

基于CRISPR的单细胞转录组测序技术研究进展

郭俊璠¹, 黄行许^{1*}, 乔云波^{2*}

(¹上海科技大学生命科学与技术学院, 上海 201210; ²广州大学生命科学学院, 广州 510006)

摘要: 基于CRISPR的单细胞转录组测序技术是将单细胞转录组测序技术与CRISPR集中筛选技术结合, 用于研究传统技术无法完成的细胞异质性、细胞谱系追踪、细胞互作及空间位置、药物作用靶点预测和鉴定等。本文简单比较了六种最常用的单细胞转录组测序技术, 重点介绍了五种基于CRISPR的单细胞转录组测序技术及应用研究, 并对其推广应用所面临的挑战作出了分析, 旨在为单细胞转录组测序技术的应用及创新提供参考。

关键词: 单细胞转录组测序技术; Perturb-Seq; CRISP-Seq; CROP-Seq; Mosaic-Seq; Direct capture Perturb-Seq

Research progress of CRISPR-based single-cell transcriptome sequencing technology

GUO Junfan¹, HUANG Xingxu^{1*}, QIAO Yunbo^{2*}

(¹School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China;

²School of Life Sciences, Guangzhou University, Guangzhou 510006, China)

Abstract: CRISPR-based single-cell transcriptome sequencing technology is to combine single-cell RNA-sequencing with CRISPR-pooled screens to study cell heterogeneity, cell lineage tracking, cell interaction and spatial location, as well as to predict and identify drug action targets that can not be completed by traditional technologies. We briefly compared the six most commonly used single-cell transcriptome sequencing technologies. Especially, we focused on five CRISPR-based single-cell transcriptome sequencing technologies and their applications and analyzed their potential challenges. We provided useful information for the applications and innovations of CRISPR-based single-cell RNA-sequencing.

Key Words: scRNA-Seq; Perturb-Seq; CRISP-Seq; CROP-Seq; Mosaic-Seq; direct capture Perturb-Seq

细胞是生物结构和功能的基本单位, 其类型和状态差异较大。多细胞生物的大部分生理功能差异反映在组成细胞间基因表达的时空变化上^[1]。最初对基因表达的研究是从大量细胞中获取足够的RNA进行测序分析, 因此测序结果是这些细胞整体的表征。然而, 由于细胞异质性, 细胞的表型相同, 其遗传信息却可能存在显著性差异,

很多低丰度的信息会在整体表征中丢失^[2]。为了弥补传统高通量测序的局限性, 单细胞测序技术应运而生。而决定细胞功能的一个重要因素是转录进程, 因此表征细胞转录水平^[3]、重建细胞谱系^[4]、探究细胞异质性^[5,6]、剖析基因功能关系^[7], 不仅对揭示组织器官正常发育进程有重要意义, 同时可为探究疾病发生发展提供新的认知。近年

收稿日期: 2021-08-25

第一作者: E-mail: guojf@shanghaitech.edu.cn

*通信作者: 黄行许, E-mail: huangxx@shanghaitech.edu.cn; 乔云波, E-mail: ybqiao@gzhu.edu.cn

来, 基因编辑技术的开发及应用呈现爆炸式增长, 由单链向导RNA(single-guide RNA, sgRNA)指导的靶向精确的基因编辑在生物学机制研究^[8]、人类疾病治疗^[9]、作物遗传改良^[10]等方面显示出巨大的应用前景。科学家们将单细胞转录组测序技术(single-cell RNA-sequencing, scRNA-Seq)与成簇的规律性间隔短回文重复序列系统(the clustered regularly interspaced short palindromic repeat, CRISPR)筛选技术结合, 促进了对基因功能和基因调控网络的系统性研究。本文介绍了单细胞转录组测序技术及CRISPR/Cas9基因编辑技术, 重点在转录组层面介绍了几种基于CRISPR的单细胞转录组测序技术及应用, 旨在为转录组测序技术的应用、开发、升级提供参考。

1 单细胞转录组测序技术

2009年, Tang等^[11]开发的单细胞转录组测序技术问世, 生物学研究正式迈入单细胞转录组测序时代。十几年来针对灵敏度、精确度、准确性、通量等问题, 单细胞转录组测序技术不断更新迭代, 细胞通量和基因通量大大提升, 成本逐渐降低, 已成为用于识别和表征细胞类型、状态、谱系和环路的重要研究手段^[12-14]。

单细胞转录组测序技术基本路线是: 细胞分离、细胞标记、基因检测、数据分析。细胞分离最初依靠人工手动分离, 如显微吸取法、激光显微切割法。这种方法工作量大、效率低, 随着技术的提升, 流式细胞术、自动化的集成流体电路、液滴微流控技术等大大提升了细胞分离效率。分离得到的单细胞需要保证拥有各自独立的微反应体系, 如微孔板利用孔间隔保证独立的反应体系、液滴微流控利用油滴包裹隔离单细胞保证独立的反应体系^[15]。细胞标记是利用数个碱基的核苷酸序列作为标签(barcode, BC), 在单细胞微反应体系的逆转录过程中对转录本进行标记, 同一个细胞内的所有转录本标记上相同的BC, 这样在后续建库时既能混合操作, 在数据分析层面又能达到区分单细胞转录数据的目的, 如CEL-Seq(cell expression by linear amplification and sequencing)^[16]等。随着技术的升级, 后来又引入了单细胞转录本独特的分子标识(unique molecular identifier,

UMI)进行标记, 以解决PCR偏好可能对基因表达定量的影响, 使基因定量更加精确, 如CEL-Seq2(cell expression by linear amplification and sequencing 2)^[17]、Drop-Seq^[18]。逆转录-标记之后得到cDNA链, 进一步PCR扩增或体外转录(*in vitro* transcription, IVT)以达到二代测序对样本量的要求。单细胞转录组测序是通过捕获携带多聚腺苷(PolyA)的RNA分子, 并且检测的区域范围具有3'端或5'端倾向性, 对中高表达的基因更为敏感^[19,20]。最后的数据分析可根据研究人员的要求进行, 如根据基因表达特征区分细胞群、探究细胞群差异表达基因、差异表达基因的功能富集分析等。

