular Index) 的凝胶微珠 (gel beads)、oligo-dT、反转录酶与细胞混合形成 GEM (gel in emulsion) 微滴, 在微滴中进行逆转录反应, 经高通量测序后进行数据分析<sup>[18]</sup>。该技术可在 18 min 内实现 1 000~10 000 个细胞的捕获与大规模单细胞分离, 具有比 Drop-seq 或基于液滴的单细

胞 RNA 测序 (indexing droplets, inDrop) 技术噪音小、稳定性好、细胞捕获率高的特点, 但具有一定局限性, 如只获得 3' 端转录本信息, 对较难分离为单细胞的样本 (如植物) 应用较差。最近, Li 等<sup>[19]</sup>开发了一种基于 i7 96 孔板的二次细胞标记法的新型 5' scRNA-seq 技术 FIPRESCI, 具有简单、有效、低成本的优势, 可以实现  $1 \times 10^5$  个超高通量的单细胞 5' 端转录本信息的捕获。

在单细胞 RNA 测序的基础上, 科学家发现, 将 scRNA-seq 与其他技术如基因编辑技术相结合, 有望实现细胞历史分子事件的记录。Chan 等<sup>[20]</sup>将 CRISPR-Cas9 技术与 scRNA-seq 结合组成“分子记录器”, 能同时记录细胞谱系的发生以及细胞具体的状态, 打破了进一步回答更高阶细胞组织的时空调控等问题技术瓶颈。由于 CRISPR-Cas9 技术可以通过不同基因扰动对细胞进行分选和测序, Replogle 等<sup>[21]</sup>优化了 CRISPRi sgRNA 库, 建立了可以对多个时间点以及不同细胞类型的 Perturb-seq 筛选技术, 可以同时读取 sgRNA 遗传扰动和转录组, 因此能同时描绘细胞表型和基因功能。

但是, 上述的单细胞 RNA 测序技术通常需要在裂解杀死细胞的情况下观测目标细胞的基因表达, 所以无法观测活细胞的变化轨迹。Chen 等<sup>[22]</sup>开发了基于流体超精密显微镜 (FluidFM) 的活细胞转录组测序技术 Live-seq, 通过对单细胞进行微创提取和转移微量细胞质, 结合超低起始量 mRNA 扩增技术, 实现对活细胞进行实时锁定和跟踪成像, 在此过程中, 细胞的状态并未发生改变。但该技术依赖于高精度的体外显微操作等, 因此, 实现在体内研究胚胎发育仍需探索。

#### 1.1.2 空间组学

scRNA-seq 可以在单细胞水平阐明细胞间的异质性, 推进了人们对细胞多样性和胚胎发育轨迹的了解。但该技术体外消化组织进行单细胞分离本身会丢失空间定位数据, 无法将单细胞信息与空间位置进行匹配, 而且细胞内的分子活动特别是转录本在长时间离体环境下可能也会发生一些细微的改变。而胚胎或组织的空间结构在正常发育、内稳态的调控以及生理病理中起重要的重用。因此保留样本的空间信息成为测序领域新的挑战。

Sladitschek 等<sup>[23]</sup>将单细胞转录组学和 4D 光片成像技术结合, 结合独创的计算框架 MorphoSeq, 绘制了脊索动物胚胎发育中单细胞基因表达 4D 图谱, 重构胚胎细胞谱系树, 独特地构建一个数字化

虚拟胚胎, 揭示了其从2-细胞到原肠胚形成的动态过程。为了进一步揭开哺乳动物早期胚胎发育过程, Srivatsan等<sup>[24]</sup>建立了保留单细胞分辨率的同时在更大的尺度上解决空间异质性的方法 (sci-Space), 获得了大约12万个小鼠胚胎细胞核的近似空间坐标和整个转录本信息, 揭开了完整胚胎单细胞水平的空间转录组学信息。这些技术为捕获样本空间信息、构建早期胚胎发育单细胞图谱提供了新的思路。

2021年《自然·方法》(*Nature Methods*)杂志将空间转录组学评为年度技术。目前空间转录组学技术主要有基于激光捕获的显微切割的LCM-seq<sup>[25]</sup>、Geo-seq<sup>[26]</sup>, 基于微阵列原位捕获的ST<sup>[27]</sup>、Slide-seq<sup>[28]</sup>、改良版的Slide-seq V2<sup>[29]</sup>, 基于微流控原位杂交技术, 可同时在多个视野中保持单细胞分辨率甚至亚细胞分辨率进行观测的split-FISH<sup>[30]</sup>。虽然这些方法可以在一定的空间尺度上实现, 但Slide-seq信号的边界与细胞的边界并没有自然对应关系, 以及seqFISH需要图像采集时间较长等从而导致空间信息获取不完全。Alon等<sup>[31]</sup>建立了可以用于非靶向以及靶向纳米级别分辨率的完整组织中RNA的原位测序技术Expansion sequencing (ExSeq技术), 更是将RNA测序技术和成像技术分辨率进行了双重提升。最近, Russell等<sup>[32]</sup>开发了密集的空间索引DNA条码磁珠, 建立了新的空间转录组技术Slide-tags, 可以在空间、单细胞分辨率下揭开不同细胞基因表达的空间变化, 几乎适用于所有的单细胞技术, 为不同样本空间转录组分析提供了通用的平台。随着空间转录组学技术的成熟, 单细胞脂质组学、单细胞代谢组学也在朝着空间层面发展。Yuan等<sup>[33]</sup>开发出单细胞空间代谢组学分析新方法——空间单核代谢组学 (spatial single nuclear metabolomics, SEAM) 实现了组织原位代谢异质性分析、系统解析组织中单细胞的“代谢指纹图谱”。此外, Liu等<sup>[34]</sup>利用一种可逆杂交链式反应 (HCR) 与DNA交换技术结合, 加上计算机辅助设计 (CAD) 技术, 开发单细胞空间蛋白质组学成像技术——CAD-HCR, 可用于亚细胞上的精确空间定位和目标蛋白质的半定量分析。如果能同时进行单细胞转录组与单细胞蛋白质组联合分析, 也是未来非常关键的发展方向。

从单细胞转录组学再到空间转录组学, 不同技术的交叉结合不断推动科学问题的解决, 同时又不

可能并不与转录组密切相关, 使得仅靠转录组信息并不能准确地辨识细胞状态, 仍需要在表观遗传、代谢组等其他组学做同步验证。

### 1.1.3 单细胞表观遗传组学

基因是遗传物质的基本单元。单细胞全基因组测序技术 (scWGS) 能够系统鉴定单个细胞点突变 (SNV)、拷贝数变异 (CNV) 等基因组中发生的变异来揭示不同细胞间的异质性, 但该技术由于二代测序 (NGS) 平台读长的限制, 难以发现基因中的结构变异。Fan等<sup>[35]</sup>利用具有长读长优势的三代测序平台开发出单细胞基因检测的SMOOTH-seq方法, 通过优化的Tn5转座子, 能够在单个细胞扩增出平均长度约6 kb的片段, 实现基因间结构变异、染色体外的环性DNA (ecDNA) 等多种关键分子事件的检验。2022年, Xie等<sup>[36]</sup>基于该技术, 对人的K562和HG002细胞测序, 并在单细胞水平上完成具有高连续性的人类基因组从头的组装, 可以更加完整和准确识别各种插入事件和复杂的结构变异。

在真核生物中, DNA与组蛋白共同构成核小体, 核小体是形成染色质基本结构单位。染色质开放区域是基因组编码生命信息的窗口, 在染色质开放区域上包含了很多重要的转录因子结合位点和基因调控的基因组元件<sup>[37]</sup>。由于染色质存在DNA酶I (DNase I) 超敏感位点, 易受到外源DNase I的切割, 说明该区域暴露、具有可开放性和调节基因组的功能。Gao等<sup>[38]</sup>采用DNase I超敏感位点测序技术 (DNase I hypersensitive site sequencing, DNase-Seq)<sup>[39]</sup>, 用少量细胞从二细胞时期至囊胚期染色质开放区域来进行表观遗传分析, 首次绘制了人早期胚胎发育过程中染色质开放性调控动态图谱。

