

# 单细胞转录组研究进展\*

周超平 李鑫辉\*\*

(上海交通大学系统生物医学教育部重点实验室, 上海 200240)

**摘要** 单细胞转录组分析以单个细胞为特定研究对象, 提取 mRNA 进行逆转录、放大和高通量测序分析, 能揭示该细胞内整体水平的基因表达状态和基因结构信息, 准确反映细胞间的异质性, 深入理解其基因型和表型之间的相互关系, 在发育生物学、基础医学、临床诊断和药物开发等领域都发挥重要作用. 本文主要介绍了单细胞转录组分析的特点和技术发展历史以及常用研究策略和不同技术的优缺点, 并就其面临挑战和未来发展前景进行了讨论, 为该技术的进一步研究与应用提供参考.

**关键词** 单细胞, 转录组, RNA-Seq

**学科分类号** Q7, Q6

**DOI:** 10.3724/SP.J.1206.2012.00597

## 1 单细胞转录组分析的特点

在某一特定阶段, 由细胞转录出来的全部 RNA 称为转录组, 包括 mRNA 和非编码 RNA (non-coding RNA, ncRNA)<sup>[1-3]</sup>. 准确深入地分析转录组, 有助于全面理解细胞基因表达和调控网络的作用. 相同组织细胞的基因型几乎一致, 但亚单位的表达往往不同, 转录组更能反映甲基化和组蛋白等修饰的差异. 不同组织的细胞各自拥有独特的转录组, 研究这些转录组能够从不同的生理学功能和表型角度阐释基因调控网络.

一般转录组实验通常需要成千上万甚至上百万个细胞. 但早期胚胎细胞和干细胞等只能收集到微量细胞, 这就要求实验中利用尽可能少的细胞, 最好是在单细胞水平进行转录组分析. 最新研究表明, 相似细胞的基因表达具有异质性, 这种异质性是由不同的衍生基因组、细胞循环和微环境等决定的<sup>[4]</sup>. 常规方法得到的是大量细胞基因表达信息的共性, 难以显示彼此的异质性. 对单细胞转录组的分析, 不仅能研究基因表达的异质性, 也能研究表达的随机性, 进一步揭示与发育、疾病等相关的一系列重要信息. 因此在单细胞水平进行转录组分析

具有重要的科学意义也是现实的迫切需要.

## 2 单细胞转录组分析的技术发展

1975 年, 英国化学家 Sanger 开创了 DNA 测序方法, 随后 Iscove 创造性地利用 PCR 方法指数性放大 cDNA, James 发明了利用 T7 RNA 聚合酶的体外转录扩增(T7-*in vitro* transcription, T7-IVT)线性放大 cDNA, 人们才便于进行核酸序列方面的基础研究. 在这几项开创性工作的基础上, 微阵列(芯片)技术应运而生, 该技术能够以少量细胞研究成千上万个基因. 芯片技术能检测稳定状态的基因表达, 定位癌细胞中拷贝数的变化, 发现与核酸相互作用的蛋白在基因组上的结合位点, 研究 DNA 间相互作用<sup>[5]</sup>. 时至今日, 已经发展出了单细胞芯片, 可以比较单个病理细胞和正常细胞中基因表达的差异, 预测基因调控的演变过程<sup>[6]</sup>. 不过芯片难以检测可变剪接形成的多个转录本, 其技术的设计

\* 国家重点基础研究发展计划(973)(2010CB529205)和国家自然科学基金重大研究计划培育项目(91229108)资助.

\*\* 通讯联系人.

Tel: 021-34206632, E-mail: xhli@sjtu.edu.cn

收稿日期: 2013-02-22, 接受日期: 2013-04-24

和功能应用多是基于已知基因, 对于未知基因的分析仍有局限.