单细胞测序技术发展迅速, 最常用的主要有六种。2013年, Picelli等^[21]开发了Smart-Seq2, 通过流式细胞仪分选细胞, 进一步裂解, 逆转录模板转换后进行PCR扩增以富集cDNA。随后使用Tn5转座酶将cDNA打断并加上通用测序接头, 再进行一次PCR扩增完成建库。Smart-Seq2能够检测mRNA的全长信息和较多的基因数目, 但是通量低、成本较高, 不适合大规模的单细胞研究。2014年, Jaitin等^[22]开发了MARS-Seq(massively parallel RNA single-cell sequencing)——一种自动化的大规模并行scRNA-Seq, 结合无监督分类算法, 用以分析数千个单细胞的体内转录状态。该技术在分子、细胞、孔板三个水平采用BC, 通过流式细胞仪将单细胞分选到孔板中, 借助BC, 首先对单细胞进行初步分类来表征细胞亚群, 然后通过整合每个无监督类别中数个细胞的数据, 以高分辨率研究转录谱。该技术不仅通过多路复用RNA-Seq对数千个细胞进行体内采样, 同时还保证了对扩增偏差和标记错误的严格控制。同年, Soumillon等^[23]开发了SCRB-Seq(single cell RNA barcoding and sequencing), 相比之前的高通量测序实验流程, SCRБ-Seq实验步骤更简单。SCRБ-Seq通过逆转录模板转换将携带多聚腺苷(polyA)的mRNA逆转录为cDNA, 并给cDNA加上BC、UMI、通用测序接头, 随后对cDNA进行扩增并使其片段化用以测序。2015年, Macosko等^[18]开发了Drop-Seq, 采用微流体装置, 其原理是装置通道中

流动的两种水溶液被油分割成微滴, 一种微滴包含单个细胞, 另一种微滴包含细胞裂解液和带有BC、UMI、PCR handle的微珠, 随后在通道中将两种微滴结合起来, 后续建库步骤与SCR-B-Seq相似。2016年, Hashimshony等^[17]开发了CEL-Seq2, CEL-Seq2借助C1单细胞自动制备系统, C1出口孔装载CEL-Seq2引物, 化学法裂解细胞得到mRNA后在其3'端添加BC、UMI、通用测序接头和T7启动子, 逆转录得到cDNA并热灭活逆转录酶, 随后进行体外转录获得测序文库。相较于之前的CEL-Seq, CEL-Seq2灵敏度更高、成本更低、操作步骤更加简单。2017年, 由10×Genomics公司研发的10×Chromium^[24]微流体平台基于GemCode问世, 其技术核心是一个乳液凝胶珠(gel bead in emulsion, GEM), 在八通道微流控芯片中封装带有BC、UMI、通用测序接头的凝胶珠, 通过微流体装置将单个细胞与凝胶珠封装, 共同形成GEM, 随后凝胶溶解、细胞裂解, 逆转录模板转换mRNA带上标签, 随后破乳, 利用cDNA进行建库。本文就样品文库制备流程, 即单细胞捕获、逆转录、第二条链的合成、扩增、建库和测序范围, 对上述六种scRNA-Seq技术进行比较(图1)。

2 CRISPR/Cas9基因编辑技术

CRISPR是自然界大多数细菌和古细菌抵抗噬菌体感染和质粒转移的防御机制^[25]。根据CRISPR/Cas基因座的分类, CRISPR系统被分为六种不同的

类型(I~VI)^[26,27], 每种类型都使用一组独特的Cas蛋白以及CRISPR RNAs(crRNAs)来进行CRISPR干扰^[28]。在第二类CRISPR系统中反式激活crRNA(tracrRNA), 与crRNA中的重复序列碱基配对, 形成一种独特的双链RNA结构^[29], 再引导DNA核酸内切酶Cas9利用不同的核酸酶结构域(HNH或RuvC)分别切割DNA的两条链^[30]。这种独特的双链RNA指导Cas9切割任何含有20 nt与之互补并且含有前间区序列邻近基序(proto-spacer adjacent motif, PAM)的DNA^[31]。随后人工合成的单链向导RNA(single-guide RNA, sgRNA)模拟tracrRNA-crRNA结构, 使系统得到简化(图2)。这种简化的双组分CRISPR/Cas9蛋白-RNA系统通过sgRNA碱基配对定位到靶DNA序列, 并造成靶DNA双链断裂(double-stranded DNA breaks, DSB)。诱导细胞或通过非同源末端连接(non-homologous end joining, NHEJ)修复DNA, 最终导致碱基的随机插入或缺失(insertions or deletions, indels), 造成移码突变; 或者在同源DNA模板存在的情况下, 通过同源定向修复(homology-directed repair, HDR)来替换切割位点周围的DNA。在应对DSBs过程中, HDR与NHEJ竞争, 通常indels比基因替换更常见^[32](图2)。