随着单细胞测序技术的发展, 人们已经可以在单细胞水平对染色质开放区域进行测序 (single cell assay for transposase-accessible chromatin using sequencing, scATAC-seq), 该技术是通过Tn5转座酶切割染色质开放区并在两端插入测序接头的方式来标记该区域, 从而获得开放染色质的位置、转录因子结合位点、核小体的调控区域以及染色质状态等信息。此类方法成为分析复杂组织样品中不同细胞类型的重要手段<sup>[40]</sup>。然而, 一些scATAC-seq方法, 例如sci-ATAC-seq<sup>[41]</sup>、scTHS-seq<sup>[42]</sup>等技术门槛较高, 商业化的scATAC-seq方法, 例如10X Genomics<sup>[43]</sup>和Bio-Rad<sup>[44]</sup>微流控平台, 价格昂贵, 且建库失败时, 实验者难以确定具体原因。因

此, Xu等<sup>[45]</sup>开发新型基于流式分选和384孔板scATAC-seq方法,即在群体水平对细胞染色质开发区域进行标记,利用流式细胞分选,两轮PCR分别在单细胞染色质开发区域DNA片段两端加入barcode,从而达到多个384孔板同时测序的目的,简单易行、成本低和流程可靠,并且测序具有更好的灵活性。Deng等<sup>[46]</sup>进一步对scATAC-seq在整个基因组上进行空间分辨测序Spatial-ATAC-seq,利用微流控技术将组织进行空间二维编码,并与scATAC-seq结合,是单细胞表观遗传与空间组学结合的又一个新的突破,实现在原位组织上全基因组尺度对染色质可及性进行研究。

在染色体开放区域上,蛋白质-DNA互作是启动基因转录调控的关键,也是启动基因转录的前提。传统的ChIP-seq需要甲醛交联后超声片段化,具有背景信号高和样品起始量高的缺点。CUT&RUN<sup>[47]</sup>是一项颠覆性技术,可以在完整的活细胞或细胞核上进行。细胞经过透膜后,目标抗体和pAG-MNase融合酶进入胞核与目标蛋白质结合。Ca<sup>2+</sup>激活pA/G-MNase对靶向染色质片段切割并释放靶蛋白结合的DNA序列,富集DNA序列进行建库测序及数据分析。Hainer等<sup>[48]</sup>进一步将该技术升级到单细胞水平,开发了超低量CUT&RUN (ultra-low input CUT&RUN)并应用于早期胚胎植入前研究。

早期胚胎发育中有大量的反转座子其中的长末端重复序列元件(long terminal repeats, LTR)活跃转录,但必须受到严格调控。Xu等<sup>[49]</sup>利用CUT&RUN技术,发现组蛋白H3K9me3修饰高度富集在LTR区域,利用小鼠模型发现人类8-细胞特异的DUX4和囊胚特异的ZNF808同源转录因子(Dux和Zfp51)调控时期特异性H3K9me3修饰,确保不同胚胎阶段表达的LTR家族精准沉默反转座子,这对于人类植入前正常胚胎发育时序的精确调控具有重要意义。同样,颀伟课题组<sup>[50]</sup>利用优化的CUT&RUN技术,发现多能性转录因子SOX2在哺乳动物早期胚胎发育中与染色质的互作不是简单的先锋因子(Pioneer)模式,而是存在“Settler”、“Pioneer”和“Pilot”等多种互作模式,而它作为Pioneer主导建立多能性网络是在晚期囊胚才会出现。也为更好地理解早期胚胎先锋转录因子SOX2调控体内发育多能性的建立和转变过程的调控机制。

由于CUT&RUN使用的MNase切割DNA后需

要补平缺口、加接头等步骤,造成稀少DNA样品的损失。Henikoff实验室<sup>[51]</sup>推出了CUT&Tag技术,用Tn5转座酶代替MNase,并且与scATAC-seq和CUT&RUN不同的是,pA/G-Tn5识别抗体后精准靶向切割目标抗体结合的DNA片段的同时在序列两端加上测序接头。经扩增后形成的文库让细胞量从1万个降到了60个甚至单个细胞,更能识别拷贝数变化对转录的影响,具有更高的信噪比和数据重复性等优势。为了进一步对单细胞内的基因表达(scRNA-seq)和组蛋白修饰(CUT&Tag)进行联合分析,Zhu等<sup>[52]</sup>开发的Paired-Tag技术,能够从复杂组织样本中更好地研究其染色体状态与基因表达调控机制,但是丢失了细胞的空间位置信息。

2022年初,Deng等<sup>[53]</sup>利用组织样本原位编码,开发了全新的特定组蛋白修饰的高空间分辨表观遗传测序技术——Spatial-CUT&Tag,能够在空间和全基因组尺度上绘制组织发育和疾病相关的表观遗传机制图谱,实现了全基因层面的空间组蛋白修饰分析。进一步,Lu等<sup>[54]</sup>同样在CUT&Tag的基础上通过Tn5转座酶将T7启动子引入修饰基因的附近后,转录产生的大量RNA被MERFISH探针识别测量,可实现在空间位置上小于1 μm高分辨率和高基因层面分辨率的单细胞原位空间表观遗传学方法Epigenomic MERFISH,特别是为回答细胞异质性和早期胚胎发育等关键生物学机制提供了强有力的空间多组学研究手段。

单细胞表观遗传学技术是除转录组学的另一个主流的研究技术,在基因水平上不断衍生出来的相关技术使研究者在从单细胞原位基因组层面研究染色质的状态和表观遗传信息,因为转录始于染色体开放,很多基因表达的调控元件以及转录因子的结合位点都位于开放区域。如果研究者能进一步将组蛋白修饰或转录因子与转录组、蛋白质组等其他组学结合形成多组学的联用,可以更好地从不同层面共同解析不同细胞间编码的错综复杂信息。

## 1.2 单细胞多维组学的交叉联合

单细胞单组学仅仅从单一层面挖掘细胞的异质性,多组学的迅速发展,极大促进了多学科、多领域的交叉合作,为深入剖析细胞状态和命运决定机制提供了机遇。单细胞多组学技术(single-cell multimodal omics)是指在同一细胞中同步测定基因组、转录组、表观遗传组、蛋白质组的多个层面信息的前沿技术,这为进一步了解其他组学如何影响细胞状态和功能提供了独特的视角<sup>[55]</sup>。例如前

几年开发的 CITE-seq<sup>[56]</sup> 和 RNA 表达及蛋白质测序 (RNA expression and protein sequencing, REAP-seq) 技术<sup>[57]</sup>, 利用寡核苷酸偶联到不同抗体, 可以同步检测同一细胞中相对稀疏的转录本和膜表面蛋白的丰度, 实现更加可靠的细胞类型识别。

在细胞核中基因组的三维染色质构象对细胞的功能至关重要, 如何理解复杂组织细胞中三维基因组 (Hi-C) 与基因表达之间的关系, Liu 等<sup>[58]</sup> 开发了一种新型的单细胞多组学技术 HiRES (simultaneous single-cell Hi-C and RNA-seq) 方法, 能高效检测单细胞染色质三维结构上 RNA 的表达, 实现单细胞水平对转录组和三维基因组的同时检测。但是, 随着三代测序的不断完善, 为了更精准和全面地了解染色质互作情况, Li 等<sup>[59]</sup> 在三维基因组 Hi-C 技术上开发了 scNanoHi-C 技术, 它能够在单细胞水平有效地检测染色质的高阶相互作用, 即可以研究多个顺式调控元件同时与一个基因启动子之间的相互作用关系, 实现在全基因组单细胞水平对增强子-启动子的三维结构直接相互作用捕获, 同时能够揭示染色体外环形 DNA 与线性基因组间的复杂相互作用。