近年来, 另一项全新的技术——RNA-Seq得到了蓬勃的发展. 最初的转录组研究很大程度上依赖于杂交芯片技术, 自 2005 年起, 以 Roche 公司的 454 技术、Illumina 公司的 Solexa 技术和 ABI 公司的 SOLiD 技术为代表的新一代测序技术(next generation sequencing, NGS)相继诞生<sup>[7-8]</sup>, 这些高通量测序技术应用到由 RNA 逆转录生成的 cDNA 上, 就产生了 RNA-Seq 技术<sup>[9]</sup>. RNA-Seq 技术旨在定量测定和描绘转录组<sup>[10]</sup>, 直接针对 RNA 或者 RNA 逆转录生成的 cDNA 进行测序分析, 比微阵列技术获得的信息更加全面具体. 理论上 RNA-Seq 技术能在单细胞水平上进行, 并且根据不同深度的分析, 最终能获得整体水平全部基因的表达信息<sup>[11]</sup>. 单细胞 RNA-Seq 具有如下优点: a. 分辨率高, 起始样品量为单个细胞; b. 灵敏度高<sup>[12]</sup>; c. 数字化信号, 可以直接检测单个碱基差异、基因家族中相似基因以及可变剪接造成的不同表达<sup>[13]</sup>; d. 检测范围广, 不仅能检测基因, 还能检测 RNA 拼接体<sup>[14]</sup>; e. 重现性好<sup>[15-16]</sup>. 虽然 RNA-Seq 技术仅仅出现了几年的时间, 但已经对基因表达和调控研究作出了卓越的贡献. 当前的 RNA-Seq 技术已经获得了一系列创新性成果: 揭示了基因组表达特性, 转录组的序列信息、复杂结构以及调控机制<sup>[17]</sup>. 可以预见, 随着技术的不断发展, RNA-Seq 技术将会发现更多振奋人心的结果.

### 3 单细胞转录组研究的一般策略

a. 分拣: 挑拣细胞至微量管, 可以手工显微操作, 也可以使用激光捕获显微切割(LCM)<sup>[18]</sup>从培养细胞或者冰冻切片中选择.

b. 纯化: 加入缓冲液使细胞降解, 获得 mRNA, 去除蛋白质和 DNA 等杂质, 纯化 mRNA.

c. 逆转录: 一般使用 oligo-dT 引物和 Superscript II 或 III 逆转录酶, 将 mRNA 逆转录出 cDNA 第一链, 并通过模板转换, 合成 cDNA 的第二条链.

d. 扩增: 对已逆转录生成的 cDNA 进行 PCR 或者 IVT 扩增. PCR 扩增的优点是 cDNA 进行指数性扩增速度很快, 缺点是容易造成引物二聚体和副产物的堆积. IVT 扩增相对于 PCR 扩增的优势是更为严格准确, 但由于 IVT 进行的是线性扩增所以相对于 PCR 过程要慢很多.

e. 分析: 可以选择高通量测序或者微阵列分析. 之前已经对两者进行过比较, 虽然微阵列分析能够较好地提供转录组信息以及调控网络信息, 但是高通量测序能够更准确地从多个角度揭示动态信息.

可以看出, 单细胞转录组研究同普通转录组研究的差异主要在于增加了一步分拣单个细胞的过程, 不过整个过程需要特别注意的是, 单细胞实验的样品量非常少, 所以对实验细节要求很高, 例如所有反应最好在同一个试管中进行, 同时加入的各种缓冲液必须保持前后不冲突<sup>[19]</sup>, 早期加入的酶必须灭活, 以免影响后续实验. 若使用随机引物, 还需要纯化 mRNA, 以免随机引物产生 rRNA/tRNA 的 cDNA, 干扰实验的进行, 扩增时的 dNTP 和引物浓度也需要注意调节. 关键步骤的质控也必不可少<sup>[4]</sup>. PCR 和 IVT, 微阵列和高通量测序各有优缺点, 实验中要注意选择<sup>[14, 20]</sup>.

## 4 不同 RNA-Seq 技术的优缺点比较

### 4.1 单细胞 mRNA-Seq 全转录组分析

该项技术的要点有: a. 使用 oligo-dT 引物逆转录生成 cDNA, 并延长了逆转录的反应时间和 PCR 反应的链延伸时间, 逆转录时间从 5 min 延长至 30 min, PCR 时间从 3 min 延长至 6 min, 尽可能地扩大阅读范围; b. 在引物的 5' 端加上氨基, 避免双链 cDNA 的 5' 端片段与 SOLiD 文库接头相连; c. 将 cDNA 的 PCR 扩增与 SOLiD 系统的 cDNA 结合, 这也是该项技术最重要的特点<sup>[4, 20]</sup>. 具体流程见图 1a.

单细胞 mRNA-Seq 全转录组分析的优点有:

a. 与微阵列技术相比, 能检测出更多的转录组, 灵敏度更高; b. 既能分析同一基因的多个转录本及其对应的蛋白质类型, 也能检测已知基因中新的剪接点; c. 准确度高, 噪音低. 然而, 该项技术也存在以下几点不足: a. 由于在转录过程中使用的是 oligo-dT 引物, 所以该技术不能检测不带 poly(A)尾巴的 mRNA 序列, 例如组蛋白的 mRNA; b. 由于逆转录酶作用时间和功能的限制, 目前大于 3 kb 的 mRNA 难以检测; c. 不能辨别正反义转录组<sup>[20]</sup>.