自最初报道来自化脓链球菌(*Streptococcus pyogenes*)的Cas9核酸酶(spCas9)具有DNA的靶向切

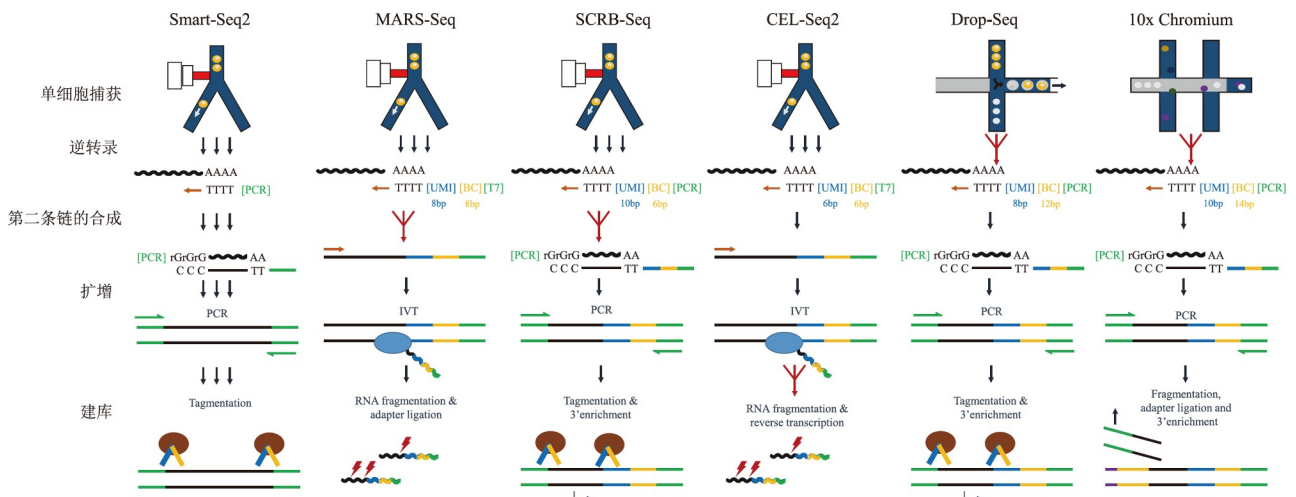


图1 scRNA-Seq文库制备比较

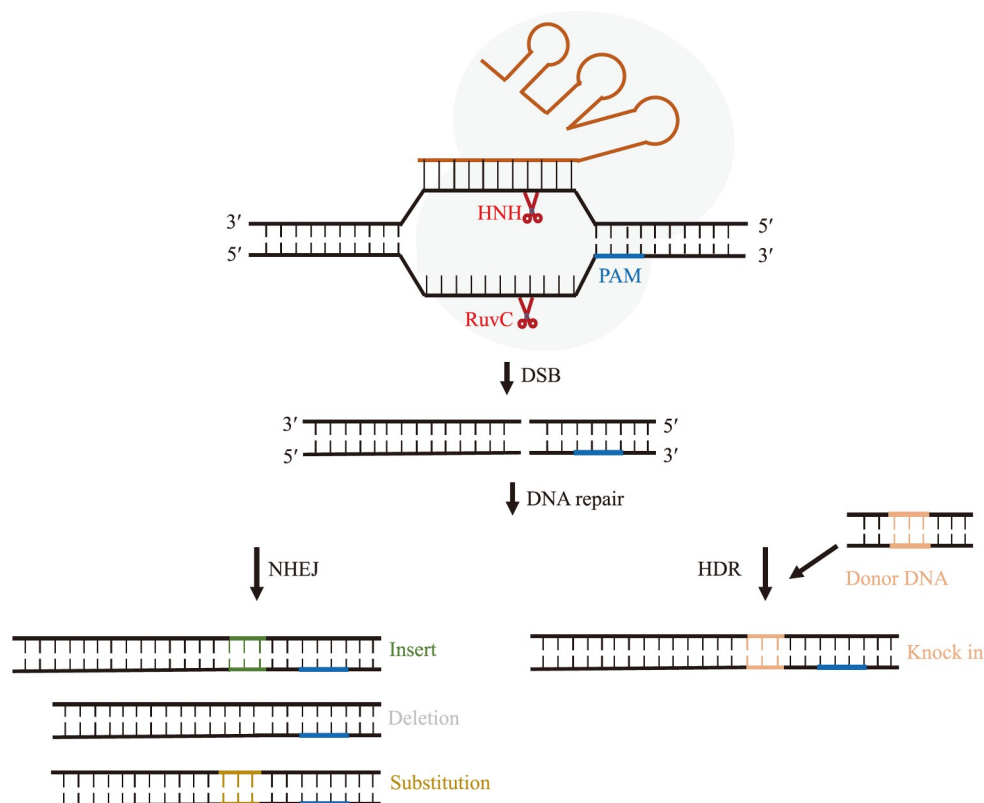


图2 CRISPR/Cas9基因编辑原理图

制作用^[33]以来,其他来源的直系同源物Cas9或工程化的Cas9也因蛋白质大小、PAM序列、sgRNA结构、最佳间隔区长度、编辑效率等优势受到科学家们的关注,并对其编辑效率进行了测试。例如,相较于1 368个氨基酸的spCas9,来源于金黄色葡萄球菌(*Staphylococcus aureus*)的saCas9^[34]、来源于嗜热链球菌(*Streptococcus thermophilus*)的stCas9^[35],因小于1 100个氨基酸,在腺相关病毒递送至体细胞组织进行基因编辑中会有优势^[36]; spCas9突变体SpRY^[37](NRN>NYN PAMs)降低了对PAM序列的限制,可靶向更多的疾病相关突变位点,更适于后续治疗应用。

因操作简单、价格低廉、编辑效率高等优势,目前CRISPR/Cas9已被广泛应用到以下领域:高通量筛选研究基因功能^[38,39];建立用于诊断和基因治疗的细胞模型^[40]和动物模型^[41];开发CRISPR干扰技术^[42](CRISPR interference, CRISPRi)和CRISPR激活技术^[43](CRISPR activation, CRISPRa)用于信号通路的揭示和药物靶点的筛选;荧光分子与催

化失活的Cas9(dead Cas9, dCas9)结合用于活细胞成像^[44]等。

3 基于CRISPR的单细胞转录组测序技术

基于CRISPR的单细胞转录组测序技术关键是将干预结果分配给不同的细胞(图3)。在获得CRISPR干扰文库及相应细胞转录组数据后,依赖sgRNA在文库中的索引(indexes)来确定表型与相关扰动基因的对应关系。为了确定表型与相关扰动基因的因果关系,scRNA-Seq筛选平台通常在表达sgRNA的质粒中加入向导标签(guide barcode, GBC),建立sgRNA/GBC库,间接确定表型与相关扰动基因的因果关系,例如Perturb-Seq^[45]、Mosaic-Seq^[46]。CROP-Seq(CRISPR droplet sequencing)^[47]通过在3'长末端重复序列(long terminal repeat, LTR)处设计直接表达sgRNA, Direct capture Perturb-Seq^[48]通过在逆转录过程中添加特定靶向引物,二者具有直接读取sgRNA的优势,直接确定表型与相关扰动基因的因果关系。