为了实现单细胞中同时进行 scRNA-seq 和 scATAC-seq 测序, Chen 等<sup>[60]</sup> 开发开放染色质接头序列 (adapter) 转换策略的 SNARE-seq 技术, 即将 ATAC-seq 插入基因组 adapter 转换成可以被 mRNA 测序引物识别的多聚 A 尾, 实现了在每个单细胞中同时进行转录组和基因组开放染色质测序。但是该技术实验步骤多、过程繁琐等限制其广泛使用, Xu 等<sup>[61]</sup> 开发了能够在同一细胞内同时检测染色质可及性以及基因表达的 ISSAAC-seq 技术, 通过原位逆转录时使用特异 TruSeq adapter 逆转录引物, 使得该序列作为标签 (index), 将 RNA-seq 和 ATAC-seq 数据分开, 这一设计可以避免常规方法中细胞的多次混合与分离过程, 因此无需再次用 index 拆分, 简化实验流程。在单细胞分选上, 该技术可根据实际需求选择孔板的低通量研究, 或通过微流控装置进行高通量实验。

在多组学同时测序上, Clark 等<sup>[62]</sup> 开发了一种染色质可及性、表观遗传和转录组分析的单细胞方法——scNMT-seq, 并应用于小鼠胚胎干细胞的分化, 发现三种组学之间分子层面的联系, 并揭示分化过程中表观遗传之间的动态变化。Stoeckius 等<sup>[63]</sup> 和 McGinnis 等<sup>[64]</sup> 开发的 CellHashing 技术和 MULTI-seq 技术, 分别通过使用通用抗体可以对多

样品进行标记和合成特定脂修饰的核酸序列, 均实现多样本单细胞 RNA 同时测序。但是都需要对细胞进行额外的预处理, 而且并没有实现多样本单细胞多组学同时测序。因此, Wang 等<sup>[65]</sup> 以封面文章报道了一种只用未经任何修饰的核苷酸就能简单且高效实现多样本同时测序的 SNUBar-ATAC 技术, 能实现多样本大规模 scATAC-seq 或 scATAC&RNA-seq 测序。随着测序通量的提高, 该技术可在人类细胞图谱 (HCA) 和癌症图谱网络 (HTAN) 绘制上发挥重大作用。

空间位置信息的保留, 对邻近细胞的相互作用以及其结构组织形式研究十分必要, 除了可以实现空间分辨率表观遗传分析技术 (Spatial-ATAC-seq 和 Spatial-CUT&Tag), 目前在此基础上, Zhang 等<sup>[66]</sup> 基于微流控技术实现了空间分辨的表观遗传组和转录组联合分析的测序技术 (Spatial ATAC-RNA-seq 和 Spatial CUT&Tag-RNA-seq 技术), 能够在空间和基因水平上研究表观遗传在组织中影响细胞基因表达、分化和基因调控等, 这工作也与近期彭广敦课题组<sup>[67]</sup> 发表的一种空间多组学技术 MISAR-seq 相近, 在组织细胞保留空间的前提下实现细胞内 ATAC 和 RNA 双组学信息的捕获。这些技术代表了空间多组学中获得信息最为丰富的工具之一。

细胞身份的鉴定不仅需要转录组和表观遗传上的信息, 更需要蛋白质表达正确。Zeng 等<sup>[68]</sup> 开发了一种三元探针方法的空间翻译组测序技术 RIBOmap, 只有与夹板 DNA 探针 (splint) 相邻的锁式探针才能被环化并生成 NDA 扩增, 实现了有目标性核糖体结合的 mRNA 才能检测, 可用于完整细胞或组织 RNA 表达和亚细胞的点位翻译的调控。在表观遗传学与蛋白质组学联合上, Mimitou 等<sup>[69]</sup> 开发和验证 ASAP-seq (ATAC with Select Antigen Profiling) 能够同时分析单细胞染色质可及性和蛋白质水平的方法。在此基础上, Fiskin 等<sup>[70]</sup> 开发了一种基于 M13 噬菌体的单细胞多模态测序方法 PHAGE-ATAC, 通过纳米抗体识别抗原表位来量化蛋白质, 可以对染色质可及性和蛋白质进行量化分析, 同时捕获线粒体 DNA (mtDNA) 用作天然克隆谱系示踪剂。更重要的是, Smibert 团队<sup>[69]</sup> 又进一步实现四模态共测量, 能够从同一细胞中多模态分析染色质可及性、基因表达和蛋白质, 实现中心法则的全解析, 真正实现单细胞水平多组学分析, 称为 DOGMA-seq。

单细胞双组学 scATAC&RNA-seq 到多样品单细胞组学测序, 再到实现遗传中心原则的真正全解析的飞跃 (图2), 单细胞多组学迎来广阔的发展

前景, 在早期胚胎发育研究领域也带来了众多新的发现和见解。

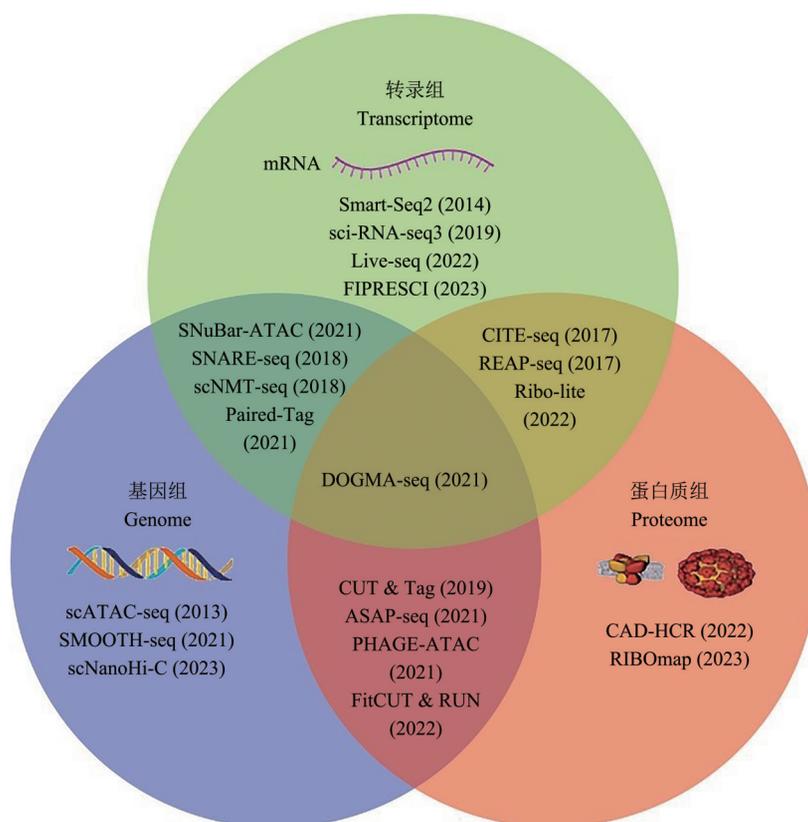


Fig. 2 The main single-cell multi-omics technology at present

图2 目前主要的单细胞多组学技术

## 2 单细胞多组学在灵长类早期胚胎发育研究的应用

### 2.1 单细胞多组学与植入前胚胎发育

受精卵是新生命的起源。受精卵在植入前发育至囊胚, 主要由位于外部的TE和位于内部的ICM细胞组成, 完成第一次细胞命运的决定。植入前胚胎发育的关键事件有卵子向胚胎的转变 (oocyte-to-embryo transition, OET)、合子基因组激活 (zygotic genome activation, ZGA)、X染色体调控、表观遗传重编程等, 本文将依据这些关键事件进行概述。

Xiong等<sup>[71]</sup>在核糖体保护片段的翻译组测序技术 (Ribo-seq) 基础上, 实现在少量细胞水平的Ribo-lite技术, 研究哺乳动物OET和着床前胚胎翻译组动态的变化, 发现在ZGA后转录和翻译的偶

