

DOI: 10.13376/j.cblls/2014035

文章编号: 1004-0374(2014)03-0228-06



文路 博士

本实验室主要围绕哺乳动物早期胚胎发育、研究胚胎干细胞和上胚层干细胞的自我更新能力和多能性调控的分子机理,特别是表观遗传学调控机理,以及相关的原始生殖细胞发育过程中的表观遗传学重编程机理。综合利用单细胞微RNA表达谱分析技术、单细胞RNA-Seq转录组高通量测序技术、单细胞DNA甲基化组高通量测序技术、单细胞基因组高通量测序技术,以及染色体免疫共沉淀-高通量测序技术、小鼠胚胎显微操作技术等分析方法,系统、全面、深入地分析多能性干细胞的基因动态表达网络及其表观遗传学调控机理。

单细胞转录组分析研究进展

文 路*, 汤富酬

(北京大学生命科学学院生物动态光学成像中心, 教育部细胞增殖与分化重点实验室, 北京 100871)

摘 要: 目前常规的转录组分析方法无法揭示单个细胞之间基因表达的异质性,也难以对极少量细胞进行分析,单细胞转录组分析技术为此提供了有效的研究工具。对单细胞转录组分析技术的历史、发展、策略、方法和应用进行综述。

关键词: 单细胞分析; 转录组; 异质性; 高通量测序

中图分类号: Q786 **文献标志码:** A

Recent progresses in single-cell transcriptome analysis

WEN Lu*, TANG Fu-Chou

(Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Biodynamic Optical Imaging Center, College of Life Sciences, Peking University, Beijing 100871, China)

Abstract: Standard transcriptome analysis approaches are not able to uncover gene expression heterogeneity among individual cells, nor can they analyze a small number of cells. Single-cell transcriptome analysis approaches provide powerful tools for these purposes. In this review, we discussed the history, recent progresses, strategies, methods and applications of single-cell transcriptome analysis.

Key words: single-cell analysis; transcriptome; heterogeneity; high-throughput sequencing

细胞的异质性 (heterogeneity) 是一个普遍存在的生物学现象。在多细胞生物个体发育过程中,单个受精卵细胞通过不断地分裂和分化发育成为具有不同形态、表型和功能的多种细胞。组织和器官是多种类型细胞高度有序地组织在一起的复合体,其中细胞种类和异质性最复杂的器官莫过于哺乳动物的脑,人脑的神经元数目可达数百亿,或许每个神经元都有独特的基因表达模式。在疾病发生的情况下,异常的细胞常常混杂于正常细胞之中。肿瘤已

知是一种具有很强的细胞异质性的疾病,恶性肿瘤组织往往由多种基因型和表型的肿瘤细胞组成,高度恶性的细胞可能只占整个肿瘤组织的一小部分。有趣的是,近年来的研究表明,看起来同质的细胞,细胞与细胞之间的基因表达也存在异质性。这种

收稿日期: 2013-12-10

基金项目: 国家自然科学基金项目(31271543)

*通信作者: E-mail: wenlu.wl@gmail.com

异质性可能来源于细胞周期的异质性、细胞微环境的异质性以及基因表达的随机性 (stochastic) 特征。已有证据表明这些异质性都可能对细胞表型产生功能性的作用, 因而已经成为目前研究的一个热点^[1-10]。

转录组 (transcriptome) 广义上是指特定细胞在某一功能状态下所有转录产物的集合, 包括信使 RNA(mRNA)、核糖体 RNA(rRNA)、转运 RNA (tRNA) 以及其他各种非编码 RNA(ncRNA); 狭义上指所有 mRNA 的集合。传统的转录组分析需要数万至数十万个细胞, 这种基于群体水平的方法将大量细胞的基因表达取平均值进行分析, 无法揭示细胞之间基因表达的异质性。另外, 当目标细胞数目非常稀少、很难获得时, 例如胚胎早期发育阶段的细胞和成体干细胞, 需要用到单细胞转录组分析技术。