### 4.2 RNA-Seq 的单细胞标记逆转录技术

RNA-Seq 的单细胞标记逆转录技术(single-cell tagged reverse transcription, STRT)与 4.1 所介绍的方法的主要区别有: a. 逆转录合成出 cDNA 第一

链后并不加上 poly(A)尾, 而是利用 cDNA 第一链生成时 3'端的 CCC 序列, 加入 GGG 序列与之配对, 生成 cDNA 的第二链; b. 利用生物素将 cDNA 固定到磁珠上, 扩增完成之后再加上 poly(A)尾进行末端修复; c. 在生成第一链的 CCC 序列旁, 引入一段 barcode 作为阅读标记, 最后的阅读顺序依次为 barcode、CCC、RNA 序列, 这也是该方法最重要的创新<sup>[21]</sup>. 具体流程见图 1b.

STRT 最显著的优点是能同时在个体和群体的水平上进行分析, 不需要已知标记和细胞类型. 单细胞 RNA 测序结果产生后形成一个双向细胞图谱, 该图谱能反映细胞群体、功能亚群、细胞个体三方面的信息<sup>[21]</sup>. 然而, 该方法最明显的缺点是除了末端由于逆转录产生的 CCC 序列, mRNA 本身的序列之中也存在着许多 CCC 序列, GGG 配对转录之后容易出现误差. 唯一的解决方法是提高引物浓度, 但是高引物浓度又会引起两个问题, 一个是价格昂贵, 另一个是引物浓度高容易出现非特异扩增. 另外, 该过程需要经历多个 PCR 循环, 产物不一定会得到线性放大, 虽然暂时可以用特异分子识别(unique molecule identification, UMIs)来解决, 但是转录效率有待提高<sup>[22]</sup>.

### 4.3 Full-length RNA 分析技术

目前最常用的有 Smart-Seq、PMA 和 SMA 三种技术.

全转录组分析和 STRT 技术存在的共同问题是由于优先扩增 mRNA 的 3'端, 转录后所得的 cDNA 包含不完整的 5'端, Smart-Seq 技术能改善这一问题. Smart-Seq 技术在用 CDS 引物(5' AAG-CAGTGGTATCAACGCAGAGTACT (30)VN 3', V 代表 A、C 或 G)逆转录带 poly(A)尾的 RNA 或者用莫洛尼鼠白血病病毒逆转录酶(moloney murine leukemia virus reverse transcriptase, MMLVRT)逆转录全部 RNA 时, 一旦逆转录进行到 RNA 分子的 5'端, MMLV 的末端转移酶发挥活性在 cDNA 的 3'端加上 C, 继而对所加的 C 碱基进行配对, 产生一个延伸模板. 模板转换之后, 所得的 cDNA 就含有 mRNA 5'端的全部序列和第二条 cDNA 链合成时的锚定序列, 之后再行扩增<sup>[23]</sup>. 具体流程见图 1c.

PMA(Phi29-mRNA amplification)是来源于全基因组扩增技术, 利用 Phi29 DNA 聚合酶进行放大的一种方法. Phi29 DNA 聚合酶具有特殊的链置换和连续合成特性, 但通常要求模板长度大于 3~4 kb 时才能进行高效放大. 为了解决这一问题, 在扩增之前, 通常会用环化酶或连接酶将逆转录后的单链或双链 cDNA 环化, 并且只有 5'和 3'端都完整的 cDNA 才能发生环化, 环化后的 5'和 3'端仍保留了初始的转录方向, 便于在测序时进行分析<sup>[24]</sup>.