联性加强, OET下调基因在卵子时期高翻译并且有长poly(A)尾巴, 随着减数分裂的恢复, poly(A)被缩短, 翻译也随之抑制, 这与去腺苷酸化因子在卵子翻译组中发挥的功能具有相关性。Zou等<sup>[72]</sup>在Ribo-lite技术基础上进一步联合Smart-seq2开发的Ribo-RNA-lite (R2-lite) 方法可以进行翻译组与转录组联合测序, 揭示人和鼠不同物种的OET过程及早期胚胎中翻译水平动态变化保守性与差异性, 并鉴定出一组同源异型盒 (HB) 转录因子TPRX1/2/L是人类ZGA的关键调控因子, 绘制了人类早期胚胎OET过程中的翻译图谱。

ZGA发生作为生命起始时的第一次转录和启动胚胎发育, 是发育生物学中重要的科学问题之一<sup>[73]</sup>。Wu等<sup>[74]</sup>利用改进版的miniATAC-seq技术实现了在极少量细胞 (最低可只用到20个) 水平上进行开放染色质区域的检测, 论证了在ZGA中

表观遗传信息在人类和小鼠中是保守的, 绘制了人类着床前胚胎发育过程中染色质局部开放区域的全景图。Liu 等<sup>[75]</sup>采用DNA与蛋白质在全基因组相互作用的 Stacc-seq 方法, 揭示了小鼠早期胚胎中 RNA 聚合酶 II 通过“三步走”的模式参与实现基因组激活的过程, 但对其启动的具体时间和特征仍知之甚少。过去 ZGA 被认为在受精后小鼠的 2-细胞晚期, 人在受精后 3 d 左右的 8-细胞阶段发生<sup>[76-77]</sup>。但 2022 年 Asami 等<sup>[78]</sup>的研究推翻了这个说法, 通过高分辨率 scRNA-seq 对人的中期 II 卵母细胞和双核单细胞胚胎的基因表达进行深度分析, 研究者认为基因首次能产生规范剪接的 mRNA 转录本是在人类胚胎受精不久后的单细胞阶段, 即 1-细胞晚期, 颠覆了人们对人源胚胎 ZGA 的认知。此外, 转录因子在调控 ZGA 机制中具有重要的作用, 但目前研究转录因子结合位点的技术大多数要依赖于高质量的特异性抗体, Wang 等<sup>[79]</sup>通过引入转录因子-免疫球蛋白 rFc 的融合表达, 建立了不依赖于特异性抗体的绘制转录因子结合位点图谱的方法 FitCUT&RUN, 这技术比传统的 ChIP-seq 和 CUT&RUN 具有更高的灵敏度和特异性, 他们发现在 ZGA 之前 Naong 在基因调控区的结合与部分早期基因的激活密切相关。

转录抑制因子 SPEN 与长非编码 RNA XIST 的相互作用在促进 X 染色体失活 (X chromosome inactivation, XCI) 中发挥核心作用。Dossin 等<sup>[80]</sup>通过敲除 SPEN 后进行 scRNA-seq, 发现敲除 SPEN 并不会影响 X 失活特异性转录本 (X inactive specific transcript, XIST) 的定位, SPEN 通过顺式作用元件以及染色体重塑因子等与 XIST 之间相互联系介导 X 染色体失活。Okamoto 等<sup>[81]</sup>通过 RNA-seq&FISH 技术对食蟹猴胚胎 XIST 以及 X 染色体连锁基因的表达进行测定, 发现父本和母本的 X 染色体均会表达 XIST, 并招募表观遗传标记 H3K27me3 和 H2AK119 泛素化信号, 但这些修饰不足以建立 XCI, 可能是 XCI 建立的前奏。以上从不同方面探讨了单细胞测序技术在 X 染色体剂量平衡的机制运用。

自 2009 年汤富酬研究组第一篇文章报道 scRNA-seq 后, 该团队在基因组水平上开发了单细胞 DNA 甲基化高通量测序技术, 发现植入前基因组范围的 DNA 甲基化重编程是高度有序的大规模去甲基化和局部再甲基化动态平衡的结果, 父源基因组 DNA 去甲基化速度快于母源基因组, 呈不平

衡分布, 描绘了人类着床前胚胎发育的 DNA 甲基化组图谱<sup>[82]</sup>。同年, 他们利用开发的高精度单细胞多组学测序技术——scCOOL-seq<sup>[83]</sup>和利用整倍体胚胎的单细胞数据发现, 不同于小鼠胚胎, 人类植入前胚胎具有更为开放、松散的染色质结构, 在发育过程中父源基因组染色质快速打开并且其开放程度迅速超过母源基因组, 更为精准地描绘了人类着床前胚胎发育过程中 DNA 甲基化组和染色质状态组的重编程过程以及二者之间的相互关系等关键问题, 实现了对同一细胞的染色质倍性和开放状态、核小体定位、DNA 甲基化及拷贝数变异等多达 5 个层面的基因组和表观遗传进行分析。以上各项研究极大丰富了人们对于人类着床前胚胎基因表达和表观遗传调控规律的认识。

## 2.2 单细胞多组学与植入后胚胎发育

由于技术和伦理的限制, 小鼠作为最常见模式生物被广泛应用于早期胚胎植入后的研究。Cao 等<sup>[84]</sup>利用研发的超高通量 sci-RNA-seq3 技术, 同时检测了来自 61 个不同胚胎发育时期超过 200 万个细胞转录组测序, 并系统地绘制了小鼠胚胎 E9.5~E13.5 的主要器官形成的细胞分化路径以及每个路径中几万个基因的表达变化。同期, Pijuan-Sala 等<sup>[85]</sup>应用 10X Genomics 对受精后小鼠胚胎 E6.5~E8.5 中原肠胚形成的 48 h 内不同时间点 10 万个单细胞的转录组进行分析, 描绘细胞相互作用的分子图谱。这两项研究系统地解释了细胞是怎样从最初的三个胚层分化成组成生物体各个系统的几百种细胞类型以及内在基因调控机制, 对研究早期出生缺陷和基因变异具有重要的意义。

在形态上, 小鼠和灵长类植入前胚胎发育过程大体相似, 然而植入后的发育过程以及发育速度却有很大的区别 (图 1)。因此, 在小鼠模型上获得的结论难以直接在人及非人灵长类模型上应用, 研究着床后胚胎发育最好能直接在灵长类胚胎上进行。但是研究者无法连续观察同一胚胎在体内不同发育时期具体的变化, 因此体外胚胎培养体系是了解着床期胚胎发育的主要途径。Zhou 等<sup>[86]</sup>结合体外模拟人类着床策略和单细胞多组学测序技术 scTrio-seq2, 证实胚胎在囊胚阶段后期具备体外自我生长能力, 并且利用单细胞转录组和 DNA 甲基化组图谱重构人类胚胎从植入到植入后的动态变化过程, 发现植入后胚胎内甲基化水平显著高于胚外, 说明不同谱系细胞的表观遗传调控过程呈现不同步的明显差异, 基因表达调控网络和 DNA 甲基

化可能共同协调决定植入后囊胚阶段的细胞谱系命运决定。

Xiang等<sup>[4]</sup>建立了人胚胎体外3D培养系统,首次将人囊胚培养至14d(原条阶段,但没有神经胚的出现,符合国际伦理),高度模拟了人胚胎前14d的发育过程。进一步地,作者通过大量单细胞转录组测序分析揭示了着床后胚内和胚外细胞谱系分化和分子调控网络,发现具有多能性的上胚层细胞着床后经历从原始(naïve)态特异*KLF4*、*KLF5*、*TFCP2L1*等基因失活,到始发(primed)态转变是由不同的因子决定的,并维持转录本到第14天,绘制了原肠前期胚胎的分子和形态发育全景图。

为了深入探索非人灵长类胚胎着床后发育过程中的分子机制,Niu等<sup>[5]</sup>和Ma等<sup>[6]</sup>同时结合胚胎体外培养系统实现了食蟹猴胚胎体外连续培养20d,成功模拟了胚胎着床后二胚盘形成、原肠运动、三胚层形成、原始生殖细胞(primordial germ cell, PGC)特化等关键发育事件。并利用scRNA-seq和scATAC-seq发现体外培养的胚胎与体内发育的高度一致,揭示了EPI、TE、PrE、PGC的基因表达特征及发育轨迹,探究了不同谱系细胞特化的调控机制及原肠运动时期复杂的分子互作关系。

