

染色质转座酶可及性测序研究进展

吴杰¹, 全建平¹, 叶勇¹, 吴珍芳¹, 杨杰¹, 杨明², 郑恩琴¹

1. 华南农业大学动物科学学院, 国家生猪种业工程技术研究中心, 广州 510642

2. 仲恺农业工程学院, 动物科技学院, 广州 510225

摘要: 染色质转座酶可及性测序(assay for transposase-accessible chromatin with high-throughput sequencing, ATAC-seq)诞生于 2013 年, 具有比脱氧核糖核酸酶 I 超敏感位点测序(deoxyribonuclease I hypersensitive site sequencing, DNase-seq)和微球菌核酸酶敏感位点测序(micrococcal nuclease sequencing, MNase-seq)更快速、灵敏、简便的优点, 是目前分析全基因组范围染色质开放区域的热点技术。通过该技术能获得染色质开放区域的相关信息, 从而映射出转录因子等调控蛋白的结合区域和核小体定位等信息, 对于研究表观遗传分子机制具有重要意义。本文比较了 5 种获取染色质开放区域技术的优缺点, 重点介绍了 ATAC-seq 的原理和主要流程, 描述了利用 ATAC-seq 技术研究染色质开放区域的发展概况以及 ATAC-seq 的相关应用, 期望对真核生物全基因组水平的染色质开放区域研究、顺式调控元件鉴定以及遗传调控网络的解析等提供借鉴。

关键词: 染色质转座酶可及性测序; 染色质开放区域; Tn5 转座酶; 表观遗传修饰; 转录因子

Advances in assay for transposase-accessible chromatin with high-throughput sequencing

Jie Wu¹, Jianping Quan¹, Yong Ye¹, ZhenFang Wu¹, Jie Yang¹, Ming Yang², Enqin Zheng¹

1. National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

2. College of Animal Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

Abstract: Assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) was developed in 2013. It has the advantages of more convenient operation and higher efficiency for DNA recovery than DNase I hypersensitive site sequencing (DNase-seq) and micrococcal nuclease sequencing (MNase-seq). ATAC-seq currently is the most popular technique of genome-wide mapping for chromatin accessibility. It provides information on binding regions of

收稿日期: 2019-11-14; 修回日期: 2020-01-29

基金项目: 广东省“扬帆计划”引进创新创业团队项目(编号: 2016YT03H062), 广东省现代农业产业技术体系生猪创新团队项目(编号: 2019KJ126)和广东省自然科学基金项目(编号: 2017A030313213)资助[Supported by Guangdong YangFan Innovative and Entrepreneurial Research Team Program (No. 2016YT03H062), Guangdong Modern Agricultural Industry Technology System Pig Innovation Team Project (No. 2019KJ126) and Guangdong Natural Science Foundation (No. 2017A030313213)]

作者简介: 吴杰, 硕士研究生, 专业方向: 分子遗传与动物育种。E-mail: wujiezi163@163.com

通讯作者: 杨明, 博士, 高级畜牧师, 研究方向: 动物遗传育种。E-mail: yangming@zhku.edu.cn

郑恩琴, 硕士, 高级实验师, 研究方向: 遗传育种。E-mail: eqzheng@scau.edu.cn

DOI: 10.16288/j.ycz.19-279

网络出版时间: 2020/2/29 8:47:16

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20200228.0936.002.html>

transcription factors and nucleosome localization on the chromatin. Thus, ATAC-seq is of great significance for studying the epigenetics and molecular mechanisms in chromatin structure. In this review, we compare the advantages and disadvantages of multiple techniques for profiling chromatin accessibility, and summarize the principles, main process, development and applications of ATAC-seq. We hope this review will provide a reference for study of genome-wide mapping for chromatin accessibility, identification of cis-regulatory elements, and dissection of the epigenetic and genetic regulatory networks using the ATAC-seq technology in eukaryotes.

Keywords: assay for transposase-accessible chromatin with high-throughput sequencing; open chromatin regions; Tn5 transposase; epigenetic modification; transcription factor

自然界中的生物根据其细胞核类型可以分为原核生物和真核生物,其中原核生物的细胞核无核膜包被,其遗传物质 DNA 裸露在外;而真核生物细胞的细胞核 DNA 并非裸露,而是以左旋超螺旋的方式(约 147 bp)绕八聚体结构的组蛋白 1.67 圈,进而形成核小体^[1,2]。相邻核小体的连接区由 10~80 bp 的游离 DNA 与组蛋白 H1 共同构成;核小体通过连接区的连接形成串珠式结构,这种串联结构进一步折叠、凝聚,形成染色质;最终多条染色质以高度螺旋化状态包裹于细胞核中^[3]。研究显示,染色质开放区域的基因组占总 DNA 序列的 2%~3%,且超过 90% 的开放区域均与转录因子(transcription factor, TF)的结合相关^[4]。以 TF 为代表的调控因子可与其他染色质结合蛋白相互作用,从而动态调控和维持染色质稳态,在发育过程的调控中发挥着不可替代的作用^[5~7]。在 DNA 复制或转录过程中, DNA 的折叠结构被打开,一些染色质区域处于开放状态,调控因子(如转录因子)会与这些裸露的无核小体结合的 DNA 部位结合,进而调控 DNA 的复制或转录过程^[8]。此外,有研究表明, DNA 折叠、凝聚形成的染色质物理结构并不是一成不变的,仍然能够发生动态的表观遗传修饰,如 DNA 甲基化、组蛋白修饰、染色质重塑等^[8~11]。因此,通过了解相关获取染色质开放信息的技术,学习技术原理和应用,明确了这些技术对于基因组调控元件的鉴定、转录因子结合位点的识别及转录调控机制等研究均具有重要意义。本文主要综述了染色质可及性研究技术的发展概况、以及染色质转座酶可及性测序(assay for transposase accessible chromatin with high-throughput sequencing, ATAC-seq)技术的原理和应用,以期表观遗传学研究提供重要的参考。

1 染色质开放区域研究技术的发展历程

染色质开放区域的研究源于人们发现某些染色质特定位点表现出对 DNase I 酶切的高度敏感性^[12~15]。后期研究表明,这些 DNase I 敏感位点(deoxyribonuclease I hypersensitive site, DHS)通常是顺式调控元件所在区域^[16],其染色质裸露、结构疏松,可与转录因子结合,从而便于 DNase I 与之结合并剪切,进而表现出高度敏感性^[17]。基于上述原理,染色质开放区域的鉴定工作也随之展开。最先开展的是 DHS 鉴定分析工作,该分析依赖 DNase I 高度敏感性特点,并与 Southern 杂交技术结合,不过很快发现该方法的灵敏性和精确性都较低,并且耗时费力^[18,19]。随着高通量测序技术(high-throughput sequencing, HTS)的发展及测序成本不断降低,衍生出一系列研究染色质开放区域的技术与方法,如脱氧核糖核酸酶 I 超敏感位点测序(deoxyribonuclease I hypersensitive site sequencing, DNase-seq)^[20]、微球菌核酸酶敏感位点测序(micrococcal nuclease sequencing, MNase-seq)^[11]、甲醛辅助性调控元件分离测序(formaldehyde-assisted isolation of regulatory elements followed by sequencing, FAIRE-seq)^[21]、核小体定位和甲基化组测序(nucleosome occupancy and methylome sequencing, NOMe-seq)^[22]和 ATAC-seq。在上述 5 种技术中,获取染色质开放信息的方式分为 3 种: DNase-seq、MNase-seq 以及 ATAC-seq 采用酶切法; FAIRE-seq 采用物理断裂法; NOMe-seq 技术则利用甲基化修饰。5 种技术的具体信息见表 1。ATAC-seq 与其他 4 种技术相比表现出更为简便和高效的优势,一经发明就被广泛采用,成为当前染色

