



染色质可及性分析的研究进展*

许兰 任立成**

(海南医学院生物学教研室, 海口 571199)

摘要 在细胞核内, 染色质可及性模式会随着外部刺激和发育线索的改变而发生动态变化。染色质可及性重构对于基因表达调控至关重要, 在建立和维持细胞特性等方面发挥着重要作用。因此开展染色质可及性的研究对染色质功能上的三维解析具有十分重要的意义。近几年, 随着高通量测序技术的进步以及测序成本的降低, 基于高通量测序技术的染色质可及性分析方法得到了迅速发展。目前观察和分析全基因组染色质开放与否的常见技术主要有脱氧核糖核酸酶I超敏位点测序(DNase-seq)、微球菌核酸酶测序(MNase-seq)、甲醛辅助分离调控元件测序(FAIRE-seq)以及转座酶可及性测序(ATAC-seq)。本文比较了这4种染色质可及性分析技术的优缺点, 详细介绍了它们的原理及主要实验流程, 并简要讨论了它们的发展及相关技术的应用, 期望通过这些互补的方法为染色质分析领域的未来发展提供一些借鉴和思路。

关键词 染色质可及, 脱氧核糖核酸酶I超敏位点测序, 微球菌核酸酶测序, 甲醛辅助分离调控元件测序, 转座酶可及性测序

中图分类号 Q2, Q3

DOI: 10.16476/j.pibb.2021.0313

真核生物的基因组DNA并不是裸露的, 而是在细胞核中与组蛋白结合形成核小体(构成染色质的基本结构单位)。核小体核心由147 bp的DNA组成, 以左手超螺旋的方式包裹着一个球状蛋白八聚体, 该八聚体由4种核心组蛋白H₂A、H₂B、H₃和H₄各两分子组成^[1]。每个核心组蛋白都有一个与DNA结合的结构域和一个无序的N端尾巴。核小体的核心颗粒再由10~80 bp左右的游离DNA与组蛋白H₁共同连接形成串珠式的染色质细丝, 染色质细丝通过紧密折叠并高度压缩形成螺旋化的染色体结构^[2]。这些高度螺旋化的染色体结构在复制和转录时需要暴露出DNA序列, 才能使转录因子和一些调控元件与之结合。这种允许启动子、增强子、绝缘子、沉默子等顺式调控元件和反式作用因子可以接近的特性, 就称为染色质可及性(chromatin accessibility), 也称为染色质开放区(图1)。

在真核生物的基因组中, 核小体分布较少的区域, 染色质呈松散状态, 压缩程度低, 有利于转录因子等调控元件与这些区域的启动子、增强子结合及相互作用, 进而调控基因表达。而核小体结构致密的区域, 其与基因表达相关的结构区域相对封闭, 从而抑制了基因的表达。因此, 在开放的染色

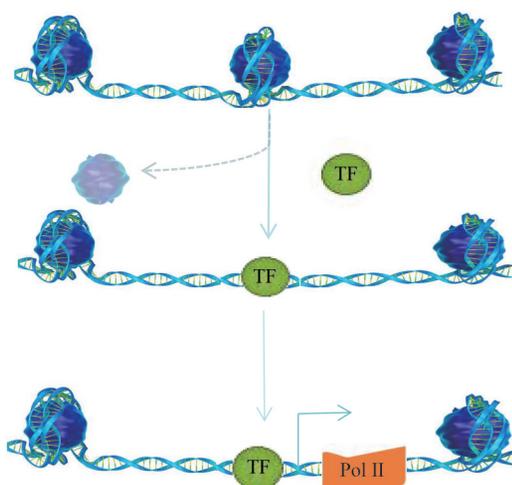


Fig. 1 The dynamic changes in chromatin accessibility

图1 染色质可及性的动态变化

TF, 转录因子(transcription factor); Pol II, RNA聚合酶II。

* 海南省自然科学基金(820RC638)和国家自然科学基金(31660318, 31260269)资助项目。

** 通讯联系人。

Tel: 0898-66893739, E-mail: lcheng08@126.com

收稿日期: 2021-10-14, 接受日期: 2021-12-17

质位点常常分布有转录活跃的基因以及各种顺式调控元件。研究显示, 染色质开放区占基因组DNA总序列的2%~3%, 而90%以上的开放区域与转录因子的结合有关^[2-4]。此外, 有研究表明, 利用能量依赖的染色质重塑酶改变核小体的位置和动态, 就可以选择性地关闭或打开细胞中的基因, 从而控制细胞的发育和功能^[5]。