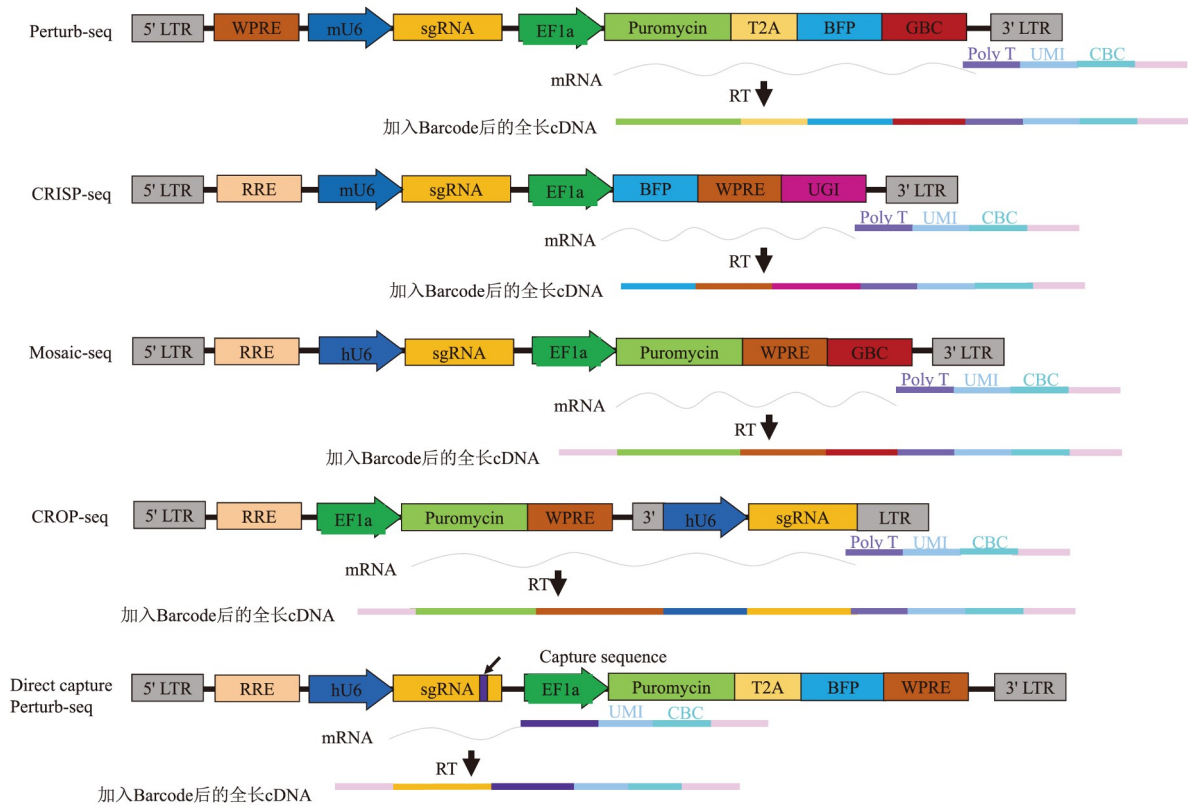


图3 基于CRISPR的单细胞转录组测序文库制备

3.1 Perturb-Seq

2015年, Aviv Regev实验室率先将大规模平行单细胞转录组测序技术^[18,49]和CRISPR/Cas9产生多位点基因扰动^[50,51]结合, 开发了一种称为Perturb-Seq^[45]的筛选技术。Perturb-Seq可以准确识别单个基因扰动或遗传相互作用下的单细胞基因特征和细胞状态。研究人员设计并使用了一种CRISPR慢病毒载体向细胞递送sgRNA, 并通过表达的GBC来报告sgRNA^[52]。通过调整病毒滴度(multiplicity of infection, MOI)来研究单个基因或上位效应。此外质粒还包括一个蓝色荧光蛋白(blue fluorescent protein, BFP)-T2A-嘌呤霉素的开放阅读框(open reading frame, ORF), 以备流式细胞仪分选或抗生素选择转导的细胞。单细胞转录组测序中使用独特的细胞标签(cell barcode, CBC)和UMI来标记并区分含有GBC的单个细胞的mRNA。研究人员还设计了一个基于正则化线性模型的计算框架, 用MPMOSCA(multi input, multi output single cell analysis)来估计扰动对基因表达的影响, 以解释基因扰动和遗传相互作用在单细胞转录进程和细胞

状态上的边际效应。

3.2 CRISP-Seq

Jaitin等^[53]融合大规模平行单细胞转录组测序技术和CRISPR集中筛选技术, 开发了CRISP-Seq (clustered regularly interspaced short palindromic repeats and sequencing), 用以重建控制髓细胞分化及其对病原体反应的调节网络。研究人员使用一个可扩展的慢病毒载体作为骨架, 还包括gRNA、多聚腺苷独特向导标签(unique guide index, UGI)和荧光选择标记。为了利用CRISP-Seq技术更好地表征单个细胞的基因型-表型关系, 并识别多重扰动, 研究人员开发了一种算法, 可以更准确地检测具有不同表型的扰动单细胞。算法的框架依赖于这样的假设, 即具有相似基因型的细胞将在表型空间中非常接近; 因此, 真正丧失功能的靶细胞群将产生靶细胞群内相似并且与非靶细胞群间不同表型。利用这一假设, 研究人员试图克服数据中假阳性和假阴性两个潜在异常值的来源。作者开发了一种基于k近邻图(k-nearest neighbor, kNN)的标记细化算法^[54]: 首先, 与邻近细胞具有

相同基因型的细胞保留UGI标记,具有不同基因型的细胞则失去UGI标记;随后,根据其邻近细胞的基因型将对应的基因型标记传递到失去标记的细胞中,由此来提高准确率。

3.3 Mosaic-Seq

转录增强子调节细胞特异性基因表达,对真核生物发育和疾病发生起着至关重要的作用^[55]。由于每个增强子都存在于三维折叠的基因组^[56]独特的遗传和表观遗传环境中^[57],而这些环境因素都影响着增强子的活性,目前基于质粒的方法^[56,58]可能不能完全概括内源性增强子的功能。Xie等^[46]开发了Mosaic-Seq技术,使用含有表观遗传修饰因子KRAB的慢病毒dCas9-KRAB病毒载体感染野生型细胞系,使用杀稻瘟菌素筛选细胞群后再使用带有BC的sgRNAs慢病毒库对细胞进行感染。使用嘌呤霉素筛选后的细胞群是经过表观遗传修饰且表达不同sgRNA的嵌合细胞,随后使用Drop-Seq技术^[18]来表征细胞。研究人员研究了71种增强子,发现其可被划分成15类超级增强子,每类超级增强子中,只有一个增强子对靶基因具有较强的作用,很多基于组蛋白修饰的增强子功能是冗余的。