表 1 5 种染色质可及性研究技术介绍
Table 1 Introduction of five chromatin accessibility assays

技术	细胞类型及数量	获取方式	目的	特点	参考文献
DNase-seq	需要 1,000,000~10,000,000 的任何细胞类型	DNase I 酶切	获取染色质开放信息	(1)比传统方法操作更简便; (2)细胞需要量大; (3)酶最优酶切浓度确定过程繁琐; (4)样品制备过程复杂且耗时	[20,24]
MNase-seq	需要 1,000,000~10,000,000 的任何细胞类型	微球菌核酸酶酶切	绘制核小体图谱以间接探测染色质可及性	(1)操作简单, 后期数据处理方便; (2)细胞样本需要量大; (3)酶浓度和切割温度难以确定	[26,27]
FAIRE-seq	需要 100,000~10,000,000 的任何细胞类型	超声波物理断裂	获取染色质开放信息	(1)不用酶切、不需要分离出细胞核; (2)没有序列切割特异性; (3)细胞需要量大; (4)甲醛最佳交联程度难以确定	[21,24,31]
NOMe-seq	至少 1,000,000 的任何细胞类型	甲基化修饰	获得内源 DNA 甲基化的信息并定位核小体	(1)不需要使 DNA 断裂, 不会产生富集偏差; (2)同时获得含 GpC 和 CpG 二核苷酸的信息; (3)细胞需要量大	[22,34]
ATAC-seq	500~50,000 个新鲜分离的细胞	Tn5 转座酶酶切	获取染色质可及性、转录因子结合以及核小体定位信息	(1)过程简便、效率高; (2)数据分析工具不够成熟; (3)线粒体、叶绿体中的 DNA 污染; (4)冷冻组织细胞 DNA 提取效率低; (5) DNA 片段损失过多	[36~38]

质开放区域获取的前沿技术。下面将对上述染色质开放区域获取技术的发展历程、作用机理以及研究进程进行描述。

1.1 DNase-seq 及其衍生技术

早在 2006 年, DNase-seq 技术便被用于 DHS 区域的探究。该技术过程可简单概括为: 先通过裂解剂裂解细胞释放细胞核, 选用最佳浓度 DNase I 消化细胞核并包埋于低熔点凝胶琼脂糖塞中, 以减少额外的随机剪切; 随后, 将 DNA 片段平末端化并在两端连接上接头, 通过 PCR 扩增目的片段完成测序文库构建^[20]。DNase-seq 鉴定大多数转录因子的结合位点的精确性主要取决于酶切片段大小, 其中短片段(<100 bp)比较长片段效果更佳^[23]。DNase-seq 技术的建立使得 DHS 的分析鉴定变得更简便, 但存在细胞量需要量大(1,000,000~10,000,000)、样品制备过程复杂且耗时、最优酶浓度确定过程繁琐等缺点^[24]。2015 年, Cumbie 等^[25]发现使用传统 DNase-seq 技术消化拟南芥(*Arabidopsis thaliana*)时, 无法获得足够的 DNA 量用于建库测序。其主要原因在于植物细胞碎片及根毛等杂质占据了凝胶琼脂糖塞, 从而

导致 DNA 产量过低。因此, 他们在传统 DNase-seq 基础上建立了 DNase I-SIM (for simplified in-nucleus method)技术。该技术在 DNase I 消化细胞核前, 先采用 Percoll 层析液对细胞核进行初步纯化; 经 DNase I 消化后, 在 T4 DNA 聚合酶的作用下, DNA 片段的双链形成平末端, 而不包埋于琼脂糖塞中。这种改进的 DNase I-SIM 技术的优点在于, Percoll 梯度纯化后的细胞核及 T4 DNA 聚合酶处理后的 DNA 片段依然具备高度完整性, 在极大缩减 DNase-seq 文库制备时间的同时, 有效提升了 DNA 片段的回收效率。

1.2 MNase-seq 技术

MNase-seq 技术发明于 2008 年, 其原理与 DNase-seq 大体相似, 不同之处在于, 该技术采用的是微球菌核酸酶替代 DNase I 酶, 进而对细胞核中的 DNA 进行切割。该技术可通过揭示由核小体和其他调节因子占据的基因组区域, 从而间接探测染色质可及性并绘制核小体图谱^[26]。与 DNase-seq 相比, MNase-seq 具有操作简单、后期数据处理更方便等优点。然而, MNase-seq 技术仍存在一些技术弊端。

首先,其灵敏度同样受到细胞样本数量(1,000,000~10,000,000)、酶浓度和切割温度等因素的影响^[27];其次,微球菌核酸酶对 A/T 碱基序列存在切割倾向性^[28],从而无法精确切割核小体边界^[29]。

1.3 FAIRE-seq 及其衍生技术

2007 年, Giresi 等^[21]开发了一种物理打断 DNA 双链的技术,即 FAIRE-seq 技术,其过程相较于 DNase-seq 更为简便。该技术是利用超声波打断已用甲醛交联的 DNA 序列,在未解交联的条件下,通过酚-氯仿抽提,提取位于水相中的游离 DNA 并测序,最终获得相应的染色质开放区域信息^[21]。FAIRE-seq 具有不需要酶、不需要分离出细胞核、不受细胞类型限制、没有序列切割特异性^[30]、以及在增强子区域具有更高的覆盖率等优点^[31],但同样面临着样品需求量大的限制(需要 100,000~10,000,000 的细胞)^[24]。值得注意的是,该技术难以确定甲醛最佳交联程度,从而成为限制该技术应用的最大瓶颈。究其原因在于, DNA 与甲醛过度交联或不充分交联,均会影响到最终的测序结果^[32]。2009 年, Auerbach 等^[33]在此基础上创建了超声处理交联染色质测序(sonication of cross-linked chromatin sequencing, Sono-seq)技术,其原理与 FAIRE-seq 相同,其主要区别在于, FAIRE-seq 是在酚-氯仿抽提之后进行大小分级,选择特定大小范围如 100~350 bp 范围进行建库测序。由于存在大小分级选择这一关键步骤, Sono-seq 与 FAIRE-seq 各自所鉴定的 Peaks 存在明显的区别。

1.4 NOMe-seq 技术

NOMe-seq 技术由 Kelly 等^[22]于 2012 年发明,该技术利用 GpC 甲基转移酶(M.CviPI)通过甲基化修饰的方式处理开放区域的 GpC 二核苷酸。因为 GpC^m 不存在于人类基因组中, M.CviPI 使 GpC 甲基化为无内源背景的 GpC^m。随后,经过亚硫酸氢盐处理和全基因组测序,可同时获得含 GpC 和 CpG 二核苷酸的相关信息,从而能在全基因组范围内确定核小体的位置,同时还能获得内源 DNA 甲基化的信息^[22,34]。NOMe-seq 需要的细胞量为 1,000,000 个,材料处理因为不需要使 DNA 断裂,因此不会产生富集偏差,从而可降低假阳性的概率。但由于 NOMe-seq 不是基于先富集目的片段,然后再测序的方法,因

此需要大量的测序读长数据以获得足够的深度及基因组覆盖率,从而获取整个基因组的开放性水平^[35]。

1.5 ATAC-seq 技术

以上 4 种技术虽然都能应用于染色质开放性的表观基因组学研究,但通病是通常需要几万至数百万个细胞作为输入材料用以平均细胞群体的异质性,涉及到复杂、耗时的样品制备过程,且不能同时探究核小体定位、染色质可接近性和 TF 结合的相互作用。而多数情况下,很多重要且稀少的细胞亚型可能很难提供足够的样品量进行全基因组染色质可及性分析。2013 年, Buenrostro 等^[36]建立了材料需求量少、过程更为简便、效率更高的 ATAC-seq 技术。该技术仅需两步就能从 500~50,000 个细胞捕获染色质开放区域^[36,37]。与其他技术方法不同的是, ATAC-seq 利用高度活跃的 Tn5 转座酶代替 DNase I 核酸酶、微球菌核酸酶 MNase 等分析染色质可接近性,能够将目的 DNA 片段化、末端修复和加上测序所需的接头(adaptor)一步完成,从而使建库步骤变得极为简便,达到投入量更低、通量更高的建库效果。Tn5 转座子的深入研究以及高通量技术的快速发展使得 ATAC-seq 技术能够成功建立并广泛应用, ATAC-seq 技术以及其衍生技术如转座子超敏位点测序(transposome hypersensitive sites sequencing, THS-seq)、Omni-ATAC 技术使染色质开放区域的获取更加精准、高效,必将成为染色质开放区获取的主流技术之一。