近年来, 获取可及性染色质图谱的技术已经以越来越高的精确度进行了改进, 极大地拓宽了染色质可及性分析的应用空间, 将生物材料需求降低到临床可用的水平, 并提高了这些检测方法的分辨率。这些技术的改进使得在单个细胞、低输入样本和异质细胞群体中识别假定的调控区成为可能。虽然启动子通常在广泛的细胞类型中都是结构可及的, 但远端增强子的可及性往往受到细胞类型的限制。了解这些调控结构域是如何随着细胞在发育阶段和激活状态之间的转变而动态建立的, 以及调控元件如何塑造基因表达程序的生物物理机制, 将是表观遗传学研究的重点。

1 全基因组水平研究染色质可及性的方法

对于染色质可及性的研究, 始于发现染色质上不同区域的DNA序列对DNase酶的敏感性存在差异, 失去核小体保护的DNA序列相对于缠绕在核小体上的DNA序列更容易与核酸酶结合并切割, 进而表现出对DNase酶的高度敏感性。这些对脱氧核糖核酸酶I (DNase I) 内切酶高度敏感的基因组区域就称为DNase I超敏位点 (DNase I hypersensitive sites, DHSs)。DHSs位点通常被认为是开放染色质区, 它可以随时被调控蛋白结合, 从而显示出真核基因组中富含顺式调控元件的区域。随着测序技术越来越成熟, 价格越来越合理, 将传统DNase I足迹法与下一代测序技术结合, 在全基因组范围内识别与DNA相互作用的调控元件成为了可能^[2, 6]。

目前研究染色质可及性的方法主要是将酶切法或物理化学方法与下一代测序技术结合, 检测染色质开放或受保护的区域。常用的方法有脱氧核糖核酸酶I超敏位点测序 (DNase I hypersensitive site sequencing, DNase-seq)、微球菌核酸酶测序 (micrococcal nuclease sequencing, MNase-seq)、甲醛辅助分离调控元件测序 (formaldehyde-assisted isolation of regulatory elements sequencing, FAIRE-seq) 以及转座酶可及性测序 (assay for targeting

accessible-chromatin with high-throughout sequencing, ATAC-seq)。下面将对这4种染色质可及性分析技术的原理及主要实验流程进行简要描述。

1.1 DNase-Seq

DNase I是一种核酸内切酶, 可以对单链或双链DNA进行非特异性的消化和切割, 将基因组DNA切割成长短不一的片段。如果DNA上面有蛋白质结合, 蛋白质覆盖的序列就不能被DNase I切割。正是因为这个特点, 通过确定DNase I消化后留下的DNA片段, 就能够获得开放染色质的信息。近年来的研究发现, DNase I敏感性与基因转录有关, 具有转录活性基因的DNA区域更容易被DNase I消化, 形成DHSs位点。在结构上, 含有DHSs位点的染色质甲基化程度低, 染色质疏松, 对DNase I的敏感度是不含此类特征基因的100倍^[6-10]。

传统上, DHSs的鉴定一直是基于DNA印迹法 (Southern blotting) 和间接末端标记, 涉及许多费时费力的步骤, 不适用于全基因组分析^[11]。为了进一步提高该方法的效率和分辨率, 科学家提出将DNase I酶足迹法与高通量测序技术结合起来在全基因组范围内检测蛋白质结合位点的DNase-Seq技术。