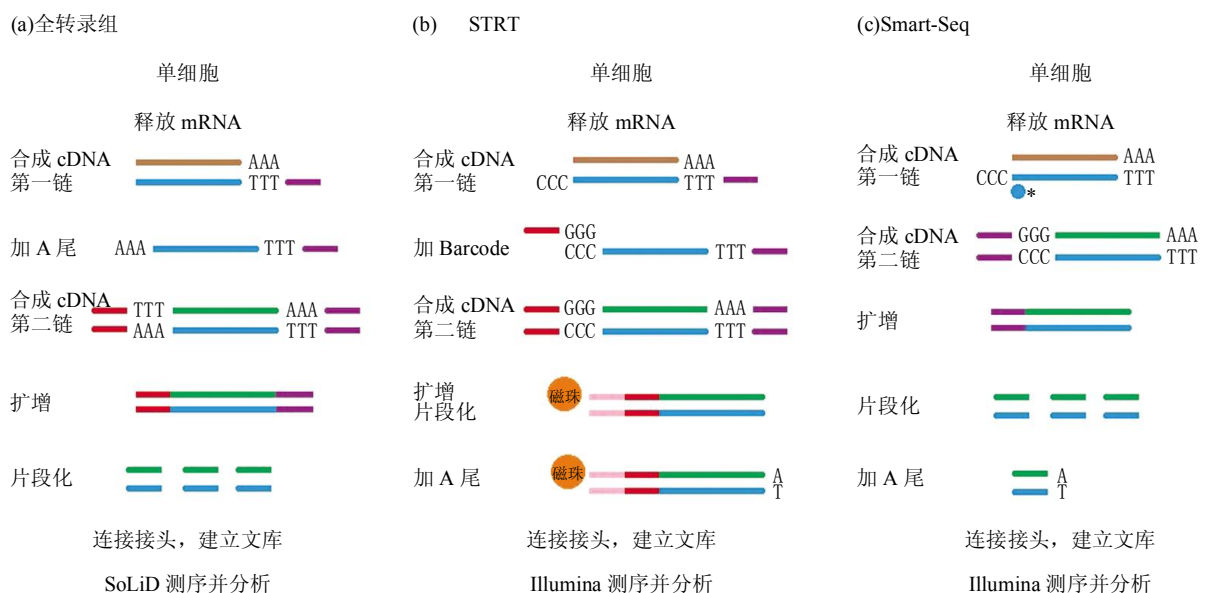


Fig. 1 Overview of three RNA-Seq methods for single cell transcriptome

图 1 三种单细胞转录组研究方法示意图

全转录组(a)、STRT(b)和 Smart-Seq(c)技术的流程示意图. \* 表示用 CDS 引物或莫洛尼鼠白血病病毒逆转录酶对 mRNA 进行逆转录时反应能进行至 mRNA 5'端, 所得到的 cDNA 包含 mRNA 5'端的全部信息.

SMA (semirandom primed PCR-based mRNA transcriptome amplification)方法利用一段带有发夹结构并在3'端含9个碱基随机序列的寡核苷酸(SMA-p1)作引物, 随机序列可以结合逆转录后的cDNA进行扩增. 获得扩增子后, 识别位点已设计在SMA-p1发夹结构中的Bci VI限制性内切酶会将所有的寡核苷酸接头除去. 该方法利用线性DNA模板和半随机引物, 获得长度相近的产物并用相同引物进行PCR扩增, 避免了cDNA分子两端信息的丢失<sup>[24]</sup>.

Smart-Seq、PMA和SMA的共同优点是扩大

了检测范围, 可以获得mRNA 5'端至3'端完整的全长信息. Smart-Seq的突出优点是可以筛选单核苷酸多态性(SNPs)和变异体, 缺点是稳定性不高<sup>[23]</sup>. PMA在理论上可以扩增出任意长度的完整cDNA, 并且在封闭的微流控体系中能对大量单个细胞进行平行扩增. 但PMA的灵敏度不如SMA高, SMA能检测更多的基因, 结果更接近于未放大的cDNA, 样品起始量低时检测的灵敏度更高, 相较之下更适用于单个细胞的分析<sup>[24]</sup>.

三种技术优缺点比较见表1.

Table 1 Brief comparison of three single cell transcriptome techniques

表1 三种单细胞转录组研究技术的简单比较

	优点	缺点	适用范围	样品准备耗时
全转录组	将检测长度提高至3 kb 准确度高, 噪音小	只能检测含 poly(A) 的 mRNA, 只可获得 3'端小于 3 kb 的片段	囊胚干细胞, 胚胎细胞, 组织细胞	6 天
STRT	可同时分析个体与群体, 不需已知标记和类型	逆转录过程易出现断片不适用任 意剪切的转录组 <sup>[25]</sup>	发育中的细胞, 病理细胞	3 天
Smart-Seq	完整阅读 mRNA 5'端 可用于筛选 SNPs	不稳定, 低水平表达的转录组可 能丢失 <sup>[25]</sup> 价格较高	哺乳动物细胞, 肿瘤细胞	-

## 5 单细胞转录组分析的应用

单细胞转录组能反映基因的差异表达, 有助于研究基因调控网络和基因表达的异质性与随机性, 结合系统生物学的方法, 能应用到肿瘤研究领域. 利用单细胞转录组分析, 能通过分析基因表达调控网络监控人类疾病进程, 持续追踪肿瘤生理学和病理学的动态基因<sup>[4]</sup>.