2 ATAC-seq 作用机理和过程

2.1 Tn5 转座酶的发现及作用机理

20 世纪 40 年代,美国遗传学家 Barbara McClintock 发现了转座子(transposon)^[39]。转座子也叫跳跃基因或转座因子,是一段可以改变其在基因组中的位置的 DNA 序列。转座因子几乎存在于所有真核生物的基因组中,其衍生物构成了基因组的很大一部分,从而在基因组功能和进化过程中发挥着重要作用^[40]。根据转座子的结构特点和转座方式可将其分为 I 型和 II 型。I 型转座子,也称作 RNA 转座子(即反转录转座子),转座方式为“复制-粘贴”型,

即转座时先将自身转录获得 RNA,再反转录回 DNA,增加自身一倍的拷贝数,增加的 DNA 再转座至新的位置。II 型转座子,也称作 DNA 转座子,直接以“剪切-粘贴”的方式剪切下自身的 DNA 序列插入到新的位置,不会增加拷贝数。

Tn5 转座子是一种细菌转座子,属于 II 型转座子的一种^[41],最早是在 *Escherichia coli* 中被发现。Tn5 由编码卡那霉素(kanamycin, KAN)、新霉素(neomycin, NEO)、链霉素(streptomycin, STR)3 种抗生素的核心序列和位于侧翼的两个高度同源且倒置的 IS50(insertion sequence, IS)序列组成,该 DNA 序列全长 5818 bp^[42~45]。其中,IS50 序列可编码参与转座的蛋白:转座酶(transposase, Tnp)和转座阻遏蛋白(transposase inhibitor, Inh)。但由于左侧末端的 IS50L 序列的第 1442 位碱基处 T/A 碱基对被 C/G 碱基对取代发生突变,导致翻译提前终止,因此仅有 IS50R 能够表达正常有活性的 Tnp 和 Inh^[43]。每个 IS50 序列具有两个 19 bp 的倒置末端:外末端(outside end, OE)和内末端(inside end, IE)。两倒置末端有 7 个 bp 不同,该末端是 Tnp 的结合位点^[46~48]。

转座发生过程大致分为 3 个步骤:(1)形成转座复合体。Tnp 分子的两个 N 末端结构域分别结合到 Tn5 转座子的两个 OE 末端,形成两个 Tnp-OE 复合体^[49];随后以末端第 2 位序列为中心弯曲约 36°~48°^[50],两个复合体发生联会,Tnp 的 C 末端结构域相互作用而二聚体化,形成一个二聚体蛋白与两分子 DNA 组成的 Tn5 转座复合体^[51,52];(2) Tnp 切割。形成转座复合体结构后,Tnp 便具备了切割 DNA 的活性^[53],且正因为上述转座复合体结构,结合在左末端的 Tnp 便会催化右末端的磷酸二酯键水解,而结合在右末端的 Tnp 负责催化左末端的磷酸二酯键水解,以此有效防止 Tnp 只对转座子 DNA 链的一端进行切割^[54,55];(3)插入靶序列。Tnp 通过活化水分子水解 DNA 链,使 Tn5 两末端分别形成 3'-OH 亲核基团,该亲核基团对 DNA 互补链进行亲核攻击,形成发夹结构,另一活化的水分子水解发夹结构,使 Tn5 的两端均变为平末端,此时整个转座复合体离开供体 DNA,向靶序列结合^[56];转座子的 3'-OH 基团以交错的方式攻击靶序列中的磷酸二酯键,使转座子插入位点间形成 9 bp 的粘性末端,通过其 3'-OH 端同靶序列的 5'-P 之间形成共价键,插入到靶 DNA 中^[57,58],随后在 DNA 聚合酶的作用下,Tn5

的两侧翼形成 9 bp 的正向重复序列补齐缺口^[59]。至此,整个转座过程完成。

通过上述转座过程不难看出,体外 Tn5 转座过程仅需 4 个条件便能完成:Mg²⁺、转座子末端序列、Tnp 和靶 DNA^[60]。ATAC-seq 过程中使用的就是简化后的二聚体转座复合物。复合物仅含有 3 个部分:转座酶、末端序列和测序接头^[61],能够保证在切割 DNA 的同时连接上接头以便后续的测序工作。同时,简化复合物的 Tnp 在 Tn5 主链上携带了特异的点突变体,使 Tnp 具有了更高的活性^[60,62]。另外,之所以需要 Mg²⁺,是由于 Mg²⁺在转座过程中能协同亲核基团,在 Mg²⁺的作用下,转座酶上催化转座子运动的 DDE 基序(天冬氨酸和谷氨酸)与 Mg²⁺配位发生突变,使原本不活跃的转座酶变成高度活跃状态^[63,64],是完成转座必不可少的因子之一。目前,Tn5 转座子以其转座的随机性好、稳定性高、插入位点容易测序等特点,已经成为分子遗传学研究的热门工具^[65,66]。随着高通量测序技术的发展和实验通量的不断增加,Tn5 转座酶因其优势被应用得越来越广泛。其中,极速建库、长读长测序技术(single tube long fragment read, stLFR)、单细胞测序、Mate Pair 文库构建、染色质转座酶可及性可视化分析(assay of transposase-accessible chromatin with visualization, ATAC-seq),以及近几年发现的 Tn5 家族对于蛋白结合区域、互作基因片段等研究的帮助都显示 Tn5 转座酶拥有不可估量的应用潜力^[67~69]。Buenrostro 等^[36]建立的 ATAC-seq 技术,正是充分利用了 Tn5 酶在测序建库中的巨大优势,能高效、精准的从基因组水平鉴别出染色质开放区域,在生命科学领域的遗传学研究中发挥着至关重要的作用。

2.2 ATAC-seq 主要过程

ATAC-seq 涉及 3 个主要的步骤^[36,38]:(1)获取细胞核。使用冷裂解缓冲液裂解细胞;(2)转座和纯化(图 1A)。细胞核提取后立即将沉淀重悬于转座酶反应混合物中,转座后使用 Qiagen MinElute PCR Purification Kit 纯化样品;(3)PCR 扩增(图 1B)。纯化后,进行 qPCR 定量分析以及 PCR 扩增。上述过程大概需要 3 h,几处细节的处理尤为重要:(1)为减小 PCR 中的片段大小和 GC 偏差影响,需要通过 qPCR 来确定 PCR 后续的循环数,在饱和前停止扩增,以此保证转座后片段大小在 40 bp~1 kb 范围而