DNase-seq试验流程可概括如下 (图2)。首先, 裂解液处理细胞, 分离细胞核, 利用DNase I对细胞核进行消化, DNase I优先切割核小体缺失的DNA序列, 也就是调控元件与之结合的位点。再分别使用RNase和蛋白酶K (proteinase K) 降解RNA和蛋白质。由于DNase I切割DNA后会留下单链突起, 还需用T4 DNA聚合酶来钝化末端以便后续测序接头的连接。之后进行酚氯仿纯化和乙醇沉淀DNA, 并通过凝胶电泳分离, 选择50~100 bp大小的DNA片段作为构建文库的模板。50~100 bp的片段比与核小体相关的147 bp的DNA要短, 不能跨越核小体, 与较长的片段相比, 此类片段在转录因子 (transcription factor, TF) 结合位点的富集度更高, 对于鉴定TF结合位点更有效。最后对筛选的短DNA片段进行高通量测序, 就可以在一次实验中获得全基因组范围内有效的DHSs位点^[6, 12-13]。

DNase I在切割DNA时存在着一种固有的偏好性, 大约每隔10 bp左右在核小体周围切割DNA的小沟序列, 这种效应被认为与小槽的宽度有关。同

时, 由于DNase I的活性随制造商、批次的不同而改变。为了获得高质量的DNase-seq数据, 通常在使用新的细胞类型或调整起始细胞数量时, 必须要滴定DNase I的浓度, 调整酶的用量, 以实现DNA的适当消化, 防止完全消化开放的染色质。这样就使实验操作涉及许多繁琐的酶滴定步骤。对于DNase-Seq存在的一些局限, 最近的一种修改使用0.1% 甲醛处理细胞以辅助鉴定, 称为XL-DNase-Seq。这种温和的交联方式有助于提取新的基因调控网络, 涉及以前无法检测到的转录因子结合位点。另一种改进技术为单细胞DNase-seq (scDNase-seq), scDNase-seq只需要对成百上千的新鲜或固定细胞进行批量分析, 不需要核分离, 优化了DNase I酶切, 最大限度地减少了样本损失^[11-13]。

总体而言, DNase-seq是一种简单可行的方法, 它可以在不需要表观遗传信息先验知识的情况下, 在整个基因组和来自测序物种的任何细胞类型中识别活跃的调控元件。它还可以在单碱基水平观察转录因子的占据情况^[9]。

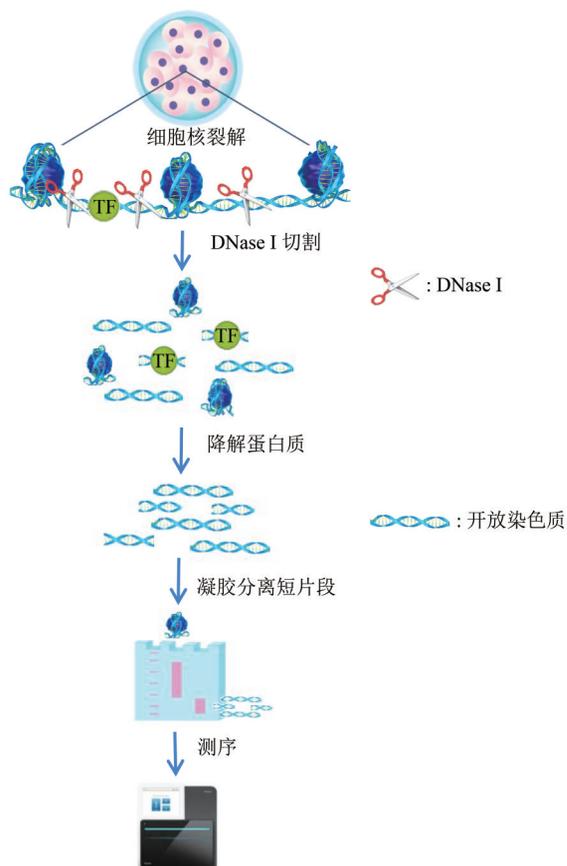


Fig. 2 Workflows of DNase-seq technology
图2 DNase-seq技术流程图

1.2 MNase-seq

微球菌核酸酶 (MNase) 于 20 世纪 60 年代从金黄色葡萄球菌中分离出来, 最初一直被用于以低通量的方式研究染色质的结构。MNase既可以作为内切酶来切割核小体连接区的DNA, 也可以作为外切酶来降解不受蛋白质保护的切割产物DNA。它优先消化核小体之间裸露的DNA, 从染色质中释放核小体, 并富集受核小体保护的DNA片段。由于核小体DNA被MNase切割的效率低于核小体间的DNA, 因此该酶成为分离单个核小体DNA的理想酶。随着基因组技术的兴起, MNase消化与下一代测序技术相结合 (MNase-seq), 在全基因组范围内确定核小体的位置和占有率, 从而间接反应出染色质的可及性区域^[11, 14-16]。