1 单细胞转录组分析技术的发展

分析单个细胞的转录组首先遇到的问题是极微量的起始样本量。单个哺乳动物细胞的直径为 10 μm 左右, 含有大约 10 pg 总 RNA, 这使得单细胞转录组分析非常困难, 具有极大的挑战性。Brady 等^[11]于 1990 年首次描述了一种基于聚合酶链式反应 (PCR) 的单细胞 cDNA 扩增方法。几乎同时, Eberwine 等^[12]提出了另一种基于体外转录的线性扩增方法。随着基因芯片技术的发展, 2002 年起多篇文章报道了单细胞转录组的基因芯片分析^[13-16]。虽然基因芯片使全面分析单细胞转录组成为可能, 但是它只能检测已知基因, 而且检测的灵敏度和动态范围均有限。2009 年, Tang 等^[17]首次报道单细胞转录组的深度测序分析 (single cell RNA-Seq, scRNA-Seq), 并于 2010 年再次发表相关研究工作^[18]。RNA-Seq 通过对测得的每条序列进行计数获得特定转录本的精确表达量, 因此是一种数码化的表达谱检测, 它能发现新的转录本和新的剪切体, 具有几乎无限的动态范围 (即能够同时检测丰度非常低和非常高的转

录本), 能确定转录本的精确序列, 这些信息都是基因芯片技术无法实现的^[19]。Tang 等^[17]对小鼠的单个卵母细胞进行了 scRNA-Seq 分析, 比基因芯片多检测到了 5 270 个基因以及 1 753 个新的剪切体。近些年来, 又涌现出多种用于单细胞 mRNA 深度测序文库的构建方法, 包括 STRT-Seq (single-cell tagged reverse transcription sequencing)^[20]、Smart-Seq^[21]、Cell-Seq (cell expression by linear amplification and sequencing)^[22] 和 PMA-Seq (Phi29-mRNA amplification and sequencing)^[23] 等, 大大拓展了单细胞转录组分析技术的应用范围。各种分析技术的比较见表 1。

2 单细胞转录组深度测序cDNA文库构建方法

构建 cDNA 文库是单细胞转录组分析的关键技术步骤。典型的哺乳动物细胞包含大约 10 pg 总 RNA, 其中 mRNA 仅有大约 0.1~0.5 pg, 而目前深度测序技术需要纳克级别的 DNA, 因此文库构建过程需要将起始的核酸扩增数十万倍。在文库构建过程中如何避免核酸丢失和如何均匀扩增 cDNA 是构建单细胞 cDNA 文库首要解决的两个关键技术问题。

单细胞 cDNA 文库构建方法包括采集细胞和细胞裂解、mRNA 逆转录成 cDNA、cDNA 扩增和文库构建四个步骤。采集细胞可以采用口吸管、激光捕获显微切割、细胞分选和微流控芯片 (microfluidics) 等方法。用口吸管手动采集单个细胞是最直接的方式, 虽然这种方式比较费时, 但是尤其适合于极稀有的细胞, 例如胚胎早期发育阶段的细胞^[15,17]。激光捕获显微切割技术需要将组织或细胞固定在特定的载玻片上, 然后在倒置显微镜下利用激光脉冲分离靶细胞^[24]。荧光激活细胞分选 (FACS) 和免疫磁珠细胞分选 (MACS) 方法则利用细胞表面标记蛋白或荧光蛋白获得纯的细胞亚群, 并可将单个细胞直接分选到 96 孔板或者 384 孔板的每个样品孔中^[25]。微流控芯片技术可以把单细胞分离、反转录和 cDNA 文库构建等步骤集成到一块毫米尺度的芯片上并同时分析大量单细胞, 极大提