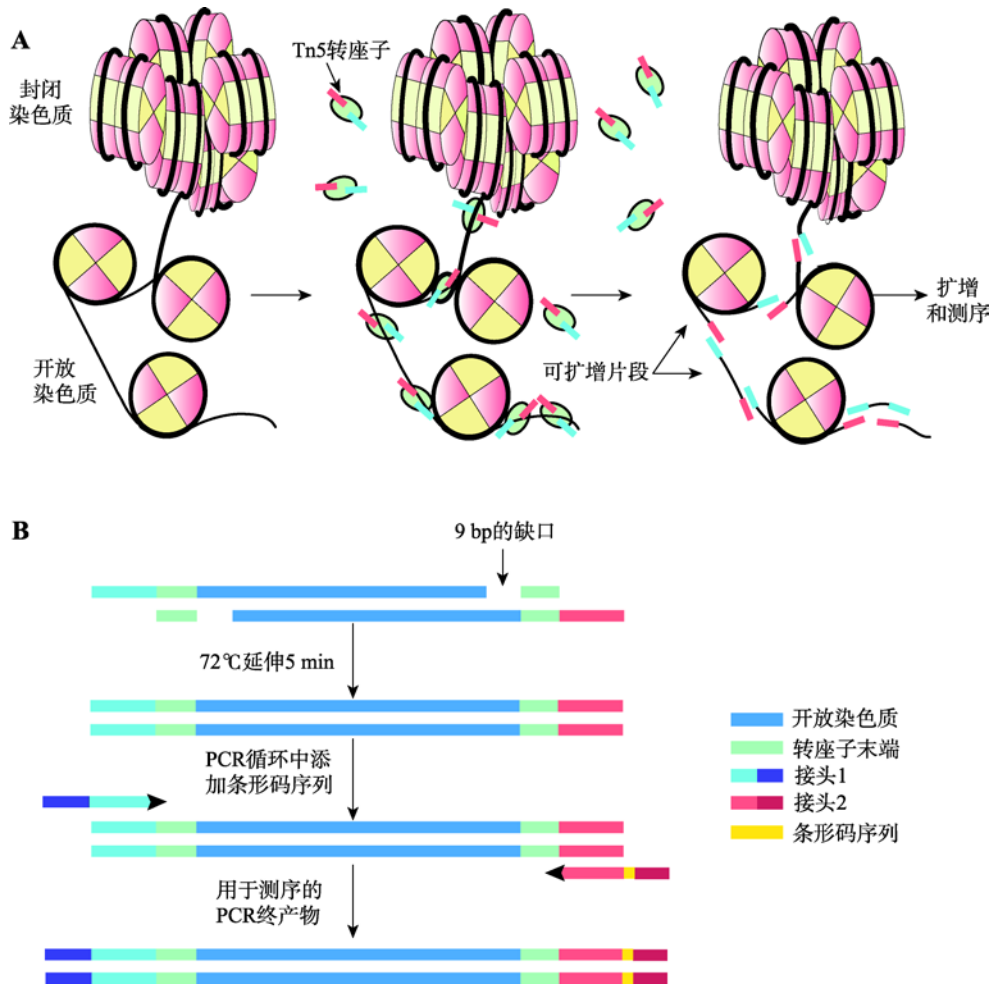


图 1 转座及扩增过程示意图

Fig. 1 The schematic transposition and PCR amplification process

A: 转座过程; B: 扩增过程。

不需要进行大小选择, 维持较高的库复杂性; (2) 因转座过程产生了 9 bp 的空隙, 因此 PCR 的第一步需要 72°C 反应 5 min 以填补该空隙, 且所用的 PCR 酶是具有链置换功能的非热启动酶; (3) 对于一个 DNA 片段而言, 两端接头连接是随机的。Tn5 酶切后会出现 3 类产物: 单端接头 1、单端接头 2 以及双端接头 1-接头 2, 只有连接不同接头的片段可用于富集扩增及测序。

2.3 数据分析

ATAC-seq 的数据分析过程大概可分为 4 个阶段^[24]: 第一阶段的数据预处理主要包括了过滤、比对和数据质量检测, 利用软件 CASAVA(Illumina) 获得 FASTQ 文件后过滤掉低质量片段, 由于相邻转座

的最小间隔为 38 bp, 通常 38 bp 以下的片段直接删除^[61], 并用 Bowtie 与参考基因组进行比对, 随后利用 SAMtools 去除重复的以及细胞器的 reads; 第二阶段是质控、数据可视化和 peak calling, 第二次的质控标准包括线粒体基因组所占比例高低和插入物大小分布图, 使用软件 IGV 可对数据进行可视化处理, 峰可以用软件 macs2、Hotspot 或者 ZINBA 寻找; 第三阶段是 peak 注释, 以获取基因组中 peak 的位置信息; 最后一步为模体(motif)注释和差异 peaks 分析, 将峰对应序列进行注释以间接确定转录因子信息, 同时, 利用 diffbind、DESeq2 等工具分析如不同的实验条件、多个时间节点、不同的发育时期等的差异区域, 最终获得转录因子结合位点的染色质可接近性状态信息。

2.4 ATAC-seq 优缺点分析

开展染色质开放区域的表现基因组学研究具有巨大的生物学意义,但过去的研究方法受到了复杂工作流程和大量细胞需求量的限制,从而导致该领域进展相对缓慢。直到 ATAC-seq 的出现,为注释开放染色质的基因组位置、DNA 结合蛋白、转录因子结合位点等基因组功能元件提供了新的契机。ATAC-seq 技术摆脱了像 DNase-seq 需要精确控制酶量以及 FAIRE-seq 需要确定甲醛交联时长等条件的限制,但依然存在影响其精确性的因素,如线粒体及植物细胞中叶绿体 DNA 的干扰、冷冻组织细胞 DNA 提取效率低、接头连接的随机性造成 DNA 片段的损失,以及大量酶切后的 DNA 片段过大而无法富集等^[24,36,38,70]。针对上述缺陷,同样产生了一系列改进措施。例如, Lu 等^[71]开发的 FANS-ATAC-seq (fluorescent activated nuclei sorting, FANS)、Roger 等^[72]开发的与细胞核基因组序列比对能达 90% 以上的 INTACT (isolation of nuclei tagged in specific cell types) 系统,以及与 INTACT 有相似结果的蔗糖沉淀法(crude)确保了在测定中使用高质量的完整细胞核的同时,能最大限度地减少线粒体和叶绿体中 DNA 的污染^[73]。Corces 等^[74]发明的 Omni-ATAC,提高了 ATAC-seq 对困难细胞系、稀少的原代细胞和临床上相关的冷冻组织中的应用普遍性。此外,针对接头的随机性和剪切后片段过大的问题, Sos 等^[75]开发了 THS-seq 技术,具有比传统 EzTn5 转座酶活性更高的新型 Tn5 超突变体(Tn5059)以及更优化的反应溶液和条件。同时设计 T7 启动子加转录引物替换原 Tn5 转座复合物中的 Adapter 1 和 2。通过转录生成单链 RNA,利用与 RNA 测序相同的原理获得 cDNA 并加上衔接子,最终完成建库。该技术避免了接头的随机连接,大大提高了转座效率,使得测序数据更为完整。随着 ATAC-seq 技术被不断改进, ATAC-seq 已逐渐成为目前染色质可及性分析的主流实验方法。

3 ATAC-seq 的应用和拓展

3.1 ATAC-seq 的应用

自 ATAC-seq 技术诞生起,该技术凭借其稳定性

和高灵敏度已广泛应用于表观基因组学研究。除了能用来确定功能基因组调控区域信息、找出组织特异基因以及预测潜在结合蛋白外,还能跟其他分析技术联合,如 RNA-seq、ChIP-seq(chromatin immunoprecipitation followed by high throughput sequencing)以及 Hi-C (high-through chromosome conformation capture)等,用以发现潜在的关键调控元件、转录因子和理解控制体内复杂过程的基因调控网络。其应用包括如下方面:

(1) DNA 调控功能元件注释。ATAC-seq 最直接的功能就是用来研究各种调控因子如 TF、启动子、增强子等的结合区域的开放状态以及对核小体进行定位,如对启动子区域开放状态的研究。Tan 等^[76]为探究寿命长且癌症发病率极低的裸鼹鼠的遗传机制,用 ATAC-seq 探测其染色质开放区域,结果显示裸鼹鼠的重编程基因的启动子区域更多的是处于关闭状态。通过对启动子区域开放状态后进行 SV40 LT (large t antigen, LT)抗原处理,发现 LT 可以抑制抑癌基因 *Rb*,从而提高重编程效率;而 LT 处理后,启动子区可接近性较处理前更高。上述结果表明裸鼹鼠细胞具有更稳定的表观基因组,提示可以利用这种稳定性为人类癌症预防和治疗提供新见解。同样,还有对转录因子结合区域开放性的研究。Scharer 等^[77]为深入了解系统性红斑狼疮(systemic lupus erythematosus, SLE)的表观遗传调控过程,对来自 SLE 和健康对照的 CD19⁺幼稚 B 细胞进行 ATAC-seq 分析。他们发现与健康对照相比,参与 B 细胞活化的基因周围的基因座,以及调节 B 细胞活化和分化的 TF 的结合位点在 SLE 细胞中可接近性更高,该结果验证了与 B 细胞活化相关的区域的开放状态,为进一步研究 SLE 疾病机制打下基础,充分体现了 ATAC-seq 确定染色质开放区域和鉴定顺式调控元件的能力。Kristofer 等^[78]利用 ATAC-seq 比较了致癌蛋白 Ras^{V12} 诱导的癌组织在早期肿瘤、晚期肿瘤与正常组织间染色质开放区域的状态。他们发现与正常组织比,肿瘤在发展过程中有数千个更容易接近的区域,从这些区域筛选并鉴定了在 Ras 依赖性肿瘤发生过程中异常活跃的调节区域,并结合 motif 分析确定出了结合这些区域的关键转录因子 AP-1 和 Stat92E。后期通过引入突变使 Stat92E 丧失功能后,发现肿瘤的严重程度降低,证明了转录因子