3.4 CROP-Seq

基于gRNAs及其诱导的细胞应答提前在单细胞内被划分的想法,Datlinger等^[47]开发了CROP-Seq技术。常用于CRISPR集中筛选的慢病毒载体因缺失400 bp的关键启动子元件而失活。研究人员假设其3'LTR可以插入大小约400 bp的hU6-gRNA盒,这个位置的gRNA成为RNA聚合酶 II (RNA polymerase II)转录嘌呤霉素mRNA的一部分。此外,在慢病毒的逆转录和整合过程中,整个hU6-gRNA盒还会被复制插入到5'LTR,但并不会干扰EF-1a启动子上游的第二个拷贝。不同于Perturb-Seq等技术使用BC来进行筛选,CROP-Seq技术使sgRNA与表型直接配对,直接读取sgRNA,极大地简化了后续操作。

3.5 Direct capture Perturb-Seq

Perturb-Seq等技术在慢病毒递送过程中带索引的sgRNA库的重组可能使索引与其指定的sgRNAs分离,这意味着此类平台仅限于阵列使用且规模有限。由于hU6-gRNA盒大小的限制,CROP-Seq

与多种sgRNAs不兼容。鉴于这些技术的缺陷,Replogle等^[48]开发了Direct capture Perturb-Seq技术。与前面几种基于液滴的scRNA-Seq类似,都是使用分子标识来识别单个细胞的转录本。这种标识的添加发生在逆转录期间,将唯一的UMI和CBC添加到mRNA序列的3'或5'端。此外,对于Direct capture Perturb-Seq,研究人员还通过在逆转录过程中添加特定靶向引物(guide-specific primers),将这种BC扩展到非多聚腺苷酸化的sgRNAs中,使表达的sgRNAs与单细胞转录组一起被测,由此将sgRNA与表型直接配对。此外,研究人员还成功实现了向单细胞中递送靶向同一个基因的数个不同sgRNAs,证明了Direct capture Perturb-Seq单个细胞单个基因多个sgRNAs能够提高CRISPR干扰和激活的效率,有助于将紧凑、高效的CRISPR文库用于单细胞筛选。

4 基于CRISPR的单细胞转录组测序技术的应用

遗传扰动研究有助于描述基因和表型之间的因果关系,而单细胞转录组提供了更丰富的分子和细胞表型数据,因此,结合遗传扰动和单细胞转录组测序的筛选已经成为描述遗传扰动后果的有力工具并被科学家们应用到生物研究的各个领域。

4.1 遗传谱系追踪

研究细胞的谱系关系对揭示细胞正常分化机制和理解疾病发生发展都至关重要。Cas9核酸酶利用gRNA靶向特定的基因组区域,产生DSBs,在修复过程中引入了长度大小和位置都可变的独特的indels并成为永久的基因组疤痕,即为基因组BC,每个带BC的细胞将BC传递给子代,从而允许克隆跟踪和谱系重建(图4)。Spanjaard等^[4]开发了LINNAEUS(lineage tracing by nuclease-activated editing of ubiquitous sequences)技术,利用红色荧光蛋白(red fluorescent protein, RFP)转基因斑马鱼受精卵,在1-细胞阶段向受精卵注射Cas9核酸酶和靶向RFP的sgRNA以产生疤痕,红色荧光也随之消失,随后分析了受精5 d后斑马鱼受精卵的单细胞转录本,发现除红细胞外,所有血细胞具有相同的疤痕。这一结果反映了胚胎红细胞的独特起