单细胞转录组分析最基础的应用是研究基因的表达水平, 检测新的基因(包括新的蛋白质编码基因和非编码RNA基因, 尤其是低表达基因)<sup>[7]</sup>. Tang等<sup>[20]</sup>利用全转录组分析的方法, 仅需6天即可对16个鼠囊胚细胞分别建库, 所建文库可以用SOLiD系统进行测序分析, 已经发现1753个未知剪接点, 同时发现8%~19%的基因在同样的卵母细胞中有两种或两种以上不同的表达——证明了单细胞转录组的多样性和复杂性. Saiful等<sup>[21]</sup>利用RNA-Seq单细胞标记逆转录技术研究复杂的组织样本, 产生一个单细胞RNA的表达谱, 聚类形成

表达数据的一个二维图谱. 所得的细胞图谱能反应群体细胞、亚群体细胞和单细胞, 不需要已知标记对细胞进行分类, 就能够发现并分析发育中的细胞和疾病中出现的细胞. Sultan等<sup>[26]</sup>通过研究人类胚胎肾脏细胞和B细胞系, 发现了4096个新的剪接位点, 证明了剪接的最普遍形式是外显子跳跃. Wang等<sup>[27]</sup>利用单细胞全基因组分析来测量个体配子基因组的基因多样性, 他们从91个单细胞得到高密度基因分型并制作出一个重组图谱, 这份重组图谱在低分辨率时与人口范围内的数据是吻合的, 但是在更高分辨率时有着显著差异, 这些数据可以进一步用于寻找基因转换的证据. 单细胞转录组分析还可以直接研究非模式生物而无需了解物种信息, 例如Wang等和Vera等已经分别对白粉虱和蝴蝶的转录组进行了研究<sup>[27-28]</sup>.

单细胞转录组分析的优势在于定量研究基因的突变情况. Tang等<sup>[20]</sup>将含有619个相同基因的野生型作为对照, 发现在Dicer<sup>-/-</sup>和Ago2<sup>-/-</sup>(Eif2c2<sup>-/-</sup>)突变小鼠卵母细胞中, 分别有1696和1553个基

因异常上调. Wang 等<sup>[29]</sup>对 31 个单细胞进行高通量测序, 揭示了大规模基因组的不稳定性. 对 8 个单细胞进行深入分析之后还发现新生突变率具有鲜明的特点. Sugarbaker 等<sup>[30]</sup>利用 mRNA 测序技术研究恶性胸膜间皮瘤(malignant pleural mesothelioma, MPMs)中的 RNA 突变体, 已经发现了 7 个点突变、3 个缺失、4 个沉默表达和 1 个 RNA 编辑位点. Gilean 等<sup>[31]</sup>使用单细胞技术测算遗传变异模型的重组率, 在研究欧洲和非洲人群单核苷酸多态性之后, 他们发现了一个极端的现象, 50% 的重组事件发生在低于 10% 的序列中. 继而说明重组热点普遍存在于人类基因组中, 平均或少于每 200 万个基因出现一次, 但是重组还是优先发生于基因外. Chepelev 等<sup>[32]</sup>利用 RNA-Seq 技术研究人类基因组外显子中的单核苷酸变异体(single nucleotide variations, SNVs), 已经在人类 Jurkat T 细胞和 CD4<sup>+</sup> T 细胞中分别检测到了 12 176 和 10 621 个单核苷酸变异体.