Stat92E 在肿瘤治疗发展中重要意义,也对 Ras 依赖性肿瘤的发展有了新的认知。上述基于 ATAC-seq 和 motif 分析的成功案例表明, ATAC-seq 可作为一种研究功能基因组调控区域和了解体内复杂基因调控网络的有效方式。另外,转录过程中,转录因子与核小体竞争结合 DNA 序列,转录因子结合处的核小体水平也因此较低^[79]。所以,获取核小体定位信息,对于了解转录调控、DNA 复制和修复等过程也很重要。Quillien 等^[80]为鉴定整个斑马鱼基因组中的细胞特异性增强子,对来自 Tg (fli1a: egfp)^{y1} 转基因胚胎的内皮细胞的细胞核进行 ATAC-seq 分析;在 FANS 技术的辅助下,用绿色荧光蛋白标记内皮细胞,通过荧光分离出细胞核获得高质量的 ATAC-seq 数据;后续通过分析短 DNA 片段(<100 bp)和长片段(180~247 bp)分别获得无核小体区及核小体结合区,以此定位核小体的位置。该研究揭示了整个基因组中转录起始位点的核小体定位模式以及与组蛋白修饰间的关联,还提供了在胚胎发育过程中控制基因表达的全基因组范围转录调控网络的动态信息。

(2) ATAC-seq 与 RNA-seq、ChIP-seq 等多组学数据的联合分析。ATAC-seq 技术更为巧妙的应用是将其获得的数据与其他表观遗传信息相结合,用以增强对科学问题的进一步解释。许多研究表明,通过联合 ATAC-seq 与 RNA-seq 数据进行分析,可发现潜在的特定性基因。Ackermann 等^[81]使用 ATAC-seq 对纯化的人 α 和 β 细胞中的开放染色质区进行了首次分析。通过与 RNA-seq 数据整合,进一步鉴定了两种细胞中已知的胰岛细胞转录因子的结合位点和已发现的 II 型糖尿病易感基因座的单核苷酸多态性(single nucleotide polymorphism, SNP)。更重要的是,发现了这两种细胞类型的新型特征基因,“组特异性蛋白质”(group specific protein)即维生素 D 结合蛋白仅存在于 α 细胞,而软骨素(chondrolectin)的免疫反应性仅存在于 β 细胞中。ATAC-seq 与 RNA-seq 的联合分析具备鉴定转录因子结合位点和发现潜在特异性基因的作用。2016 年, Garcia 等^[82]通过 ATAC-seq 比对分析了经典食道癌细胞系(OE33)、一种新发现的食管癌细胞系(MFD-1)、以及正常细胞系(HET1A)三种细胞系的特异染色质开放位点。他们发现和正常对照组相比, MFD-1 特异的染色质开放位点显著富集着 CTCF、NFY、Meis3、

Nrf2 的 motif,且这些位点附近基因富集在和食道癌及消化道肿瘤相关的功能基因区域内。后续结合全基因组测序和 RNA-seq 技术,发现了 MFD-1 的基因表达特性,进一步表明了 MFD-1 作为食管腺癌的临床疾病模型的可行性。同样,通过 ATAC-seq 与 RNA-seq 技术联合,可验证基因表达量与 ATAC-seq 信号的相关性,从而找出对应的转录因子。Ho 等^[83]为了找出特异表征间充质干细胞(mesenchymal stem cells, MSCs)的分子特征,从 8 周龄小鼠的四个组织(股骨和椎骨的骨髓、脂肪、肺)中分离出 MSCs 进行 ATAC-seq 和 RNA-seq 联合分析。首先,得出了 ATAC-seq 比转录组分析更加适合用来研究细胞特异性的结论,随后鉴定出可能有助于区分具有不同特征的 MSCs 的潜在转录因子,并推测这种通过研究染色质可及性来分析 MSCs 的方法,也可用于表征人类的 MSCs 和人类 MSCs 的临床应用。值得注意的是,虽然启动子可访问性和基因表达之间呈正相关,但也有许多研究表明,低表达或表达降低并不总是由于缺乏可访问区导致^[84,85]。因此, Starks 等^[86]再一次联合使用 ATAC-seq 和 RNA-seq 技术,探索染色质开放性与基因表达间的关联。试验通过分析妊娠中期小鼠胎盘的表达水平和启动子覆盖率将基因分成 3 组:高启动子覆盖率和表达的基因组(HA-HE)、中低覆盖率和表达的基因组(HA-ME)、低覆盖率和低表达的基因组(MA-ME)。结果发现 HA-HE 组基因可能是管家基因,而 HA-ME 组基因可能是组织特异性基因。随后,他们通过 motif 富集分析鉴定出了抑制 HA-ME 组基因的潜在转录因子,并发现 HA-ME 组的基因与胎盘的功能密切相关。该试验证明了 ATAC-seq 和 RNA-seq 的联合分析可用于小鼠和人的多种组织和细胞类型,用以鉴定活性抑制因子和组织特异性基因。

ATAC-seq 与 ChIP-seq 的联合使用常用于验证转录因子与目标开放区域的结合。2018 年, Rajbhandari 等^[87]联合 ATAC-seq、ChIP-seq 和 RNA-seq 揭示了抗炎因子 IL-10 的作用机理。为探究 IL-10 抑制脂肪细胞产热和能量输出的潜在机制,他们先利用 ATAC-seq 和 RNA-seq,验证了 IL-10 能改变脂肪细胞的染色质开放状态和降低产热相关基因 *UCP1* 的表达量。最后利用 ChIP-seq 揭示了 IL-10 通过抑制产热转录因子 ATF 和 C/EBP β 向增强子区

域募集进而达到抗炎作用的机制。Denny 等^[88]为探究小细胞肺癌的转移机制,联合 ATAC-seq、ChIP-seq 与 RNA-seq 对肺原发癌和肝脏转移癌进行比较分析。ATAC-seq 分析发现转移癌中的染色质开放性普遍增高,且差异的染色质开放性区域主要在基因远端调控元件上。随后通过 motif 富集分析、RNA-seq 和 ChIP-seq,发现显著富集的 *Nfib* 基因在转移癌中表达量增高,其 ChIP 信号与开放信号正相关,从而最终锁定关键基因 *Nfib*。

此外,还有联合 Hi-C 分析寻找调控元件如增强子。2018 年, Wang 等^[89]利用 ATAC-seq 和 Hi-C 研究多倍体棉花的三维基因组结构和转录调控之间的关系。他们发现一些由 DNase-seq 数据获得的 DHSs 与启动子之间存在互作,并推测这些处于开放状态的区域是潜在增强子,与 RNA-seq 数据整合进一步确认这些关键候选增强子具有转录活性。Mas 等^[90]联合 Hi-C、ATAC-seq、ChIP-seq 和 RNA-seq 四种技术,研究敲除甲基转移酶复合亚基基因 *MLL2* 后对基因组三维结构、染色质可接近性、组蛋白修饰以及表达水平的影响。首先利用 ChIP-seq 和 reChIP-seq 检测二价修饰,鉴定二价启动子。再通过 Hi-C 分析发现 *MLL2* 的敲除会引起二价基因转录起始位点与上下游互作模式的改变。随后 ATAC-seq 研究发现敲除 *MLL2* 引起二价启动子区的可接近性减少。最终揭示了 *MLL2* 通过作用于二价启动子,维持相对集中的二价基因间互作,可接近性状态和转录水平正常的作用。总之,ATAC-seq 能够与各种组学构建不同的关联模式,从不同的分析思路获取表观遗传信息或是探究转录调控机制,充分展现了其巨大的应用前景和无限的可能性。