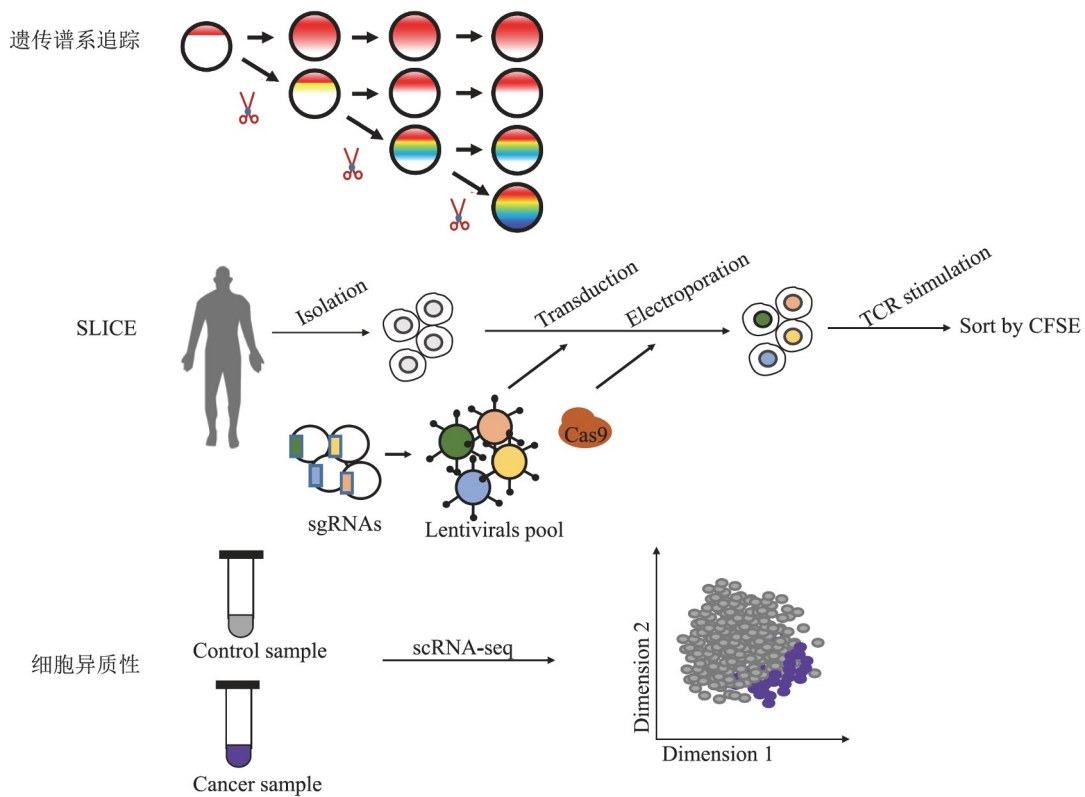


图4 基于CRISPR的单细胞转录组测序技术应用示意图

源, 它是在造血的启示阶段产生的, 而所有的成人血细胞都是在造血的终末期由内皮细胞产生的。此外还发现胰腺中, 胰岛内分泌细胞 α 细胞、 β 细胞、 δ 细胞有着共同的起源。Alemany等^[59]使用带八个串联组蛋白-绿色荧光蛋白(green fluorescent protein, GFP)的转基因斑马鱼。通过受精卵注射Cas9蛋白和靶向GFP的sgRNA或受精卵卵黄注射Cas9 RNA和靶向GFP的sgRNA, 使用ScatTrace揭示了部分多能胚胎祖细胞在骨髓中产生所有造血细胞, 胚胎祖细胞在穹顶期后不久就进入右眼或左眼, 尾鳍中的表皮细胞和间充质细胞来自相同的祖细胞, 成骨细胞的前体细胞在再生过程中可以产生间充质细胞, 鳍中的常驻免疫细胞具有不同于其他血细胞的克隆起源。Chan等^[60]利用piggyBac将合成DNA整合到小鼠卵母细胞, 合成DNA包括一个8 bp整合标签(integration barcode, intBC)、三个靶向位点序列, 以及受三个不同启动子启动的gRNAs, 通过精细胞中注射Cas9核酸酶, 体外受精后在胚胎第8.5天或第9.5天进行分析, 建立了从受精到原肠胚形成的小鼠细胞命运图谱。

4.2 功能基因筛选与免疫治疗

结合scRNA-seq和CRISPR集中筛选, 可用于揭示疾病发生发展的分子机制, 寻找药物治疗靶点。Shifrut等^[61]设计建立了靶向已知细胞表面蛋白和TCR信号通路关键组分的sgRNA文库, 使用SLICE(sgRNA lentiviral infection with Cas9 protein electroporation)技术, 即利用慢病毒对人类原代T细胞进行转导, 随后对T细胞进行电穿孔递送Cas9蛋白, TCR刺激4 d后通过流式分选将细胞分为非增殖细胞和强增殖细胞以筛选人类原代T细胞增殖的正、负调节因子, 随后研究人员建立了靶向筛选出来的负调节因子基因、7个已知免疫检查点基因和非靶向对照的文库, 并用上述实验方法对T细胞进行遗传干扰, 结合Crop-Seq表征了人类原代T细胞免疫反应的关键基因, 揭示了靶向敲除增殖的负调控因子可以诱导特异性刺激依赖的效应基因表达, 从而增强T细胞的功能。Cohen等^[62]利用接受硼替佐米(Bortezomib)治疗的多发性骨髓瘤患者的肿瘤细胞进行MARS-Seq, 筛选到一些高风险多发性骨髓瘤相关的新基因, 揭示了一种新的多发

性骨髓瘤耐药特征。其中*PPIA*基因编码的亲环素A可促进蛋白质的有效折叠,研究人员通过CRISPR/Cas9技术敲除多发性骨髓瘤细胞系*PPIA*基因和使用*PPIA*基因小分子抑制剂(CsA),确定了*PPIA*在蛋白酶体抑制剂抗性中的直接作用,并确定了卡非佐米(Carfilzomib)和CsA的协同治疗具有高效特异性的肿瘤杀伤作用。为了研究疾病相关基因在不同细胞类型以及细胞发育过程中的基因功能,Jin等^[63]开发了*in vivo* Perturb-Seq,利用慢病毒向组成型表达Cas9小鼠的E12.5胚胎前脑脑室递送靶向自闭症谱系障碍/神经发育迟缓(autism spectrum disorder/neurodevelopmental delay, ASD/ND)相关基因sgRNA,随后在小鼠出生后第7天分选BFP阳性细胞进行Perturb-Seq,利用加权基因相关网络分析(weighted gene correlation network analysis, WGCNA)和结构主题建模(structural topic modeling, STM)在皮质投射神经元、皮质抑制神经元、星形胶质细胞、少突胶质细胞和小胶质细胞/巨噬细胞中定义出14种基因模块,ASD/ND相关基因*Adnp*、*Ank2*、*Ash1l*、*Chd8*、*Gatad2b*、*Pogz*、*Scn2a1*、*Stard9*和*Upf3b*对其中五个模块有显著影响。此外,还对比了人类组织细胞scRNA-Seq数据,这14个基因模块具有保守性。*In vivo* Perturb-Seq可用于研究单/多疾病基因对发病的影响,揭示疾病基因在疾病发生发展过程中形成的功能网络。

4.3 细胞异质性

基因表达在细胞之间是高度异质的,因此scRNA-Seq可用于表征表型异常的细胞、鉴定多能细胞(图4)。Ji等^[64]使用10× Genomics Chromium对人皮肤鳞状细胞癌细胞(cutaneous squamous cell carcinoma, cSCC)和正常皮肤细胞进行scRNA-Seq,构建了细胞亚群的单细胞转录组图谱,并对细胞亚群进行了划分,在cSCC细胞中发现了正常细胞没有的肿瘤特异性角质形成细胞(tumor-specific keratinocyte, TSK),TSKs表达经典的上皮间质转化(epithelial mesenchymal transition, EMT)标志物,提示cSCC在表皮分化过程中失调导致细胞无法完全参与分化,基底细胞迅速增殖,表达EMT连锁基因的TSK亚群出现。通过对SCC异种移植肿瘤细胞进行CRISPR/Cas9筛选,确定了TSKs