原肠运动是早期胚胎发育最关键的生物学阶段之一,深入了解这段时间胚胎的发育及其调控机理对多种出生缺陷和发育源性疾病的了解极为重要。Bergmann等<sup>[87]</sup>通过激光显微切割的空间三维转录组学和胚胎干细胞模型,发现猕猴原肠运动时期前内脏内胚层的出现是保守的转录因子HHEX、LEFTY2、LHX1和灵长类特有因子POSTN、SDC4、FZD5的共同参与。羊膜特化响应BMP信号,通过ID1/2/3在胚盘的边界发生,描绘了早期灵长类动物妊娠第二周子宫内原肠运动的空间图谱。同样基于非人灵长类,Cui等<sup>[88]</sup>采用Geo-seq空间转录组技术,建立了E17~E21五个发育节点3D胚胎模型,首次构建了食蟹猴原肠运动时期精度的空间转录组图谱,展示了灵长类在原肠运动时期胚层发生的关键调控网络,揭示BMP、Hippo和IGF信号通路对于灵长类的原肠运动起始具有重要作用,并跨物种分析FoxO家族在鼠、食蟹猴、人等不同物种的原肠运动时期存在物种特异的转录调控网络。

为进一步解析原肠运动中晚期和胚胎器官发生的核心事件,一方面通过体内研究,Tyser等<sup>[89]</sup>有

幸获得一枚宝贵的在体胚胎,首次揭示人CS7时期(第16~19天)原肠胚的细胞群体分子特征,发现PGC和幼红血细胞,但没有早期神经元诱导的SOX3、PAX6和神经元分化TUBB3、NEUROD1等相关标记物存在,所以这将可能影响人体胚胎研究领域的“14天规则”。接着,Zeng等<sup>[90]</sup>在人CS7基础上,通过scRNA-seq和空间转录组技术研究了从CS7到CS10(受精后20d)人类原肠胚和早期神经系统发育不同阶段之间细胞类型的发育轨迹、空间分布模式等。Xu等<sup>[91]</sup>利用同样的技术,构建人类胚胎第四周原肠运动后器官发生的初期,即各个成体器官最早的特异的祖先状态转录图谱,对头部中胚层、四肢芽、神经管进行细致的空间图谱研究。对于非人灵长类,Zhai等<sup>[92]</sup>收集了CS8-CS11(E20-E29)时期的在体食蟹猴胚胎,利用10X Genomics技术,揭示了晚期原肠运动和三胚层分化中神经管、体节和肠管等早期器官发生中谱系发生的调控机制和重要的细胞类群的特征,并比较了Hippo信号通路等在灵长类与啮齿类相同发育阶段之间转录组的差异。这些重要的体内研究数据为灵长类胚胎模型和不同组织细胞的体外诱导研究提供了必要的在体数据参考。另一方面,Gong等<sup>[7]</sup>和Zhai等<sup>[8]</sup>在体外培养猴胚至20d基础,开发了一种3D培养系统,可支持食蟹猴体外受精后发育至25d的长时程培养,结合单细胞多组学测序技术(scChARM-seq),对胚胎转录组、DNA甲基化和染色质可及性进行分析,与已报道的在体胚胎特征相似。借助scRNA-seq解析了胚胎早期神经的分化和造血谱系的特化,并且高度模拟了体内卵巢两个波次造血的发生,系统研究了非人灵长类从原肠运动晚期到早期器官发生过程重要的细胞组成和谱系特化轨迹。这些发现都加深了人们对原肠运动时期精细调控的理解,为深入了解灵长类器官发生以及神经系统发育和功能等重要科学问题提供了新的思路和启示。

在PGC的特化方面,Tang等<sup>[93]</sup>利用scRNA-seq、scATAC-seq和超低量染色质免疫沉淀测序,建立hPGC发育的表观遗传学机制,发现形成素刺激的中内胚层(mesendoderm, ME)增强子重塑能瞬时赋予hPGC命运,证明SOX17、TFAP2C和Blimp1作为主要的转录激活因子发挥功能,对于hPGC命运建立至关重要。小鼠和非人灵长类PGC发育路径和信号调控方式有差别<sup>[94]</sup>,在小鼠当中Blimp1激活TFAP2C,并且与PRDM14三者以相互

依赖的方式建立 PGC 的基因表达程序。而在人和非人灵长类中, *Blimp1* 上调来自 *SOX17*, 与 *TFAP2C* 二者以一种相互依赖的方式建立 hPGC 的基因表达。

关于羊膜细胞, Yang 等<sup>[95]</sup> 在不同时间点对胚胎进行单细胞转录组测序分析中发现, 羊膜细胞 *ISL1* 基因敲除使 *BMP4* 表达显著降低而引起中胚层发育缺陷, 羊膜细胞作为发育调节信号源在原肠胚形成中的发挥重要作用。Rostovskaya 等<sup>[96]</sup> 利用体外人类多能干细胞从着床前 naïve 到着床后 primed 的过渡模拟上胚层的发育, 整合人和灵长类动物原肠运动前后的 scRNA-seq 数据集, 发现早期和晚期羊膜细胞是遵循类似 TE 和非神经外胚层样的两个不同时间分离的转录谱系。这些工作发现都推进了人们对于灵长类胚胎着床后羊膜形成机制的认识。

因此, 通过灵活运用新兴发展的各类单细胞多组学测序技术, 深入阐明灵长类胚胎在发育过程中的原肠运动、羊膜起源与形成、PGC 起源、神经胚层发育等关键事件, 必将逐步揭开胚胎发育调控的神秘面纱。

### 3 总结与展望

从单个细胞发育成复杂的生命个体, 胚胎发育是发育生物学中贯穿的核心问题。由于“14 天伦理”规则的限定, 目前研究者只能够在体外将人类囊胚培养至 14 d, 为研究人原肠运动前胚胎发育提供重要理论基础<sup>[4]</sup>。而对于着床后 4 周的研究可以从流产的胎儿中获得。因此, 胚胎从第 2 周到第 4 周中间发生的许多关键事件仍然是一个未解之谜。目前对于人胚胎这个时期的发育研究主要以啮齿类为参照, 但是由于种间差异, 在发育速度上、形态发生和分子调控网络等方面存在种种不同, 例如着床后小鼠胚胎发育成卵圆筒状结构, 而人胚胎发育成双胚盘结构 (图 1)。

因此, 非人灵长类作为人类亲缘关系上的近亲, 与人类基因组同源性高达 95% 以上<sup>[97]</sup>, 是研究胚胎发育中不可或缺的模式。借助单细胞测序技术, 研究者可以更好地解析胚胎发育中目标表型背后复杂分子网络机制。单细胞测序从转录组学开始, 经历了单细胞三维基因组学、蛋白质组学, 到 scATAC&RNA-seq 等双组学, 再到 DNA、RNA、蛋白质水平上多组学的交叉结合, 最终回归于中心法则的全解析, 实现了从单一层面到全面解析细胞的飞跃, 这对于灵长类胚胎发育中稀有细胞群关键

谱系分化图谱的构建有重要意义及广泛应用前景, 使研究者能在单细胞分辨率下揭示胚胎发育中 ZGA 激活、X 染色体剂量平衡效应、羊膜起源与形成、PGC 起源等事件等关键事件过程的具体机制 (表 1)。