除此之外, 在肿瘤研究方面也取得了重要成果. Daniel 等<sup>[23]</sup>运用 Smart-Seq 技术成功地对单个循环系统肿瘤细胞(circulating tumor cells, CTCs)进行 RNA 分析, 利用各个细胞类型中微量的个体细胞找到了多个基因的不同表达, 成功检测出前列腺淋巴结癌(lymph node carcinoma of prostate, LNCaP)细胞中的大多数活跃基因, 并发现了不同的基因表达模式和生物标记物. Nicholas 等<sup>[33]</sup>结合全基因组扩增和新一代测序方法, 准确地测量出单个核内基因组拷贝的数量, 又用单核测序的方法研究两种人类乳腺癌肿瘤的结构和进化情况. 分析 100 个单基因原发肿瘤和肝转移的单细胞后发现, 是一个单一的克隆形成了原发肿瘤并导致了转移. 分析 100 个多基因肿瘤单细胞后显示连续的克隆扩增可能是由三个特定的克隆亚群引起的. 另外, 在上述两种原发肿瘤中, 还发现了一种基因遗传多样性相当丰富的细胞, 并且该种细胞并不会移动至转移部位. 他们认为肿瘤生长是由克隆扩增和少量持久中间体造成的, 这与以往的逐步发展模型是相反的<sup>[33]</sup>.

除了一般的组织细胞, 单细胞转录组分析已经应用到人脑单个神经细胞中, 分析是否因为神经细胞中基因组的变异导致功能多样化或者神经系统疾病. Gilad 等<sup>[34]</sup>对三个正常个体的 300 个颅内皮质和尾状核神经细胞进行全基因组 L1(long interspersed nuclear elements-1)插入分析, 随后发现每个神经细

胞只有不足 0.6 个 L1 插入, 大多数神经细胞没有检测到 L1 插入, 表明 L1 并不是导致大脑皮质和尾状核神经细胞多样化的主要原因. 此外, 他们还利用单核测序得到了单皮层神经细胞 AKT3 突变的镶嵌特征. Alysson 等<sup>[35]</sup>还发现神经细胞中的 L1 逆转录转座作用受 MeCP2 的调控, 具体表现为, 在啮齿动物中, 如果缺少 MeCP2, 则 L1 转录和逆转录转座作用就增强. 患有 Rett 综合征(rett syndrome, RTT)的病人携带 MeCP2 的突变体, 增加了 L1 逆转录转座的灵敏度, 继而说明了相关疾病的基因突变会影响到 L1 逆转录转座的发生频率. Nicole 等<sup>[36]</sup>从人类胎儿大脑和人类胚胎干细胞得到了海马祖先细胞(hippocampus progenitor cells, NPCs), 证明了 L1 片段体外逆转录的可信度, 同时也表明 L1 逆转录转座有可能发生在人类的大脑中, 也可能促进个体体细胞的嵌合.

## 6 单细胞转录组分析的前景

通过上面的介绍, 不难看出单细胞转录组分析在疾病研究方面的重要作用, 随着技术的不断发展和完善, 单细胞转录组分析可以揭示更多基因表达网络、异质性、随机性表达的信息, 从而尽可能更多更完善地揭示某些疾病的机理, 为人类战胜疾病奠定理论基础. 鉴于单细胞转录组分析在人类攻克疾病方面有如此重要的作用, 美国国立卫生研究所已经宣布在未来五年内投入 9 千万美元到单细胞研究中. Life Technologies 公司甚至设立了百万美元奖金用作攻克人类单个癌细胞全基因组和总 RNA 测序<sup>[37]</sup>. 该技术的研究前景之广, 意义之大, 可见一斑.

单细胞转录组学分析还能作为桥梁, 将基因型、表现型和功能联系起来. 借助基因表达网络、表现型之间的关系, 可以研究细胞谱系, 完善物种起源和进化理论. 另外, 序列转录的方向信息对于转录组注释十分重要, 所以单链测序(single-strand sequencing)和特异链测序(strand-specific sequencing)技术也是单细胞转录组技术发展的一个重要方向<sup>[1, 38-39]</sup>.

除去单细胞转录组学本身的应用, 未来该技术还可以结合其他的技术, 发挥更强大的生命力. 例如结合细胞成像技术和 RNA-Seq 技术, 能更好地揭示基因表达网络如何造成细胞差异以及如何进行动态调控. 单细胞转录组技术发展至少应有三个方面: a. 高通量. 未来将不仅针对单个样品或单个