3.2 单细胞 ATAC-seq 测序技术

染色质状态是以细胞类型特异性的方式动态调节的^[91]。虽然 DNase-seq、ATAC-seq 等技术用于测定全基因组水平染色质特定区域的可接近性,但测量获得的是群体细胞平均染色质状态,掩盖了细胞类型间和细胞内的异质性。因此,基于单细胞测序的 ATAC-seq (single-cell assay for transposase-accessible chromatin, scATAC-seq)被用于探测单细胞水平的染色质可及性^[92]。scATAC-seq 可应用于分析细胞亚群的基因调控网络、研究细胞异质性、发现生物

标志物、研究单细胞表观基因组学等^[93,94]。

目前 scATAC-seq 主要通过两种技术手段来高效获取单细胞全基因组范围染色质开放信息,分别为微流控技术^[92]和 ATAC-seq 组合标签方法 (single-cell combinatorial indexed ATAC-seq, sciATAC-seq)^[95]。单细胞 ATAC-seq 目前最常用的平台为 10× Genomics。其核心的微流控技术原理是将带有 barcode 信息的凝胶珠与转座酶处理后的细胞核混合,包裹在油滴中形成 GEMs (Gel Beads-in-emulsion)。一个特定 barcode 序列标记一个细胞核的所有序列以此区别各个细胞。单细胞 ATAC-seq 微流控技术因其捕获率高、容纳量大和价格相对较低等优点,被广泛应用于探索由表观遗传变化引起的细胞异质性、探索生物标志物、了解基因表达上游的基因调控网络等方面^[93]。另外,2015 年,华盛顿大学的 Jay Shendure 团队开发了 ATAC-seq 组合标签技术 sciATAC-seq^[95]。该方法不必依赖微流控平台,而是利用细胞标签技术对细胞核进行分子标记,通过两次稀释标记-混匀-再稀释标记-再混匀,使单个细胞能够被唯一标记而无需物理分离细胞。以此获得大量单细胞的染色质开放信息。该技术与微流控 scATAC-seq 相比,每个实验可共同测定数百万个单细胞,获得更多的单细胞信息。sciATAC-seq 的局限性是由于数据的稀疏性导致产生的数据集难以分析,无法获得较高的精确性^[96,97]。不过,相信随着数据集分析工具的不断改进,能够有效地提高分析精确性。如最新的分析工具 Scasat 可将开放的染色质信息作为二进制数据进行处理,并在保持数据的二进制特性的同时校正批次效应,使得该工具在分析 scATAC-seq 数据方面优于其他工具^[98]。sciATAC-seq 常用于获取单细胞的染色质调控信息以研究转录调控机制^[99,100]。

单细胞测序也同样发展到了单细胞水平的多组学整合分析,这对于准确解析细胞群中的细胞间差异至关重要。通过单细胞多组学联合分析,如通过联合 scATAC-seq 和 RNA-seq 技术,同时获得单细胞的表观基因组和转录组学信息,能够鉴定引起这些不同细胞表型的致病的顺式和反式作用元件,揭示基因表达调控特异性。如用于研究肿瘤异质性^[93]、揭示造血系统异质性^[101]等。Cusanovich 等^[94]为深入研究细胞亚群的基因表达调控机制,应用组合标签

技术 sciATAC-seq 分析 13 个成年小鼠组织的 10 万个单细胞的基因组范围内染色质的可接近性, 采用了一种 label-transfer 的方法, 整合 RNA-seq 和 ATAC-seq 数据对细胞进行聚类分析, 共鉴定出 30 个主要细胞亚群, 随后确定出 85 种不同的染色质可接近性模式, 以及近 40 万个差异可接近性元件。并使用这些数据将调节元件与其靶基因建立联系, 鉴定出了许多组织特异性的转录因子。之后, 通过将小鼠染色质开放性与人全基因组相关联, 揭示了部分人类遗传病与开放染色质间的潜在关系, 拓展了该研究的意义。整个实验为组织结构, 发育和分化, 各组织器官的调控网络的研究提供了丰富的参考。scATAC-seq 组合标签的方式进一步提高了单细胞 ATAC-seq 的通量, 标签组合作为单细胞基因组学的一种推广策略^[102], sciATAC-seq 具有不可估量的发展潜能。总之, 单细胞多组学联合分析无论是在疾病研究还是了解基因组功能等领域上都有广泛的应用前景。

4 结语与展望

鉴定染色质开放区域并对其进行精确定位, 对表观遗传学的研究具备重要意义。随着高通量测序技术的不断发展, 以 ATAC-seq 为代表的染色质开放区域获取技术, 将能系统发掘全基因组上的启动子、增强子、绝缘子和转录因子等重要调控元件的结合位点, 对深入了解整个基因调控网络具有重要意义。除了应用于寻找组织特异基因、定位核小体外, 将 ATAC-seq 的染色质开放信息进一步整合基因组、转录组、甲基化组等多组学数据, 可更加立体、直观地了解复杂基因之间的相互作用及其对表型的影响效应。

当前, ATAC-seq 技术已成为研究表观遗传调控的重要手段, 其更加简便的操作过程和更易满足的实验材料, 已在染色质开放区获取方面展现出无可比拟的优势和应用潜能。尽管该技术目前的数据分析工具还不够成熟, 但不容置疑的是, 其已成为表观遗传学研究的突破性技术。随着相应实验技术的进一步提高, 可以预期, ATAC-seq 将成为复杂性状遗传解析的研究利器之一, 从而进一步推动人类、小鼠及其他动植物等生命科学领域的稳步向前发展。

参考文献(References):

- [1] Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*, 1974, 184(4139): 868–871. [\[DOI\]](#)
- [2] Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. Structure of the nucleosome core particle at 7 Å resolution. *Nature*, 1984, 311(5986): 532–537. [\[DOI\]](#)
- [3] Zhou YB, Gerchman SE, Ramakrishnan V, Travers A, Muyltermans S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature*, 1998, 395(6700): 402–405. [\[DOI\]](#)
- [4] Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoiyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoiyannopoulos JA. The accessible chromatin landscape of the human genome. *Nature*, 2012, 489(7414): 75–82. [\[DOI\]](#)
- [5] Poirier MG, Bussiek M, Langowski J, Widom J. Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol*, 2008, 379(4): 772–786. [\[DOI\]](#)
- [6] Fedor MJ. Chromatin structure and gene expression. *Curr Opin Cell Biol*, 1996, 4(18): 9384–9388. [\[DOI\]](#)
- [7] John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, Hager GL, Stamatoiyannopoulos JA. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet*, 2011, 43(3): 264–268. [\[DOI\]](#)
- [8] Bell O, Tiwari VK, Thomä NH, Schübeler D. Determinants and dynamics of genome accessibility. *Nat Rev Genet*, 2011, 12(8): 554–564. [\[DOI\]](#)
- [9] Kouzarides T. Chromatin modifications and their function. *Cell*, 2007, 128(4): 693–705. [\[DOI\]](#)
- [10] Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*, 2009, 10(3): 161–172. [\[DOI\]](#)
- [11] Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang ZB, Wei G, Zhao KJ. Dynamic regulation of nucleosome positioning in the human genome. *Cell*,