大多数的MNase-Seq方案是以甲醛为交联剂进行的, 甲醛交联被用来捕捉蛋白质与DNA的相互作用, 避免在MNase消化过程中核小体位置发生移动。这种交联使结合蛋白质的DNA受到保护不被MNase消化。交联完成后, 细胞用MNase进行裂解和消化, 经RNase和蛋白酶K处理后解交联。酚氯仿提取分离DNA, 并在琼脂糖凝胶上进行分析, 以确保获得150 bp左右的单个完整核小体DNA的条带。最后对得到的未消化的DNA进行高通量测序, 分析MNase-seq测序数据, 并将其映射到参考基因组, 获得全基因组核小体定位的精确图谱 (图3)^[11-12, 15, 17-18]。

MNase-seq技术的关键在于确定MNase酶的消化水平, 即精确控制酶的用量和消化时间, 获得主要由单个核小体DNA组成的染色质片段。基因组的不同区域可能被MNase以不同的速率消化。具体地说, 在消化开始时, 所有的核小体都随着消化水平的提高而增加 (更多的核小体从染色质中释放出来), 在新的消化阶段, 一些核小体随着消化水平的提高而减少 (释放的核小体开始被MNase破坏)^[16]。况且不同的细胞存在着不同的核小体组织特征, 无法先验确定合适的酶量, 因此, 就需要大量实验来探索严格的酶解条件。

事实上, MNase消化染色质时, 在A或T上游切割DNA的速度大约是G或C上游的30倍。这可能导致富含A/T序列的核小体DNA过度消化, 获得的核小体分布图谱人为地偏向于基因组上富含G/C的位置, 从而低估了富含A/T的核小体^[11, 16]。为了克服这一缺点, 许多研究者对这一技术进行了改进, 可以最大限度地减少由于MNase切割偏好

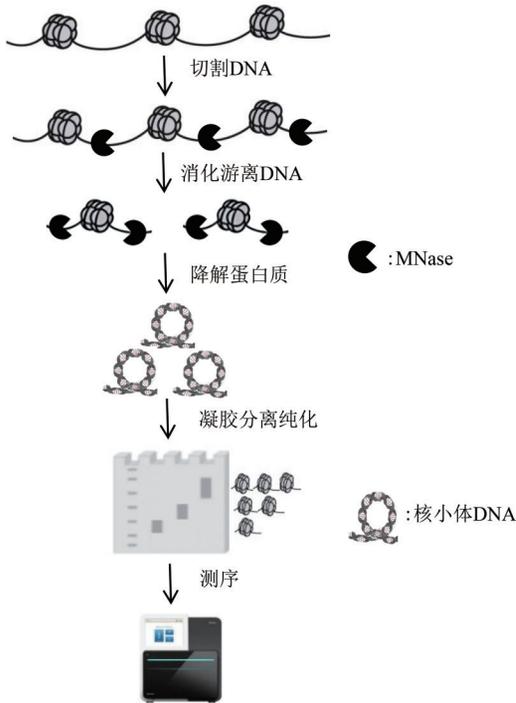


Fig. 3 Workflows of MNase-seq technology
图3 MNase-seq技术流程图

造成的数据偏差, 例如, Ramani V 实验室^[19]介绍了一种用核酸酶消化染色质的单链测序文库制备方法 (MNase-SSP), 可同时绘制 TF 和核小体位置, 与现有的 MNase-seq 方案相比, MNase-SSP 显示出低的序列偏差并显著富集了短的 DNA 片段。

总体而言, MNase-seq 用于核小体定位具有较高的分辨率, MNase 可以高效去除连接 DNA, 获得长度较均一的核小体 DNA 片段, 从而得到精确的核小体位置信息。适用于多种类型的细胞, 已广泛应用于绘制从酵母到人类等许多生物体的核小体图谱。