然而, 灵长类胚胎发育仍存在一系列尚未解决的关键科学问题, 如胚胎植入后的细胞分化、原肠运动、胚层特化的发育动态和机制、人 PGC 的起源和迁移、神经细胞的发育与分化、早期组织器官的形成机制等。由于伦理限制、样本稀缺、人鼠的种间差异, 以及单细胞转录组学不能完全解析上述问题的发生机制等原因, 因此, 研究者应当整合单细胞多组学的前沿技术应用于灵长类胚胎发育的研究。但是如何选择单细胞测序或整合多模式组学, 本身就是极具挑战性的工作。不同的单细胞多组学测序技术各有优缺点, 如果整个细胞群的异质性预期很小, 或者知道异质性价值不高, 那么单细胞的方法就不是最好的选择, 所以在科研和临床的实际应用中应做到具体问题具体分析。其次, 如何将前沿的单细胞多组学技术真正的广泛运用到科研临床工作中, 从发现到应用如何打通“最后一公里”, 针对目前仍然存在的可重复技术精确性、建库成本问题、专业技术人员的培训与验证等。因此, 不断优化建库技术以及自动化高通量设备的开发、降低成本同时提高准确性, 最重要的是加强不同领域研究者间的合作与数据共享, 多管齐下, 真正实现供应链的国产化。

此外, 随着单细胞多组学技术的不断进步, 生物信息分析也面临诸多的机遇和挑战。在有限的计算资源下, 如何将庞大的多组学数据合并, 并进行有效的整合与分析, 降低样品间的批次效应同时提高数据的信噪比等, 这都需要不断开发相关的新算法。Hao 等<sup>[98]</sup> 开发的利用前沿的“加权最近邻”(WNN) 对多组学数据的有效整合, 实现对细胞类型、功能和行为进行联合定义的分析方法, Turkalj 等<sup>[99]</sup> 开发了基因型分型结合转座酶可及性染色质区域分析方法 (genotyping with the assay for transposase-accessible chromatin, GTAC), 可以实现在单细胞多个位点层面进行基因分型同时进行 ATAC 分析等等。

未来生命组学前沿的空间多组学也将迎来井喷式发展。将细胞的空间位置信息有机整合到日益发展的多组学技术平台中, 从基因的线性调控研究扩展到基因网络调控的研究, 使得研究者有机会在单

个细胞层面以及多重维度下推动精准医学的发展，逐渐打破灵长类胚胎发育生物医学研究的“无人区”，为临床上先天性早期出生缺陷、不孕症、胚

胎流产、胎儿宫内生长受限等疾病的诊疗提供科学有效的理论依据。

**Table 1 The main results of single-cell multiomics applied to primate embryo development**

**表1 单细胞多组学应用于灵长类胚胎发育的主要研究结果**

发育时期	背景	单细胞多组学技术	结论
植入前	ZGA启动的具体时间和特征	高分辨率scRNA-seq	人胚ZGA具体启动在1-细胞晚期 <sup>[78]</sup>
	哺乳类OET过程和启动人类ZGA的关键转录因子	Ribo-RNA-lite (R2-lite)	人和鼠不同物种早期胚胎中翻译组动态变化保守性与差异性 Homeobox转录因子TPRX1/2/L是人类ZGA的关键调控因子 <sup>[72]</sup>
	人类植入前胚胎发育过程中表观基因组多层面的动态变化机制	scCOOL-seq	DNA甲基化组和染色质状态组的重编程的过程和二者间关系 五个层面的基因组和表观遗传的分析 <sup>[83]</sup>
	灵长类动物中的X染色体剂量补偿机制	RNA-seq & FISH	灵长类父本和母本的X染色体均会表达XIST X染色体上组蛋白修饰不足以建立XCI <sup>[81]</sup>
	胚胎着床过程中多维度基因表达谱和调控机制的解析	scTrio-seq2	人胚胎着床过程基因表达调控网络 着床过程DNA甲基化动态变化图谱 <sup>[86]</sup>
植入后	体外培养模拟体内人着床后胚胎发育	SMART-Seq2&10X Genomics	3D胚胎培养系统的建立 原肠前人胚胎的发育全景图谱 <sup>[4]</sup>
	非人灵长类胚胎着床后发育过程中的分子机制	scATAC&scRNA-seq	体外培养20 d胚胎与体内发育高度一致 EPI、TE、PrE、PGC等基因表达及发育轨迹 不同谱系细胞特化的调控机制 <sup>[5-6]</sup>
	灵长类胚胎中晚期原肠运动分子特征	scChARM-seq	3D培养猴胚重现体内胚胎的转录组特征 灵长类早期神经发育过程中神经板弯曲、折叠、神经管闭合等核心事件的研究 <sup>[8]</sup>
	人类E16~E19原肠胚的细胞群体分子特征	SMART-Seq2	人CS7时期具有PGC和幼红细胞 无神经元或中枢神经系统相关细胞存在 <sup>[89]</sup>
	人类三胚层的形成和早期神经系统的发育机制	scRNA-seq&空间转录组	原肠胚阶段从CS7到CS10的细胞类型发育轨迹推断 人类早期神经系统的细胞亚型、分子特征、空间分布以及分化机制 <sup>[90]</sup>
	非人灵长类胚胎植入子宫后第二周发育及其调控机理	三维空间转录组	猕猴前内脏内胚层的出现是保守的转录因子和灵长类特有因子共同参与 灵长类子宫内原肠运动的空间剖析 <sup>[88]</sup>

## 参 考 文 献

- [1] Rossant J, Tam P P L. New insights into early human development: lessons for stem cell derivation and differentiation. *Cell Stem Cell*, 2017, **20**(1): 18-28
- [2] Shahbazi M N, Jedrusik A, Vuoristo S, *et al.* Self-organization of the human embryo in the absence of maternal tissues. *Nat Cell Biol*, 2016, **18**(6): 700-708
- [3] Deglincerti A, Croft G F, Pietila L N, *et al.* Self-organization of the *in vitro* attached human embryo. *Nature*, 2016, **533**(7602): 251-254
- [4] Xiang L, Yin Y, Zheng Y, *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature*, 2020, **577**(7791): 537-542
- [5] Niu Y, Sun N, Li C, *et al.* Dissecting primate early post-implantation development using long-term *in vitro* embryo culture. *Science*, 2019, **366**(6467): eaaw5754
- [6] Ma H, Zhai J, Wan H, *et al.* *In vitro* culture of cynomolgus monkey embryos beyond early gastrulation. *Science*, 2019, **366**(6467): eaax7890
- [7] Gong Y, Bai B, Sun N, *et al.* Ex utero monkey embryogenesis from blastocyst to early organogenesis. *Cell*, 2023, **186**(10): 2092-2110
- [8] Zhai J, Xu Y, Wan H, *et al.* Neurulation of the cynomolgus monkey embryo achieved from 3D blastocyst culture. *Cell*, 2023, **186**(10): 2078-2091