表1 单细胞转录组分析技术

分析技术	扩增方法	检测RNA数目	检测内容	分析细胞数目	文献
scRNA-Seq	PCR	全转录组	全长mRNA	数十个	[17-18, 27]
Smart-Seq	PCR	全转录组	全长mRNA	数个	[21]
STRT-Seq	PCR	全转录组	mRNA 5'端	数十个	[20]
Cell-Seq	体外转录	全转录组	mRNA 3'端	数十个	[22]
PMA-Seq	Phi29聚合酶	全转录组	全长mRNA	数个	[23]
微流控芯片定量PCR	PCR	96个mRNA	mRNA特定序列	数百千	[26, 33]

高了单细胞分析的通量和自动化程度^[26]。

为了避免在构建单细胞 cDNA 文库过程中的核酸丢失,从采集细胞之后到 cDNA 被扩增之前,所有反应步骤最好被整合到一个试管中完成,这些反应的缓冲液是互相兼容的,利用加热灭活前一步反应的酶^[18]。

目前有三种策略来扩增单细胞 cDNA:(1) PCR 指数扩增;(2) 体外转录线性扩增;(3) Phi29 聚合酶扩增。通常来说,单细胞被采集之后,细胞膜在去污剂的作用下溶解,随后 RNA 被释放出来并在反转录酶的作用下合成第一链 cDNA。目前所有的单细胞 cDNA 文库构建方法都采用 oligo(dT) 引物进行反转录获得第一链 cDNA,这样可以避免 rRNA 和 tRNA 的干扰,但是无法检测不带 poly(A) 尾的各种 RNA。

第一种扩增策略需要首先在 cDNA 的两端加上锚定序列,然后利用锚定序列进行 PCR 扩增。转录本 3' 端的锚定序列可以通过 oligo(dT) 引物在反转录的过程中引入。Brady 等^[11]采用了一种简单有效的方法引入转录本 5' 端的锚定序列:首先用末端脱氧核糖核酸转移酶在第一链 cDNA 的 3' 末端加上一连串脱氧腺苷酸(A),然后用一条 3' 端为 oligo(dT) 序列、5' 端为锚定序列的引物合成第二链。这种方法灵敏度非常高,例如 Tang 等^[27]改进了该方法用于 RNA-Seq 分析,能够从单个小鼠胚胎干细胞中检测到 10 800 个基因。

最近报道的 STRT-Seq^[20]和 Smart-Seq^[21]采用另一种方法引入转录本 5' 端的锚定序列。这种方法利用了反转录酶的末端转移酶和模板转换(template switch)活性。当到达 mRNA 的 5' 末端时,反转录酶通过末端转移酶活性可以在第一链 cDNA 的 3' 端加上数个脱氧胞苷酸(C),以一段 3' 端含有鸟苷酸残基(G)及 5' 端含锚定序列的寡核苷酸与胞苷酸退火形成模板,反转录酶可以随后转换模板,以该寡核苷酸为模板继续延伸从而引入转录本 5' 端的锚定序列。这种方法的优势在于能够相对富集 mRNA 的 5' 端序列,但是由于反转录酶的末端转移酶和模板转换活性不能达到 100%^[28],这种方法的灵敏度比 scRNA-Seq 低。虽然从转录本丰度极高的卵母细胞中,Smart-Seq 检测到的基因数目接近 scRNA-Seq^[21],但是从较小的单个人类胚胎干细胞中,scRNA-Seq 可以检测到大约 12 000 个基因,而 Smart-Seq 只能检测到大约 8 000 个基因,说明大量低丰度的基因被丢失^[29]。

第二种策略利用体外转录反应进行扩增,其方法是首先用一条结构为 5'-T7 RNA 聚合酶启动子序列 - oligo(dT)-3' 的引物进行反转录并随后合成第二链 cDNA,这样所合成的第二链 cDNA 可以作为 T7 RNA 聚合酶的模板用于体外转录^[13,22]。体外转录扩增的优势在于线性扩增,能够减少 PCR 指数扩增所造成的偏差,但是缺点在于扩增效率远低于 PCR 扩增,需要进行三轮体外扩增,且扩增的片段较短。最近报道的 Cell-Seq 改进了这一策略,它改用一条结构为 5'-T7 RNA 聚合酶启动子序列 - Illumina 5' 接头序列 - 编码序列 - oligo(dT)-3' 的引物,并在完成第二链 cDNA 合成后,将数十个样本样品混合,共同进行体外转录反应,所扩增的 RNA 随后被片段化并加上 Illumina 3' 端接头序列,最后利用接头序列进行 PCR 扩增完成建库。Cell-Seq 只需要进行一轮体外扩增,而且能够同步分析多个单细胞^[22]。