- 2008, 132(5): 887–898. [DOI]
- [12] Hewish DR, Burgoyne LA. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun*, 1973, 52(2): 504–510. [DOI]
- [13] Scott WA, Wigmore DJ. Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. *Cell*, 1978, 15(4): 1511–1518. [DOI]
- [14] Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SCR. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell*, 1979, 16(4): 797–806. [DOI]
- [15] Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, Weintraub H. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I. *Cell*, 1980, 20(2): 451–460. [DOI]
- [16] Mcghee JD, Wood WI, Dolan M, Engel JD, Felsenfeld G. A 200 base pair region at the 5' end of the chicken adult β -globin gene is accessible to nuclease digestion. *Cell*, 1981, 27(1, Part 2): 45–55. [DOI]
- [17] Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem*, 2003, 57(57): 159–197. [DOI]
- [18] Wu C. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature*, 1980, 286(5776): 854–860. [DOI]
- [19] Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, Bouffard G, Young A, Masiello C, Green ED, Wolfsberg TG, Collins FS. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci USA*, 2004, 101(4): 992–997. [DOI]
- [20] Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 2008, 132(2): 311–322. [DOI]
- [21] Giresi PG, Kim J, Mcdaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*, 2007, 17(6): 877–885. [DOI]
- [22] Kelly TK, Liu YP, Lay FD, Liang GN, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res*, 2012, 22(12): 2497–2506. [DOI]
- [23] He HH, Meyer CA, Hu SS, Chen MW, Zang C, Liu Y, Rao PK, Fei T, Xu H, Long H, Liu XS, Brown M. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods*, 2014, 11(1): 73–78. [DOI]
- [24] Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*, 2014, 7(1): 33. [DOI]
- [25] Cumbie JS, Filichkin SA, Megraw M. Improved DNase-seq protocol facilitates high resolution mapping of DNase I hypersensitive sites in roots in *Arabidopsis thaliana*. *Plant Methods*, 2015, 11(1): 42. [DOI]
- [26] Rizzo JM, Sinha S. Analyzing the global chromatin structure of keratinocytes by MNase-seq. *Methods Mol Biol*, 2014, 1195: 49–59. [DOI]
- [27] Telford DJ, Stewart BW. Micrococcal nuclease: its specificity and use for chromatin analysis. *Int J Biochem*, 1989, 21(2): 127–137. [DOI]
- [28] Chung HR, Dunkel I, Heise F, Linke C, Krobisch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M. The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One*, 2010, 5(12): e15754. [DOI]
- [29] Clark DJ. Nucleosome Positioning, Nucleosome spacing and the nucleosome code. *J Biomol Struct Dyn*, 2010, 27(6): 781–793. [DOI]
- [30] Zentner GE, Henikoff S. Surveying the epigenomic landscape, one base at a time. *Genome Biol*, 2012, 13(10): 250. [DOI]
- [31] Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, Prabhakar S. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol*, 2013, 31(7): 615–622. [DOI]
- [32] Simon JM, Giresi PG, Davis IJ, Lieb JD. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat Protoc*, 2012, 7(2): 256–267. [DOI]
- [33] Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrancois P, Struhl K, Gerstein M, Snyder M. Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci USA*, 2009, 106(35): 14926–14931. [DOI]
- [34] Rhie SK, Schreiner S, Farnham PJ. Defining regulatory elements in the human genome using nucleosome occupancy and methylome sequencing (NOME-Seq). *Methods Mol Biol*, 2018, 1766: 209–229. [DOI]
- [35] Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*, 2019, 20(4): 207–220. [DOI]
- [36] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Met*, 2013, 10(12): 1213–1218. [DOI]
- [37] Wu JY, Huang B, Chen H, Yin QZ, Liu Y, Xiang YL,

- Zhang BJ, Liu BF, Wang QJ, Xia WK, Li WZ, Li YY, Ma J, Peng X, Zheng H, Ming J, Zhang WH, Zhang J, Tian G, Xu F, Chang Z, Na J, Yang XR, Xie W. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*, 2016, 534(7609): 652–657. [DOI]
- [38] Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*, 2015, 109: 21–29. [DOI]
- [39] McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA*, 1950, 36(6): 344–355. [DOI]
- [40] Bucher E, Reinders J, Mirouze M. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr Opin Plant Biol*, 2012, 15(5): 503–510. [DOI]
- [41] Huang CR, Burns KH, Boeke JD. Active transposition in genomes. *Annu Rev Genet*, 2012, 46: 651–675. [DOI]
- [42] Berg DE, Davies J, Allet B, Rochaix JD. Transposition of R factor genes to bacteriophage lambda. *Proc Natl Acad Sci USA*, 1975, 72(9): 3628–3632. [DOI]
- [43] Rothstein SJ, Jorgensen RA, Postle K, Reznikoff WS. The inverted repeats of Tn5 are functionally different. *Cell*, 1980, 19(3): 795–805. [DOI]
- [44] Lovell S, Goryshin IY, Reznikoff WR, Rayment I. Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Biol*, 2002, 9(4): 278–281. [DOI]
- [45] Auerswald EA, Ludwig G, Schaller H. Structural analysis of Tn5. *Cold Spring Harb Symp Quant Biol*, 1981, 45 Pt 1: 107. [DOI]
- [46] Berg DE. Julian Davies and the discovery of kanamycin resistance transposon Tn5. *J Antibiot*, 2017, 70(4): 339–346. [DOI]
- [47] Reznikoff WS. The TN5 transposon. *Annu Rev Microbiol*, 1993, 47(1): 945–963. [DOI]
- [48] Reznikoff WS, Bhasin A, Davies DR, Goryshin IY, Mahnke LA, Naumann T, Rayment I, Steiniger-White M, Twining SS. Tn5: A molecular window on transposition. *Biochem Biophys Res Commun*, 1999, 266(3): 729–734. [DOI]
- [49] Zhou M, Reznikoff WS. Tn5 transposase mutants that alter DNA binding specificity. *J Mol Biol*, 1997, 271(3): 362–373. [DOI]
- [50] York D, Reznikoff WS. DNA binding and phasing analyses of Tn5 transposase and a monomeric variant. *Nucleic Acids Res*, 1997, 25(11): 2153–2160. [DOI]
- [51] Steiniger-White M, Reznikoff WS. The C-terminal alpha helix of Tn5 transposase is required for synaptic complex formation. *J Biol Chem*, 2000, 275(30): 23127–23133. [DOI]
- [52] Bhasin A, Goryshin IY, Steiniger-White M, York D, Reznikoff WS. Characterization of a Tn5 pre-cleavage synaptic complex. *J Mol Biol*, 2000, 302(1): 49–63. [DOI]
- [53] Kale SB, Landree MA, Roth DB. Conditional RAG-1 mutants block the hairpin formation step of V(D)J recombination. *Mol Cell Biol*, 2001, 21(2): 459–466. [DOI]
- [54] Naumann TA, Reznikoff WS. Trans catalysis in Tn5 transposition. *Proc Natl Acad Sci USA*, 2000, 97(16): 8944–8949. [DOI]
- [55] Lisa AMB, Goryshin IY, Reznikoff WS. A mechanism for Tn5 inhibition. *J Biol Chem*, 1999, 274(1): 86. [DOI]
- [56] Bhasin A, Goryshin IY, Reznikoff WS. Hairpin formation in Tn5 transposition. *J Biol Chem*, 1999, 274(52): 37021–37029. [DOI]
- [57] Mizuuchi K, Adzuma K. Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: Evidence for a one-step transesterification mechanism. *Cell*, 1991, 66(1): 129–140. [DOI]
- [58] Mizuuchi K. Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu Rev Biochem*, 1992, 61: 1011–1051. [DOI]
- [59] Crellin P, Chalmers R. Protein-DNA contacts and conformational changes in the Tn10 transpososome during assembly and activation for cleavage. *EMBO J*, 2001, 20(14): 3882–3891. [DOI]
- [60] Goryshin IY, Reznikoff WS. Tn5 *in vitro* transposition. *J Biol Chem*, 1998, 273(13): 7367–7374. [DOI]
- [61] Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, Mackenzie AP, Caruccio NC, Zhang X, Shendure J. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol*, 2010, 11(12): R119. [DOI]
- [62] Reznikoff WS. Transposon Tn5. *Annu Rev Genet*, 2008, 42: 269–286. [DOI]
- [63] Davies DR, Mahnke Braam L, Reznikoff WS, Rayment I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. *J Biol Chem*, 1999, 274(17): 11904–11913. [DOI]
- [64] Peterson G, Reznikoff W. Tn5 transposase active site mutations suggest position of donor backbone DNA in synaptic complex. *J Biol Chem*, 2003, 278(3): 1904–1909. [DOI]
- [65] Reznikoff WS. Tn5 as a model for understanding DNA transposition. *Mol Microbiol*, 2003, 47(5): 1199–1206. [DOI]
- [66] Sakamoto H, Thiberge S, Akerman S, Janse CJ, Carvalho TG, Ménard R. Towards systematic identification of Plasmodium essential genes by transposon shuttle mutagenesis.