- [9] Eberwine J, Sul J Y, Bartfai T, *et al.* The promise of single-cell sequencing. *Nat Methods*, 2014; **11**(1): 25-27
- [10] No authors listed. Method of the year 2013. *Nat Methods*, 2014, **11**(1): 1
- [11] Briggs J A, Weinreb C, Wagner D E, *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, 2018, **360**(6392): eaar5780
- [12] Wagner D E, Weinreb C, Collins Z M, *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 2018, **360**(6392): 981-987
- [13] Farrell J A, Wang Y, Riesenfeld S J, *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 2018, **360**(6392): eaar3131
- [14] Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*, 2017, **541**(7637): 331-338
- [15] Ramskold D, Luo S, Wang Y C, *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*, 2012, **30**(8): 777-782
- [16] Picelli S, Faridani O R, Björklund A K, *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*, 2014, **9**(1): 171-181
- [17] Macosko E Z, Basu A, Satija R, *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 2015, **161**(5): 1202-1214
- [18] Zheng G X, Terry J M, Belgrader P, *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 2017, **8**: 14049
- [19] Li Y, Huang Z, Zhang Z, *et al.* FIPRESCI: droplet microfluidics based combinatorial indexing for massive-scale 5'-end single-cell RNA sequencing. *Genome Biol*, 2023, **24**(1): 70
- [20] Chan M M, Smith Z D, Grosswendt S, *et al.* Molecular recording of mammalian embryogenesis. *Nature*, 2019, **570**(7759): 77-82
- [21] Replogle J M, Saunders R A, Pogson A N, *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 2022, **185**(14): 2559-2575
- [22] Chen W, Guillaume-Gentil O, Rainer P Y, *et al.* Live-seq enables temporal transcriptomic recording of single cells. *Nature*, 2022, **608**(7924): 733-740
- [23] Sladitschek H L, Fiuza U M, Pavlinic D, *et al.* MorphoSeq: full single-cell transcriptome dynamics up to gastrulation in a chordate. *Cell*, 2020, **181**(4): 922-935
- [24] Srivatsan S R, Regier M C, Barkan E, *et al.* Embryo-scale, single-cell spatial transcriptomics. *Science*, 2021, **373**(6550): 111-117
- [25] Espina V, Wulfkühle J D, Calvert V S, *et al.* Laser-capture microdissection. *Nat Protoc*, 2006, **1**(2): 586-603
- [26] Chen J, Suo S, Tam P P, *et al.* Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat Protoc*, 2017, **12**(3): 566-580
- [27] Ståhl P L, Salmén F, Vickovic S, *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 2016, **353**(6294): 78-82
- [28] 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**(6434): 1463-1467
- [29] Stickels R R, Murray E, Kumar P, *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol*, 2021, **39**(3): 313-319
- [30] Goh J J L, Chou N, Seow W Y, *et al.* Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat Methods*, 2020, **17**(7): 689-693
- [31] Alon S, Goodwin D R, Sinha A, *et al.* Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science*, 2021, **371**(6528): eaax2656
- [32] Russell A J C, Weir J A, Nadaf N M, *et al.* Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature*, 2024, **625**(7993): 101-109
- [33] Yuan Z, Zhou Q, Cai L, *et al.* SEAM is a spatial single nuclear metabolomics method for dissecting tissue microenvironment. *Nat Methods*, 2021, **18**(10): 1223-1232
- [34] Liu X, Mao D, Song Y, *et al.* Computer-aided design of reversible hybridization chain reaction (CAD-HCR) enables multiplexed single-cell spatial proteomics imaging. *Sci Adv*, 2022, **8**(2): eabk0133
- [35] Fan X, Yang C, Li W, *et al.* SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol*, 2021, **22**(1): 195
- [36] Xie H, Li W, Hu Y, *et al.* De novo assembly of human genome at single-cell levels. *Nucleic Acids Res*, 2022, **50**(13): 7479-7492
- [37] Tsompana M, Buck M J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*, 2014, **7**(1): 33
- [38] Gao L, Wu K, Liu Z, *et al.* Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell*, 2018, **173**(1): 248-259
- [39] Jin W, Tang Q, Wan M, *et al.* Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature*, 2015, **528**(7580): 142-146
- [40] Buenrostro J D, Giresi P G, Zaba L C, *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 2013, **10**(12): 1213-1218
- [41] Cusanovich D A, Reddington J P, Garfield D A, *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, 2018, **555**(7697): 538-542
- [42] Lake B B, Chen S, Sos B C, *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*, 2018, **36**(1): 70-80
- [43] Satpathy A T, Granja J M, Yost K E, *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol*, 2019, **37**(8): 925-936
- [44] Lareau C A, Duarte F M, Chew J G, *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol*, 2019, **37**(8): 916-924
- [45] Xu W, Wen Y, Liang Y, *et al.* A plate-based single-cell ATAC-seq

- workflow for fast and robust profiling of chromatin accessibility. *Nat Protoc*, 2021, **16**(8): 4084-4107
- [46] Deng Y, Bartosovic M, Ma S, *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*, 2022, **609**(7926): 375-383
- [47] Skene PJ, Henikoff JG, Henikoff S. Targeted *in situ* genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc*, 2018, **13**(5): 1006-1019
- [48] Hainer S J, Bošković A, McCannell K N, *et al.* Profiling of pluripotency factors in single cells and early embryos. *Cell*, 2019, **177**(5): 1319-1329
- [49] Xu R, Li S, Wu Q, *et al.* Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos. *Cell Stem Cell*, 2022, **29**(7): 1051-1066
- [50] Li L, Lai F, Hu X, *et al.* Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos. *Science*, 2023, **382**(6676): eadi5516
- [51] Kaya-Okur HS, Wu SJ, Codomo CA, *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*, 2019, **10**(1): 1930
- [52] Zhu C, Zhang Y, Li Y E, *et al.* Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat Methods*, 2021, **18**(3): 283-292
- [53] Deng Y, Bartosovic M, Kukanja P, *et al.* Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science*, 2022, **375**(6581): 681-686
- [54] Lu T, Ang C E, Zhuang X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell*, 2022, **185**(23): 4448-4464
- [55] Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. *Nat Methods*, 2020, **17**(1): 11-14
- [56] Stoeckius M, Hafemeister C, Stephenson W, *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*, 2017, **14**(9): 865-868
- [57] Peterson V M, Zhang K X, Kumar N, *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol*, 2017, **35**(10): 936-939
- [58] Liu Z, Chen Y, Xia Q, *et al.* Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science*, 2023, **380**(6649): 1070-1076
- [59] Li W, Lu J, Lu P, *et al.* scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. *Nat Methods*, 2023, **20**(10): 1493-1505
- [60] Chen S, Lake B B, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*, 2019, **37**(12): 1452-1457
- [61] Xu W, Yang W, Zhang Y, *et al.* ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. *Nat Methods*, 2022, **19**(10): 1243-1249
- [62] 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**(1): 781
- [63] Stoeckius M, Zheng S, Houck-Loomis B, *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol*, 2018, **19**(1): 224
- [64] McGinnis C S, Patterson D M, Winkler J, *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods*, 2019, **16**(7): 619-626
- [65] Wang K, Xiao Z, Yan Y, *et al.* Simple oligonucleotide-based multiplexing of single-cell chromatin accessibility. *Mol Cell*, 2021, **81**(20): 4319-4332
- [66] Zhang D, Deng Y, Kukanja P, *et al.* Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature*, 2023, **616**(7955): 113-122
- [67] Jiang F, Zhou X, Qian Y, *et al.* Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nat Methods*, 2023, **20**(7): 1048-1057
- [68] Zeng H, Huang J, Ren J, *et al.* Spatially resolved single-cell transcriptomics at molecular resolution. *Science*, 2023, **380**(6652): eadd3067
- [69] Mimitou E P, Lareau C A, Chen K Y, *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol*, 2021, **39**(10): 1246-1258
- [70] Fiskin E, Lareau C A, Ludwig L S, *et al.* Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat Biotechnol*, 2022, **40**(3): 374-381
- [71] Xiong Z, Xu K, Lin Z, *et al.* Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol*, 2022, **24**(6): 968-980
- [72] Zou Z, Zhang C, Wang Q, *et al.* Transcriptome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science*, 2022, **378**(6615): abo7923
- [73] Schulz K N, Harrison M M. Mechanisms regulating zygotic genome activation. *Nat Rev Genet*, 2019, **20**(4): 221-234
- [74] Wu J, Xu J, Liu B, *et al.* Publisher correction: chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature*, 2018, **557**(7704): 256-260
- [75] Liu B, Xu Q, Wang Q, *et al.* The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature*, 2020, **587**(7832): 139-144
- [76] Jukam D, Shariati S A M, Skotheim J M. Zygotic genome activation in vertebrates. *Dev Cell*, 2017, **42**(4): 316-332
- [77] Braude P, Bolton V, Moore S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature*, 1988, **332**(6163): 459-461
- [78] Asami M, Lam B Y H, Ma M K, *et al.* Human embryonic genome activation initiates at the one-cell stage. *Cell Stem Cell*, 2022, **29**(2): 209-216
- [79] Wang X, Wang W, Wang Y, *et al.* Antibody-free profiling of transcription factor occupancy during early embryogenesis by FitCUT&RUN. *Genome Res*, 2022, **32**(2): 378-388
- [80] Dossin F, Pinheiro I, Żylicz J J, *et al.* SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*,