第三种扩增策略是利用了 Phi29 聚合酶进行扩增。Phi29 聚合酶是一种从 *Bacillus subtilis* 噬菌体 phi29 中克隆出的 DNA 聚合酶,具有链置换和很强的连续合成特性,被广泛用于单细胞基因组扩增^[30]。Kang 等^[31]和 Pan 等^[23]利用该酶分别成功地扩增了单个细菌和单个哺乳动物细胞的转录组。二者的方法均在反转录之后将第一链 cDNA 环化,然后利用 Phi29 聚合酶进行连续的滚环扩增。利用这种策略,Pan 等^[23]可以从单个 K562 细胞中检测到大约 5 000 个转录本。

3 高通量的单细胞分析

在很多情况下,需要同时分析大量单细胞才能对细胞群体有全面认识。因此,如何提高分析通量是单细胞转录组分析的另一个技术问题。深度测序技术为解决这个问题提供了一种独特而有效的策略:条码法(barcoding),即对每个单细胞的转录组加上一段条码序列,随后将它们混合在一个试管中进行建库和测序,通过辨别条码序列将属于不同单细胞的测序数据区分开。条码法将数十到数百个单细胞建库操作简化至一个操作,明显地提高了效率。STRT-Seq 和 Cell-Seq 均采用了这种策略,所不同的是 STRT-Seq 利用模板转换将条码序列加在转录本的 5' 端,而 Cell-Seq 通过 oligo(dT) 引物和体外转录将条码序列加在转录本的 3' 端^[20,22]。

高通量单细胞基因表达分析的另一个值得提及的方法是基于微流控芯片的高通量单细胞定量 PCR 分析平台。斯坦福大学的 Quake 实验室于 2003 年

首次报道了应用微流控技术在纳升级的反应体系中进行基因表达分析的方法^[32]。随后, Spurgeon 等^[33]报道了 48.48 的基因表达分析芯片, 能够同步进行两千多个定量 PCR 反应, 可针对 48 个单细胞的 48 个基因同时进行分析。2011 年, White 等^[26]报道的分析平台更将自动化细胞采集、裂解、反转录和定量 PCR 反应全部整合到一张微流控芯片中。基于微流控芯片的单细胞定量 PCR 分析平台目前虽然只能分析数十个基因的表达, 但其高通量和高自动化的特征具有独特的优势。

4 单细胞转录组分析技术的应用

以下将通过三个例子来说明单细胞转录组分析技术的应用, 这些例子展示了如何利用该技术来分析稀有细胞类型的转录组, 以及揭示单个细胞之间基因表达的异质性。

第一个例子是哺乳动物着床前胚胎的单细胞分辨率转录组分析。哺乳动物着床前胚胎的细胞数目相当稀少, 人类着床前胚胎的细胞则不仅数目稀少而且珍贵, 因此常规的转录组分析技术难以对其进行分析。Kurimoto 等^[15]通过基因芯片技术首次以单细胞分辨率在转录组水平分析了 E3.5 d 小鼠的内细胞团, 结果显示看起来同质的胚胎细胞已经能够在转录组水平分为两群细胞, 分别具有类似原始内胚层和类似上胚层细胞的表达谱。Tang 等^[27]通过所建立的 scRNA-Seq 技术分析了小鼠的内细胞团, 并跟踪分析了内细胞团转变为体外培养的胚胎干细胞的过程, 发现细胞在转变过程中发生了一系列包括基础代谢、microRNA 和可变剪接在内的转录组全局改变。他们还发现在细胞转变过程中, 抑制性的表观遗传调控基因表达上调, 同时激活性的调控基因表达下调, 这种变化可能参与了小鼠胚胎干细胞全能性的维持。2013 年, Yan 等^[29]和 Xue 等^[34]分别报道了人类胚胎从受精卵到着床前的单细胞分辨率转录组分析, 是人类胚胎早期发育阶段转录组的首次全面描述, 为理解人类胚胎早期发育和相关疾病, 以及人类胚胎干细胞全能性的分子机理提供了有用的信息。