- nesis. *Nucleic Acids Res*, 2005, 33(20): e174. [DOI]
- [67] Caruccio N. Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. *Methods Mol Biol*, 2011, 733: 241–255. [DOI]
- [68] Chen CY, Xing D, Tan LZ, Li H, Zhou GY, Huang L, Xie XS. Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). *Science*, 2017, 356(6334): 189–194. [DOI]
- [69] Chen XQ, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, Doudna JA, Greenleaf WJ, Liphardt JT, Chang HY. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods*, 2016, 13(12): 1013–1020. [DOI]
- [70] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf W. Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nat Methods*, 2013, 10(12): 1213–1218. [DOI]
- [71] Lu Z, Hofmeister BT, Vollmers C, Dubois RM, Schmitz RJ. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res*, 2017, 45(6): e41. [DOI]
- [72] Deal RB, Henikoff S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev Cell*, 2010, 18(6): 1030–1040. [DOI]
- [73] Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, Queitsch C, Bailey-Serres J, Sinha N, Brady SM, Deal RB. Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell*, 2018, 30(1): 15–36. [DOI]
- [74] Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*, 2017, 14(10): 959–962. [DOI]
- [75] Sos BC, Fung HL, Gao DR, Osothpraprop TF, Kia A, He MM, Zhang K. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. *Genome Biol*, 2016, 17(1): 20. [DOI]
- [76] Tan L, Ke Z, Tomblin G, Macoretta N, Hayes K, Tian X, Lv R, Ablaeva J, Gilbert M, Bhanu NV, Yuan Z F, Garcia BA, Shi YG, Shi Y, Seluanov A, Gorbunova V. Naked mole rat cells have a stable epigenome that resists iPSC reprogramming. *Stem Cell Reports*, 2017, 9(5): 1721–1734. [DOI]
- [77] Scharer CD, Blalock EL, Barwick BG, Haines RR, Wei C, Sanz I, Boss JM. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. *Sci Rep*, 2016, 6(1): 27030. [DOI]
- [78] Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, Halder G, Aerts S. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet*, 2015, 11(2): e1004994. [DOI]
- [79] Liu YJ, Zhang F, Liu HD, Sun X. The application of next-generation sequencing techniques in studying transcriptional regulation in embryonic stem cells. *Hereditas(Beijing)*, 2017, 39(8): 717–725.
刘亚军, 张峰, 刘宏德, 孙啸. 下一代测序技术在干细胞转录调控研究中的应用. *遗传*, 2017, 39(8): 717–725. [DOI]
- [80] Quillien A, Abdalla M, Yu J, Ou JH, Zhu LJ, Lawson ND. Robust identification of developmentally active endothelial enhancers in zebrafish using FANS-assisted ATAC-Seq. *Cell Rep*, 2017, 20(3): 709–720. [DOI]
- [81] Ackermann AM, Wang ZP, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab*, 2016, 5(3): 233–244. [DOI]
- [82] Garcia E, Hayden A, Birts C, Britton E, Cowie A, Pickard K, Mellone M, Choh C, Derouet M, Duriez P, Noble F, White MJ, Primrose JN, Strefford JC, Rose-Zerilli M, Thomas GJ, Ang Y, Sharrocks AD, Fitzgerald RC, Underwood TJ. Authentication and characterisation of a new oesophageal adenocarcinoma cell line: MFD-1. *Sci Rep*, 2016, 6: 32417. [DOI]
- [83] Ho YT, Shimbo T, Wijaya E, Ouchi Y, Takaki E, Yamamoto R, Kikuchi Y, Kaneda Y, Tamai K. Chromatin accessibility identifies diversity in mesenchymal stem cells from different tissue origins. *Sci Rep*, 2018, 8(1): 17765. [DOI]
- [84] Nelson AC, Mould AW, Bikoff EK, Robertson EJ. Mapping the chromatin landscape and Blimp1 transcriptional targets that regulate trophoblast differentiation. *Sci Rep*, 2017, 7(1): 6715–6793. [DOI]
- [85] Scott-Browne JP, López-Moyado IF, Trifari S, Wong V, Chavez L, Rao A, Pereira RM. Dynamic changes in chromatin accessibility occur in CD8⁺ T cells responding to viral infection. *Immunity*, 2016, 45(6): 1327–1340. [DOI]
- [86] Starks RR, Biswas A, Jain A, Tuteja G. Combined

- analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. *Epigenetics Chromatin*, 2019, 12(1): 16. [DOI]
- [87] Rajbhandari P, Thomas B J, Feng A, Hong C, Wang JX, Vergnes L, Sallam T, Wang B, Sandhu J, Seldin M M, Lusis AJ, Fong LG, Katz M, Lee R, Young SG, Reue K, Smale ST, Tontonoz P. IL-10 signaling remodels adipose chromatin architecture to limit thermogenesis and energy expenditure. *Cell*, 2018, 172(1–2): 218–233.e17. [DOI]
- [88] Denny SK, Yang D, Chuang CH, Brady JJ, Lim JS, Grüner BM, Chiou S, Schep AN, Baral J, Hamard C, Antoine M, Wislez M, Kong CS, Connolly AJ, Park K, Sage J, Greenleaf WJ, Winslow MM. Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell*, 2016, 166(2): 328–342. [DOI]
- [89] Wang MJ, Wang PC, Lin M, Ye ZX, Li GL, Tu LL, Shen C, Li JY, Yang QY, Zhang XL. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants*, 2018, 4(2): 90–97. [DOI]
- [90] Mas G, Blanco E, Ballaré C, Sansó M, Spill YG, Hu D, Aoi Y, Le Dily F, Shilatifard A, Marti-Renom MA, Di Croce L. Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat Genet*, 2018, 50(10): 1452–1462. [DOI]
- [91] Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, Vernot B, Cheng JB, Thurman RE, Sandstrom R, Haugen E, Heimfeld S, Murry CE, Akey JM, Stamatoyannopoulos JA. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*, 2013, 154(4): 888–903. [DOI]
- [92] Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 2015, 523(7561): 486–490. [DOI]
- [93] Liu LQ, Liu YC, Quintero A, Wu L, Yuan Y, Wang MY, Cheng MN, Leng LZ, Xu LQ, Dong GY, Li R, Liu Y, Wei XY, Xu JS, Chen XW, Lu HR, Chen DS, Wang QL, Zhou Q, Lin XX, Li GB, Liu SP, Wang Q, Wang HR, Fink JL, Gao ZL, Liu X, Hou Y, Zhu SD, Yang HM, Ye YM, Lin G, Chen F, Herrmann C, Eils R, Shang ZC, Xu X. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat Commun*, 2019, 10(1): 470. [DOI]
- [94] Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, Dewitt WS, Lee C, Regalado SG, Read DF, Steemers FJ, Disteche CM, Trapnell C, Shendure J. A single-cell atlas of *in vivo* mammalian chromatin accessibility. *Cell*, 2018, 174(5): 1309–1324. [DOI]
- [95] Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, 2015, 348(6237): 910–914. [DOI]
- [96] Zamanighomi M, Lin ZL, Daley T, Chen X, Duren Z, Schep A, Greenleaf WJ, Wong WH. Unsupervised clustering and epigenetic classification of single cells. *Nat Commun*, 2018, 9(1): 2410. [DOI]
- [97] Schep AN, Wu BJ, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*, 2017, 14(10): 975–978. [DOI]
- [98] Baker SM, Rogerson C, Hayes A, Sharrocks AD, Rattray M. Classifying cells with Scasat, a single-cell ATAC-seq analysis tool. *Nucleic Acids Res*, 2019, 47(2): e10. [DOI]
- [99] Preissl S, Fang RX, Huang H, Zhao Y, Raviram R, Gorkin DU, Zhang YX, Sos BC, Afzal V, Dickel DE, Kuan S, Visel A, Pennacchio LA, Zhang K, Ren B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci*, 2018, 21(3): 432–439. [DOI]
- [100] Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, Trapnell C, Shendure J, Furlong EEM. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, 2018, 555(7697): 538–542. [DOI]
- [101] Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell*, 2018, 173(6): 1535–1548. [DOI]
- [102] Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J. Massively multiplex single-cell Hi-C. *Nature Methods*, 2017, 14(3): 263–266. [DOI]