细胞, 目前 STRT 技术采用 96 孔板并利用 barcode 一次可分析 96 种细胞, 未来也许可以发展至 384 孔板甚至更多, 实现集成化和规模化分析, 大大节约时间和经费. b. 自动化. 目前单细胞转录组技术的操作相对繁琐, 未来应朝着专业化和自动化的方向继续发展, 类似搭载于配套专业设备的微流控芯片, 实现整个实验过程的自动操作, 并使用专门的试剂盒, 降低对实验者操作技术的依赖, 提高效率和稳定性. c. 与上下游技术的无缝衔接. 在简化流程基础上与上游的样品获取技术如 LCM 或流式细胞术, 以及下游高通量测序的建库步骤衔接, 辅以对应的质控步骤, 将能显著提高实验效率. 然而, 单细胞转录组学正面临着几个挑战: 首先是实验中针对群体细胞的标记物并不一定适用于所有的细胞类型; 其次, 细胞与细胞之间转录组的差别甚大<sup>[21]</sup>; 第三, 庞大的数据量增加了信息处理的困难<sup>[40]</sup>; 最后也是最关键的是对于不同丰度的转录组在放大时仍未能解决线性扩增的问题, 即部分低拷贝的基因信息可能会在放大过程中丢失, 这在整体水平的转录组分析过程中将会引起偏差. 对此, 我们正在研发的技术利用生物素标记的 oligo-dT 为引物, 将单细胞 mRNA 逆转录成为 cDNA 并连接到磁珠上以后, 通过一系列在磁珠上进行的定点酶切和连接, 制成约含有 20 个碱基的未知序列标签, 并在两端加上相对较长的固定接头, 依靠多轮 PCR 扩增或者 T7 IVT 方法实现放大. 与本文介绍的其他三种单细胞转录组研究方法相比, 我们获得的是一段固定序列的 DNA 中仅包含约 2% 的未知序列, 因此有望实现线性程度很好的放大, 能减少低拷贝基因转录信息的丢失. 当然其他可供借鉴的线性放大方法还有很多<sup>[41-43]</sup>, 相信未来人们还将继续改善实验技术, 不断推进单细胞转录组的研究向前发展.

总之, 单细胞转录组技术在发育进化和疾病尤其是肿瘤发生发展方面的应用将会取得越来越多的成果, 成为人们攻克疾病, 探索生命科学领域的一项重要技术.

### 参 考 文 献

- [1] 祁云霞, 刘永斌, 荣威恒. 转录组研究新技术: RNA-Seq 及其应用. 遗传, 2011, **33**(11): 1191-1202  
Qi Y X, Liu Y B, Rong W H. Hereditas(Beijing), 2011, **33**(11): 1191-1202
- [2] Costa V, Angelini C, De F I, *et al.* Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol, 2010(2010): 853916
- [3] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 2009, **10**(1): 57-63
- [4] Tang F C, Lao K Q, Surani M A. Development and applications of single cell transcriptome analysis. Nat Methods, 2011 (4 Suppl): S6-11
- [5] Brian T W, Josette R L. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods, **48**(2009): 249-257
- [6] John C M, Christopher E M, Shrikant M M, *et al.* RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 2008, **18**(9): 1509-1517
- [7] 王 曦, 汪小我, 王立坤, 等. 新一代高通量 RNA 测序数据的处理与分析. 生物化学与生物物理进展, 2010, **37**(8): 834-846  
Wang X, Wang X W, Wang L K, *et al.* Prog Biochem Biophys, 2010, **37**(8): 834-846
- [8] Margulies M, Egholm M, Altman W E, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. Nature, 2005, **437**(7057): 376-380
- [9] Fatih O, Patrice M M. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet, 2011, **12**(2): 87-98
- [10] Wang Z, Mark G, Michael S. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 2009, **10**(1): 57-63
- [11] Tang F C, Catalin B, Ellen N, *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc, 2010, **5**(3): 516-535
- [12] Sean C B, Garry P N. From single cells to deep phenotypes in cancer. Nat Biotechnol, 2012, **30**(7): 639-647
- [13] Wilhelm B T, Marguerat S, Watt S, *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 2008, **453**(7199): 1239-1243
- [14] Ali M, Brian A W, Kenneth M, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 2008, **5**(7): 621-628
- [15] Cloonan N, Forrest A R R, Kolle G, *et al.* Stem cell transcriptome profiling *via* massive-scale mRNA sequencing. Nature Methods, 2008, **5**(7): 613-619
- [16] Nagalakshmi U, Wang Z, Waem K, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 2008, **320**(5881): 1344-1349
- [17] Samuel M, Jürg B. RNA-seq: from technology to biology. Cell Mol Life Sci, 2010, **67**(4): 569-579
- [18] Fredrik K, Jessica Z, Luo L, *et al.* Single-cell laser-capture microdissection and RNA amplification. Methods Mol Med, 2004, **99**: 215-223
- [19] Kurimoto K, Yabuta Y, Ohinata Y, *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucl Acid Res, 2006, **34**(5): e42
- [20] Tang F C, Catalin B, Wang Y Z, *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods, 2009, **6**(5): 377-382