第二个例子是树突状细胞被脂多糖激活的单细胞转录组分析。脂多糖是革兰氏阴性细菌细胞壁中的一种成分, 能够通过激活树突状细胞的 Toll 样受体引发显著的转录组改变。最近 Shalek 等^[35]在单细胞水平对这一转录激活过程进行了分析。在脂多糖刺激 4 h 后, 他们采集了 18 个小鼠骨髓来源的树

突状细胞, 利用 Smart-Seq 构建单细胞 cDNA 文库并随后进行测序分析。尽管脂多糖的转录激活在群体细胞水平已经被详细研究, 但在单细胞水平, 他们发现树突状细胞的反应具有相当高的异质性。许多已知的炎症反应基因在部分细胞中被强烈激活, 而在部分细胞中仅被低度激活或完全未被激活, 呈现一种双峰模式。他们进一步发现炎症反应的异质性源于两个方面, 一方面是树突状细胞的成熟状态, 另一方面则是基因调控网络的随机性特征。基于后者, 他们鉴定了一个由 137 个炎症反应基因组成的调控模块。另外, 他们还意外地发现了大量具有细胞异质性的可变剪接形式。

第三个例子是对肿瘤组织的高通量单细胞基因表达分析。传统方法, 如免疫组织化学, 能够显示单个或数个基因在肿瘤组织中表达的细胞异质性, 但是难以对单个肿瘤细胞中的更多基因同时进行检测。最近, Dalerba 等^[36]利用微流控芯片高通量单细胞定量 PCR 方法对正常结肠上皮组织和结肠肿瘤进行了分析。他们通过荧光激活细胞分选 (FACS) 分选出相当于正常结肠隐窝底部的 EpCAM^{hi}/CD44⁺ 细胞, 该群细胞包含了结肠上皮的干细胞。他们选择了 56 个基因进行分析, 包括 53 个结肠上皮组织特异表达的基因、3 个看家基因和 3 个增殖相关基因。从正常结肠上皮组织中, 他们可以从 EpCAM^{hi}/CD44⁺ 细胞群中分辨出 5 类细胞亚群。他们发现, 来自肿瘤组织的 EpCAM^{hi}/CD44⁺ 细胞也能分出类似的细胞亚群。通过将单个 EpCAM^{hi}/CD44⁺ 细胞接种到裸鼠皮下, 他们得到了一个来自单个细胞的移植瘤, 分析显示该移植瘤组织内也包含着类似的细胞亚群。这些结果表明, 至少在部分肿瘤中, 瘤细胞之间基因表达的异质性源自细胞的多谱系分化过程。他们进一步发现这种与分化过程相关的基因表达模式与肿瘤的恶性程度有关, 并提出了新的结肠癌分型方法, 与现有的分型方法相比可以更好地预测患者的预后。

5 单细胞转录组研究展望

单细胞转录组分析技术已取得可喜的进展, 但是仍有很大的提升空间, 包括检测方法的灵敏度和精确性, 分析内容的广泛性, 分析的通量和自动化程度, 以及分析软件的开发等。利用单细胞技术在转录组水平全面揭示细胞异质性的研究工作目前才刚刚开始, 随着技术的进一步完善, 相信单细胞水平的转录组研究将有可能导致新的细胞分类, 发现

细胞转化过程中新的中间类型, 以及全面地揭示单个细胞在发育过程和对环境应激过程中的各种行为。

[参 考 文 献]