以*PCNA*和*GINS2*为中心的最大基因调控网络。Nguyen等^[65]使用WTC-CRISPRi hiPSCs(人类诱导多能干细胞)^[66]进行scRNA-Seq,成功把细胞分为四个亚群:核心多能性、多能增殖性、早期始发态、晚期始发态。亚群一和亚群二占总细胞的96.1%,两者的主要区别是控制多能和分化的相关基因表达水平不同,揭示了细胞增殖相关基因网络。该网络由两个转录因子*SALL4*、*ZIC1*和*NR5A1*信号受体,以及多能性调节因子*POU5F1*、*SOX2*和*NANOG*协同调控。亚群三、亚群四和亚群一、亚群二相比,细胞从多能性向谱系决定过渡,多能性相关基因表达显著下调,多能干细胞调控网络被逐步破坏,基因表达表现出高变异性特征。Zhao等^[67]对MIN6细胞系进行CRISPR递送并结合scRNA-Seq分析,确定了MIN6细胞系随着细胞传代次数的增加而产生遗传异质性,其中神经分泌细胞高丰度表达lncRNA,通过实验证明了未表征的lncRNA GM26917、Cenpw与胰岛素转录有关。

5 总结与展望

在单细胞水平,高通量转录组分析技术促进了对细胞群体、疾病状态、发育谱系的理解。然而,scRNA-Seq技术复杂的流程以及产生的庞大数据,给后续分析和解释带来了挑战,总的来说有以下几点。(1)实验材料有限。对于许多基因来说,每个细胞都只有几十个转录本,总RNA约10 pg,其中仅有0.1 pg是mRNA^[68]。因此需要通过一个扩增过程产生足够多的cDNA用于测序。但因为GC含量不一致,导致基因比例扩增失衡。(2)单细胞捕获困难。目前将器官、组织分离成单细胞的方法有:显微操作分离,此方法能确保样品是单一细胞,但劳动密集、产量低^[69];基于流式细胞术的选择,这种方法使细胞空间信息在分离过程中丧失;酶解分离,用于组织分离的酶都需要在37℃进行孵育,但37℃也是哺乳动物体内多种酶的最适温度,孵育过程中引起基因表达的改变,造成基因表达假象。(3)数据稀疏性和噪声。基因不是以稳定的方式表达的,对于一个给定的细胞,各种基因转录本的水平处于不断变化的状态,基因表达的脉冲性、爆发性引起生物噪音^[70-72]。此外,逆转录和扩增步骤的低效率也给单细胞基因表达

研究增大了技术噪音。(4)数据集成。数据集成对于研究单细胞的不同模式尤为重要,但不同研究模式间还不能进行数据集成。例如目前的单细胞ATAC-Seq(scATAC-Seq)可以揭示增强子区域及调节关系,但目前无法像转录组学那样获得无监督细胞类型的发现能力^[73]。

正如二十世纪人类三大科学计划之人类基因组计划,通过测定人类染色体中所有核苷酸序列,成功绘制人类基因组图谱,破译人类遗传信息。随着单细胞测序技术不断发展与成熟,可检测的细胞总量及单个细胞可检测的参数不断提升,科学家们也希望跨实验或模式集成单细胞数据,建立一个全面的、包括人体所有细胞的人类细胞图谱^[74]。目前Han等^[75]对60种人体组织样品和7种细胞培养样品进行了Microwell-Seq,建立了70多万万个单细胞的转录组数据库,鉴定了人体100余种细胞大类和800余种细胞亚类并系统性地绘制了跨越胚胎和成年两个时期、涵盖八大系统的人类细胞图谱。但完整的人类细胞图谱还应该整合空间信息、多组学数据和人群分析,以期通过研究单个细胞内的多模式数据类型之间的关系,揭示细胞功能的潜在基础,推断模式之间的因果关系。

综上所述,本文对比了现有的六种单细胞测序技术的文库制备流程,回顾了CRISPR/Cas9原理,介绍了基于CRISPR的单细胞转录组测序技术的研究现状。未来,随着细胞分离技术的提升以及转录组扩增、测序和计算方法的改进,将有助于我们对生命现象的理解与解释,为人类疾病治疗作出重大贡献。

参考文献

- [1] Brown DD. Gene expression in eukaryotes. *Science*, 1981, 211(4483): 667-674
- [2] Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*, 2018, 14(8): 479-492
- [3] Szabo PA, Levitin HM, Miron M, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun*, 2019, 10(1): 4706
- [4] Spanjaard B, Hu B, Mitic N, et al. Simultaneous lineage tracing and cell-type identification using CRISPR Cas9-induced genetic scars. *Nat Biotechnol*, 2018, 36(5): 469-473
- [5] Peng J, Sun BF, Chen CY, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res*, 2019, 29(9): 725-738
- [6] Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*, 2018, 18(1): 35-45
- [7] Katzenelenbogen Y, Sheban F, Yalin A, et al. Coupled scRNA-Seq and intracellular protein activity reveal an immunosuppressive role of TREM2 in cancer. *Cell*, 2020, 182(4): 872-885.e19
- [8] Haapaniemi E, Botla S, Persson J, et al. CRISPR Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med*, 2018, 24(7): 927-930
- [9] Zhao H, Li Y, He L, et al. *In vivo* AAV-CRISPR/Cas9 mediated gene editing ameliorates atherosclerosis in familial hypercholesterolemia. *Circulation*, 2020, 141(1): 67-79
- [10] Debernardi JM, Tricoli DM, Ercoli MF, et al. A GRF GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat Biotechnol*, 2020, 38(11): 1274-1279
- [11] Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 2009, 6(5): 377-382
- [12] Enge M, Arda HE, Mignardi M, et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell*, 2017, 171(2): 321-330.e14
- [13] Haque A, Engel J, Teichmann SA, et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*, 2017, 9(1): 75
- [14] Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*, 2017, 541(7637): 331-338
- [15] Hindson BJ, Ness KD, Masquelier DA, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem*, 2011, 83(22): 8604-8610
- [16] Hashimshony T, Wagner F, Sher N, et al. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*, 2012, 2(3): 666-673
- [17] Hashimshony T, Senderovich N, Avital G, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*, 2016, 17(1): 77
- [18] Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 2015, 161(5): 1202-1214
- [19] Ziegenhain C, Vieth B, Parekh S, et al. Comparative