- [21] Saiful I, Una K, Annalena M, *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res*, 2011, **21**(7): 1160–1167
- [22] Saiful I, Una K, Annalena M, *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc*, 2012, **7**(5): 813–828
- [23] Daniel R, Luo S J, 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
- [24] Pan X H, Russell E D, Zhu H Y, *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): 594–599
- [25] Jillian J G, Jeffrey M T. Transcriptome sequencing of single cells with Smart-Seq. *Nature Biotechnol*, 2012, **30**(8): 763–765
- [26] Sultan M, Schulz M H, Richard H, *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 2008, **321**(5891): 956–960
- [27] Wang X W, Luan J B, Li J M, *et al.* De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, 2010, **11**(1): 400
- [28] Vera J C, Wheat C W, Fescemyer H W, *et al.* Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol*, 2008, **17**(7): 1636–1647
- [29] Wang J B, Christina F, Barry B, *et al.* Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell*, 2012, **150**(2): 402–412
- [30] Sugarbaker D J, Richards W G, Gordon G J, *et al.* Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc Natl Acad Sci USA*, 2008, **105**(9): 3521–3526
- [31] Gilean A T M, Simon R M, Sarah H, *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science*, 2004, **304**(5670): 581–584
- [32] Chepelev I, Wei G, Tang Q S, *et al.* Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucl Acid Res*, 2009, **37**(16): e106
- [33] Nicholas N, Jude K, Jennifer T, *et al.* Tumour evolution inferred by single-cell sequencing. *Nature*, 2011, **472**(7341): 90–94
- [34] Gilad D E, Cai X Y, Lee E J, *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell*, 2012, **151**(3): 483–496
- [35] Alysson R M, Maria C N M, Nicole G C, *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, 2010, **468**(7322): 443–446
- [36] Nicole G C, José L G, Grace E P, *et al.* L1 retrotransposition in human neural progenitor cells. *Nature*, 2009, **460**(7259): 1127–1131
- [37] Brian O. The single life. *Nature*, 2012, **491**(7422): 27–29
- [38] Croucher N J, Fookes M C, Perkins T T, *et al.* A simple method for directional transcriptome sequencing using Illumina technology. *Nucl Acid Res*, 2009, **37**(22): e148
- [39] Vivancos A P, Güell M, Dohm J C, *et al.* Strand-specific deep sequencing of the transcriptome. *Genome Res*, 2010, **20**(7): 989–999
- [40] Vliet V A. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol Lett*, 2010, **302**(1): 1–7
- [41] Liu C L, Bernstein B E, Schreiber S L. *Whole Genome Amplification: Methods Express*. UK: Scion publishing Ltd, 2005: 77–98
- [42] Molly C. Linear amplification sequencing, a powerful method for sequencing DNA. *Methods*, 1991, **3**(1): 1–70
- [43] Liu C L, Schreiber S L, Bernstein B E. Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics*, 2003, **4**(1): 19–30

## Research Progress in Analyzing Transcriptome at The Level of Single Cells\*

ZHOU Chao-Ping, LI Xin-Hui\*\*

(Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract** Analysis of the transcriptomes of single cells can provide a comprehensive understanding of the heterogeneous expression and gene regulatory networks within a population of cells. Through mRNA isolation, reverse transcription, amplification and deep-sequencing analysis, the transcriptional landscape can be obtained for subsequent detailed quantitative investigation of the levels and variation of gene expression underlying the relationship between the genotype and phenotype of a single cell. Not surprisingly, it has been applied within many fields such as developmental biology, basic medicine, clinical research and drug discovery. This review introduces the history, progress, and technological characteristics of whole transcriptome analyses of single cells, as well as detailed strategies and advantages of the methodology. The technical challenges, potential applications, and future developments are also discussed.

**Key words** single-cell, transcriptome, RNA-Seq

**DOI:** 10.3724/SP.J.1206.2012.00597

---

\* This work was supported by grants from National Basic Research Program of China (2010CB529205) and The Major Research Plan of National Natural Science Foundation of China(91229108).

\*\*Corresponding author.

Tel: 86-21-34206632, E-mail: xhli@sjtu.edu.cn

Received: February 22, 2013 Accepted: April 24, 2013