- 2020, **578**(7795): 455-460
- [81] Okamoto I, Nakamura T, Sasaki K, *et al.* The X chromosome dosage compensation program during the development of cynomolgus monkeys. *Science*, 2021, **374**(6570): eabd8887
- [82] Tang F, Barbacioru C, Wang Y, *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009, **6**(5): 377-382
- [83] Li L, Guo F, Gao Y, *et al.* Single-cell multi-omics sequencing of human early embryos. *Nat Cell Biol*, 2018, **20**(7): 847-858
- [84] Cao J, Spielmann M, Qiu X, *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 2019, **566**(7745): 496-502
- [85] Pijuan-Sala B, Griffiths JA, Guibentif C, *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, 2019, **566**(7745): 490-495
- [86] Zhou F, Wang R, Yuan P, *et al.* Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature*, 2019, **572**(7771): 660-664
- [87] Bergmann S, Penfold CA, Slatery E, *et al.* Spatial profiling of early primate gastrulation in utero. *Nature*, 2022, **609**(7925): 136-143
- [88] Cui G, Feng S, Yan Y, *et al.* Spatial molecular anatomy of germ layers in the gastrulating cynomolgus monkey embryo. *Cell Rep*, 2022, **40**(9): 111285
- [89] Tyser R C V, Mahammadov E, Nakanoh S, *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature*, 2021, **600**(7888): 285-289
- [90] Zeng B, Liu Z, Lu Y, *et al.* The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell*, 2023, **30**(6): 851-866
- [91] Xu Y, Zhang T, Zhou Q, *et al.* A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol*, 2023, **25**(4): 604-615
- [92] Zhai J, Guo J, Wan H, *et al.* Primate gastrulation and early organogenesis at single-cell resolution. *Nature*, 2022, **612**(7921): 732-738
- [93] Tang W W C, Castillo-Venzor A, Gruhn W H, *et al.* Sequential enhancer state remodeling defines human germline competence and specification. *Nat Cell Biol*, 2022, **24**(4): 448-460
- [94] Kojima Y, Sasaki K, Yokobayashi S, *et al.* Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. *Cell Stem Cell*, 2017, **21**(4): 517-532
- [95] Yang R, Goedel A, Kang Y, *et al.* Amnion signals are essential for mesoderm formation in primates. *Nat Commun*, 2021, **12**(1): 5126
- [96] Rostovskaya M, Andrews S, Reik W, *et al.* Amniogenesis occurs in two independent waves in primates. *Cell Stem Cell*, 2022, **29**(5): 744-759
- [97] Gibbs R A, Rogers J, Katze M G, *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 2007, **316**(5822): 222-234
- [98] Hao Y, Hao S, Andersen-Nissen E, *et al.* Integrated analysis of multimodal single-cell data. *Cell*, 2021, **184**(13): 3573-3587
- [99] Turkalj S, Jakobsen NA, Groom A, *et al.* GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. *Cell Stem Cell*, 2023, **30**(5): 722-740

## Early Primate Embryo Development Meets Single-cell Multi-omics\*

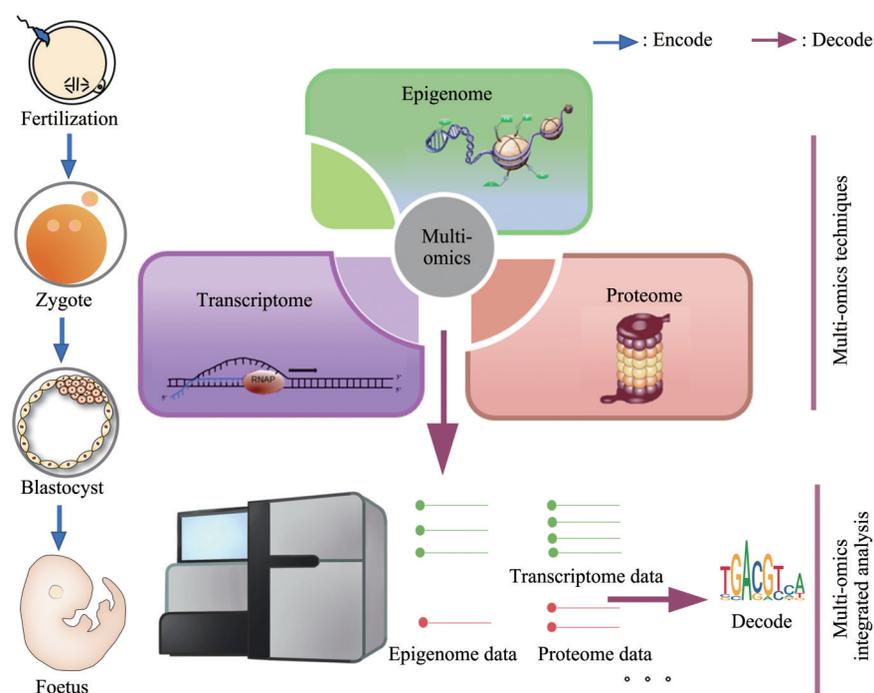
PAN Zhi-Hui<sup>1,2)</sup>, GUO Rong-Rong<sup>1)</sup>, ZHANG You-Yue<sup>1)</sup>, TAN Tao<sup>1,2)\*\*</sup>

<sup>1)</sup>State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine,

Kunming University of Science and Technology, Kunming 650500, China;

<sup>2)</sup>Southwest United Graduate School, Kunming 650032, China)

### Graphical abstract



**Abstract** Given the constraints imposed by the “14-day ethics” rule, numerous critical events occurring between the second and fourth weeks of embryonic development remain poorly understood. This underscores the necessity of a detailed understanding of embryonic development and regulation during this period, which is indispensable for preventing pregnancy failure, treating birth defects, and promoting human reproductive health. Rodents, characterized by their small size, rapid growth, strong reproductive capacity, and fully sequenced genomes, are widely used as crucial models for studying embryonic development. However, the substantial physiological differences between rodents and primates due to evolutionary divergence make it challenging to directly apply findings from rodent studies to primates. Besides, primates, our closest relatives in terms of

\* This work was supported by a grant from The National Natural Science Foundation of China (GZ2021GB001A).

\*\* Corresponding author.

Tel: 86-871-65946625, E-mail: tant@lpbr.cn

Received: January 26, 2024 Accepted: April 1, 2024

evolutionary phylogenetics and physiological characteristics, share more than 95% genetic homology with humans, underscoring the urgent need for primate research. Furthermore, early-stage embryonic cells are both scarce and diverse, making their regulatory mechanisms and developmental pathways typically elucidated through single-cell sequencing. For instance, three significant articles published in *Science* in 2018 mapped the complete atlas of organ and tissue development from fertilization and captured dynamic gene expression profiles in zebrafish and frogs through single-cell transcriptomics. Unfortunately, relying solely on single-cell omics analysis falls short in effectively and comprehensively deciphering the intricate cellular network information. Single-cell multi-omics empower researchers to systematically decode cell heterogeneity and developmental trajectories at the individual cell level by combining transcriptomics, epigenomics, proteomics, and metabolomics analyses. These emerging technologies play a significant role in life sciences, enabling the elucidation of critical early primate embryonic development events from a multi-dimensional perspective, including zygotic genome activation (ZGA), X-chromosome dosage compensation, origins of primordial germ cells (PGCs), mechanisms of cell fate determination, and pivotal events in gastrulation and early organogenesis. This article chronicles the advancement of pivotal technologies, from single-cell histology to multi-omics, beginning with the single-cell transcriptome and culminating in a comprehensive analysis according to the central dogma of molecular biology. It highlights the transition from a singular to a holistic perspective in cellular analysis and reviews the application of multi-omics techniques in unveiling early primate embryonic development. Finally, it delves into the application of multi-omics technologies in enhancing our understanding of early primate embryonic development and explores future possibilities, directions, and challenges in this rapidly evolving field. In doing so, it emphasizes the critical role of interdisciplinary approaches, combining insights from genetics, molecular biology, and bioinformatics to foster innovations in reproductive medicine and developmental biology. The integration of such technologies offers the promise of breakthroughs in understanding complex biological processes, potentially leading to novel therapeutic strategies and advancements in reproductive health and medicine.

**Key words** single-cell sequencing, multi-omics technologies, primates, early embryo development

**DOI:** 10.16476/j.pibb.2024.0036