- [1] Chambers I, Silva J, Colby D, et al. Nanog safeguards pluripotency and mediates germline development. *Nature*, 2007, 450(7173): 1230-4
- [2] Singh AM, Hamazaki T, Hankowski KE, et al. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells*, 2007, 25(10): 2534-42
- [3] Chang HH, Hemberg M, Barahona M, et al. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 2008, 453(7194): 544-7
- [4] Hayashi K, Lopes SM, Tang F, et al. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*, 2008, 3(4): 391-401
- [5] Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 2008, 135(2): 216-26
- [6] Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*, 2010, 467(7312): 167-73
- [7] Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. *Science*, 2010, 327(5965): 542-5
- [8] Li GW, Xie XS. Central dogma at the single-molecule level in living cells. *Nature*, 2011, 475(7356): 308-15
- [9] Gupta PB, Fillmore CM, Jiang G, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 2011, 146(4): 633-44
- [10] Rocco M, Schmitter D, Knobloch M, et al. Predicting stem cell fate changes by differential cell cycle progression patterns. *Development*, 2013, 140(2): 459-70
- [11] Brady G, Barbara M, Iscove NN. Representative *in vitro* cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol Cell Biol*, 1990, 2: 17-25
- [12] Eberwine J, Yeh H, Miyashiro K, et al. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA*, 1992, 89(7): 3010-4
- [13] Klein CA, Seidl S, Petat-Dutter K, et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol*, 2002, 20(4): 387-92
- [14] Tietjen I, Rihel JM, Cao Y, et al. Single-cell transcriptional analysis of neuronal progenitors. *Neuron*, 2003, 38(2): 161-75
- [15] Kurimoto K, Yabuta Y, Ohinata Y, et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res*, 2006, 34(5): e42
- [16] Hartmann CH, Klein CA. Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Res*, 2006, 34(21): e143
- [17] Tang F, Barbacioru C, Wang Y, et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009, 6: 377-82
- [18] Tang F, Barbacioru C, Nordman E, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc*, 2010, 5(3): 516-35
- [19] Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57-63
- [20] Islam S, Kjällquist U, Moliner A, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res*, 2011, 21(7): 1160-7
- [21] Ramsköld D, Luo S, Wang YC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*, 2012, 30(8): 777-82
- [22] Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, 2012, 2(3): 666-73
- [23] Pan X, Durrett RE, Zhu H, et al. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc Natl Acad Sci USA*, 2013, 110(2): 594-9
- [24] Schütze K, Lahr G. Identification of expressed genes by laser-mediated manipulation of single cells. *Nat Biotechnol*, 1998, 16(8): 737-42
- [25] Galbraith DW, Elumalai R, Gong FC. Integrative flow cytometric and microarray approaches for use in transcriptional profiling. *Methods Mol Biol*, 2004, 263: 259-80
- [26] White AK, VanInsberghe M, Petriv OI, et al. High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci USA*, 2011, 108(34): 13999-4004
- [27] 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. *Cell Stem Cell*, 2010, 6(5): 468-78
- [28] Schmidt WM, Mueller MW. CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res*, 1999, 27(21): e31
- [29] Yan L, Yang M, Guo H, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol*, 2013, 20(9): 1131-9
- [30] Dean FB, Nelson JR, Giesler TL, et al. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*, 2001, 11(6): 1095-9
- [31] Kang Y, Norris MH, Zarzycki-Siek J, et al. Transcript amplification from single bacterium for transcriptome analysis. *Genome Res*, 2011, 21(6): 925-35
- [32] Liu J, Hansen C, Quake SR. Solving the "world-to-chip" interface problem with a microfluidic matrix. *Anal Chem*, 2003, 75(18): 4718-23
- [33] Spurgeon SL, Jones RC, Ramakrishnan R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One*, 2008, 3(2): e1662
- [34] Xue Z, Huang K, Cai C, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, 2013, 500(7464): 593-7
- [35] Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and

splicing in immune cells. *Nature*, 2013, 498(7453): 236-40
[36] Dalerba P, Kalisky T, Sahoo D, et al. Single-cell dissection

of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol*, 2011, 29(12): 1120-7

汤富酬报告讨论

陈润生(中国科学院生物物理研究所)

Q: 我提一个关于技术的问题。你的报告中提到你在做反转录的时候考虑了 poly(A) 尾巴, 因为还有很大一部分的非编码 RNA (noncoding RNA, ncRNA) 是没有 poly(A) 的, 那么你是否所有的反转录都只考虑到 poly(A) ?