- analysis of single-cell RNA sequencing methods. *Mol Cell*, 2017, 65(4): 631-643.e4
- [20] Ding J, Adiconis X, Simmons SK, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*, 2020, 38(6): 737-746
- [21] Picelli S, Björklund ÅK, Faridani OR, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*, 2013, 10(11): 1096-1098
- [22] Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, 2014, 343(6172): 776-779
- [23] Soumillon M, Cacchiarelli D, Semrau S, et al. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv*, 2014: 003236
- [24] Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 2017, 8(1): 14049
- [25] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, 315(5819): 1709-1712
- [26] Shmakov S, Abudayyeh OO, Makarova KS, et al. Discovery and functional characterization of diverse class 2 CRISPR Cas systems. *Mol Cell*, 2015, 60(3): 385-397
- [27] Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR Cas systems. *Nat Rev Microbiol*, 2015, 13(11): 722-736
- [28] Wright AV, Nuñez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, 2016, 164(1-2): 29-44
- [29] Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011, 471(7340): 602-607
- [30] Garneau JE, Dupuis MÉ, Villion M, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010, 468(7320): 67-71
- [31] Gasiunas G, Barrangou R, Horvath P, et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*, 2012, 109(39): E2579-E2586
- [32] Komor AC, Kim YB, Packer MS, et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 2016, 533(7603): 420-424
- [33] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337(6096): 816-821
- [34] Ran FA, Cong L, Yan WX, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature*, 2015, 520(7546): 186-191
- [35] Esvelt KM, Mali P, Braff JL, et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*, 2013, 10(11): 1116-1121
- [36] Jiang F, Doudna JA. CRISPR Cas9 structures and mechanisms. *Annu Rev Biophys*, 2017, 46(1): 505-529
- [37] Walton RT, Christie KA, Whittaker MN, et al. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, 2020, 368(6488): 290-296
- [38] Michels BE, Mosa MH, Streibl BI, et al. Pooled *in vitro* and *in vivo* CRISPR-Cas9 screening identifies tumor suppressors in human colon organoids. *Cell Stem Cell*, 2020, 26(5): 782-792.e7
- [39] Wang T, Wei JJ, Sabatini DM, et al. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 2014, 343(6166): 80-84
- [40] Lentsch E, Li L, Pfeffer S, et al. CRISPR/Cas9-mediated knock-out of Kras^{G12D} mutated pancreatic cancer cell lines. *Int J Mol Sci*, 2019, 20(22): 5706
- [41] Li H, Yang Y, Hong W, et al. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Sig Transduct Target Ther*, 2020, 5(1): 1
- [42] Gilbert LA, Larson MH, Morsut L, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 2013, 154(2): 442-451
- [43] Maeder ML, Linder SJ, Cascio VM, et al. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*, 2013, 10(10): 977-979
- [44] Chen B, Gilbert LA, Cimini BA, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 2013, 155(7): 1479-1491
- [45] Dixit A, Parnas O, Li B, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, 2016, 167(7): 1853-1866.e17
- [46] Xie S, Duan J, Li B, et al. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell*, 2017, 66(2): 285-299.e5
- [47] Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*, 2017, 14(3): 297-301
- [48] Replogle JM, Norman TM, Xu A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol*, 2020, 38(8): 954-961
- [49] Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 2015, 161(5): 1187-1201
- [50] Cong L, Ran FA, Cox D, et al. Multiplex genome

- engineering using CRISPR/Cas systems. *Science*, 2013, 339(6121): 819-823
- [51] Qi LS, Larson MH, Gilbert LA, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 2013, 152(5): 1173-1183
- [52] Adamson B, Norman TM, Jost M, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*, 2016, 167(7): 1867-1882.e21
- [53] Jaitin DA, Weiner A, Yofe I, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell*, 2016, 167(7): 1883-1896.e15
- [54] Neupane K, Rayamajhee B, Acharya J, et al. Comparison of nasal colonization of methicillin-resistant staphylococcus aureus in HIV-Infected and Non-HIV patients attending the national public health laboratory of central nepal. *Canadian J Infect Dis Med MicroBiol*, 2018, 2018: 1-9
- [55] Smith E, Shilatifard A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol*, 2014, 21(3): 210-219
- [56] Downen JM, Fan ZP, Hnisz D, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 2014, 159(2): 374-387
- [57] Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*, 2015, 518(7539): 317-330
- [58] Dostie J, Richmond TA, Arnaout RA, et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 2006, 16(10): 1299-1309
- [59] Alemany A, Florescu M, Baron CS, et al. Whole-organism clone tracing using single-cell sequencing. *Nature*, 2018, 556(7699): 108-112
- [60] Chan MM, Smith ZD, Grosswendt S, et al. Molecular recording of mammalian embryogenesis. *Nature*, 2019, 570(7759): 77-82
- [61] Shifrut E, Carnevale J, Tobin V, et al. Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell*, 2018, 175(7): 1958-1971.e15
- [62] Cohen YC, Zada M, Wang SY, et al. Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing. *Nat Med*, 2021, 27(3): 491-503
- [63] Jin X, Simmons SK, Guo A, et al. *In vivo* Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science*, 2020, 370(6520): eaaz6063
- [64] Ji AL, Rubin AJ, Thrane K, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 2020, 182(2): 497-514.e22
- [65] Nguyen QH, Lukowski SW, Chiu HS, et al. Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. *Genome Res*, 2018, 28(7): 1053-1066
- [66] Mandegar MA, Huebsch N, Frolov EB, et al. CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell*, 2016, 18(4): 541-553
- [67] Zhao R, Lu J, Li Q, et al. Single-cell heterogeneity analysis and CRISPR screens in MIN6 cell line reveal transcriptional regulators of insulin. *Cell Cycle*, 2021, 20(19): 2053-2065
- [68] Petukhov V, Guo J, Baryawno N, et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol*, 2018, 19(1): 78
- [69] Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res*, 2017, 27(3): 491-499
- [70] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*, 2019, 20(1): 296
- [71] Fan J, Salathia N, Liu R, et al. Characterizing transcriptional heterogeneity through pathway and gene set over-dispersion analysis. *Nat Methods*, 2016, 13(3): 241-244
- [72] Brennecke P, Anders S, Kim JK, et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods*, 2013, 10(11): 1093-1095
- [73] Cusanovich DA, Hill AJ, Aghamirzaie D, et al. A single-Cell atlas of *in vivo* mammalian chromatin accessibility. *Cell*, 2018, 174(5): 1309-1324.e18
- [74] Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet*, 2019, 20(5): 257-272
- [75] Han X, Zhou Z, Fei L, et al. Construction of a human cell landscape at single-cell level. *Nature*, 2020, 581(7808): 303-309