A: 是这样的, 目前采用反转录方法研究的干细胞相关 ncRNA 都含有 poly(A) 尾巴。不含有 poly(A) 尾巴的 RNA 鉴定比较难, 因为它们跟 rRNA 很难区分开 (二者都没有 poly(A) 尾巴, 最近我们在发展借助于 cDNA 文库的方法来发现不含有 poly(A) 尾巴的 ncRNA。

鲁志(清华大学生命科学学院)

Q: 从单细胞中提取 RNA 时, 去除 rRNA 的技术是否成熟?

A: 要除去 rRNA 需要 1 μ g 到 10 μ g 的总 RNA 才能操作, 而从单细胞中只能抽提到大约 10 pg 总 RNA, 不能达到去除 rRNA 所需要的量。

付向东(美国加州大学圣地亚哥分校)

Q: 那么从单细胞中提取 RNA 时, 不用去除 rRNA 吗?

A: 如果研究的是含有 poly(A) 的 ncRNA, 就可以不去除 rRNA。但是如果研究的是不含有 poly(A) 的 ncRNA, 技术难度就要更大。目前国际上也没有很成熟的方法和系统在单细胞水平上来区分两者, 我们也在发展此项技术。

鲁志(清华大学生命科学学院)

Q: 既然你们正在发展此项技术, 那么有什么好的方法可以优化呢?

A: 现在我们做的方法是, 将总 RNA 反转录后测序, 根据测序结果来归类, 以区别 ncRNA 与 rRNA。

Q: 你做的是单细胞的研究, 那么就有一个统计上的问题。你在研究中得出一个结论是 ncRNA 在胚胎发育中的差异性很大, 那么你是如何统计得出的这个结论呢? 你就只做一个细胞吗?

A: 我研究中所有的关于单细胞的数据都是同一种细胞至少做 3 个单细胞的 RNA-Seq, 然后计算它们的平均值以反映变化的趋势。

陈润生(中国科学院生物物理研究所)

Q: 有一个更普遍的问题, 研究单细胞和多细胞其实各有各的优势。生物体有多种行为, 当然这是以单细胞为基础的, 在单细胞水平上研究发现的 RNA 或者蛋白质水平的变化很多时候并不表现出整体水平功能的变化, RNA 或者蛋白质水平可能只是生物体内在稳态的起伏。所以如何能找到与生物学功能相关的变化是一个值得认真考虑的问题。

A: 首先同一胚胎阶段的细胞至少会做 3 个单细胞的独立分析, 这样可以比较客观地反映此细胞阶段的 RNA 或者蛋白质水平, 这是横向的平均。在纵向比较不同胚胎阶段的 RNA 或者蛋白质水平时, 我们会选择相差最大的作为研究对象。

刘 晓(清华大学生命科学学院)

Q: 我就接着刚刚陈老师的问题, 提一个关于发育的问题, 发育过程是一种细胞分化成另一种细胞, 也就是两种细胞状态终端的转换, 而做单细胞的时候相当于就是某一个中间状态了, 所以这个如何反映发育的变化?

A: 我们在研究的时候, 会研究每个胚胎阶段的细胞, 从胚胎卵裂中的两细胞、四细胞到八细胞。我们发现在两细胞时, ncRNA 的相关性很高, 达 0.99。而到达八细胞时, 其相关性就降低至 0.96。这就说明到达八细胞胚胎阶段这些差异变化的 ncRNA 可能与发育相关。