1.3 FAIRE-seq

FAIRE 技术是直接探测基因组核小体缺失区域的一种简单方法。它可以直接富集活性染色质区域, 而且不需要细胞的初始准备和费力的酶滴定过程。FAIRE 技术被认为是通过利用组蛋白和转录因子与 DNA 交联之间的差异进行有效工作的。由于组蛋白中有大量的氨基酸残基与甲醛发生反应, 而且它们与 DNA 有广泛的接触, 因此组蛋白比转录因子能更有效地与 DNA 交联^[20]。

在这种方法中, DNA 通过交联剂甲醛共价结合到染色质蛋白质上, 核小体紧密堆积的染色质区

域会有丰富的蛋白质与 DNA 交联, 而没有或有较少核小体的 DNA 区域几乎没有交联的蛋白质。通过超声波将染色质剪切成碎片, 苯酚-氯仿抽取后, 与蛋白质未交联的游离 DNA 留在水相, 而与蛋白质交联的 DNA 将被捕获在水相和有机相之间, 提取到水相中的 DNA 可以通过下一代测序技术 (FAIRE-Seq) 进行分析, 对染色质的开放区进行鉴定 (图 4)^[21-23]。如果只分析少量位点, 则可以通过基因组位点特异性定量 PCR (FAIRE-qPCR) 来分析^[24]。

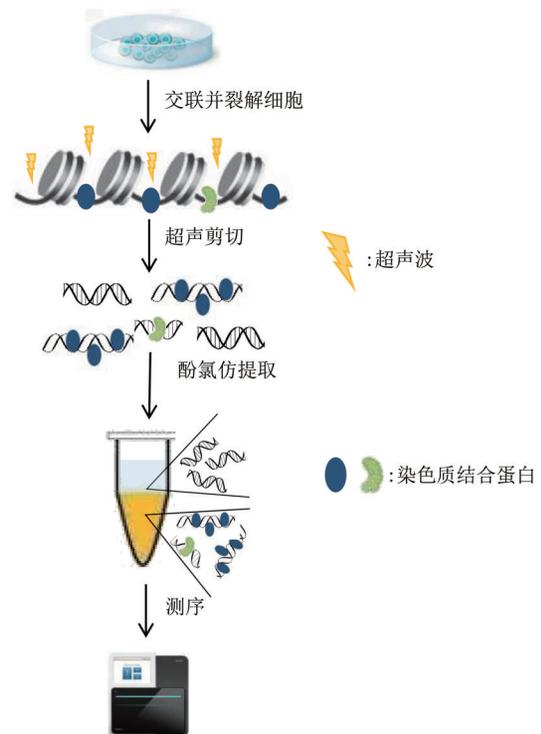


Fig. 4 Workflows of FAIRE technology
图4 FAIRE技术流程图

与其他方法相比, FAIRE 不依赖于酶活性, 使用超声波剪切染色质, 避免了酶切割存在的偏好性。FAIRE 实验的成功在很大程度上取决于适当的甲醛固定效率。这种固定效率根据细胞的渗透性、成分和各种其他生理因素不同而改变。对于大多数培养细胞来说, 通常只需要 5~7 min 的固定时间, 真菌和植物可能需要显著增加固定时间, 才能让甲醛到达所有组织样本中的细胞^[11, 25]。FAIRE 方法依赖于超声波对 DNA 的最佳剪切。如果片段太长, 任何无核小体的区域都会被相邻的染色质结合的蛋白质所掩盖。如果片段太小, 深度测序的分析就会

变得非常困难。因此，超声处理后DNA片段的理想大小在200~300 bp，相当于1~2个核小体的大小^[21]。

总的来说，FAIRE-seq是一种程序简单、成功率很高、易于实现的识别染色质开放区的方法。与MNase和DNase I方法相比需要更少的细胞量，不需要酶介导的DNA切割，避免了酶切割存在的偏好性以及繁琐的酶滴定优化实验。对全基因组范围内分离染色质开放区提供了一种无偏性的分析。

1.4 ATAC-Seq

基于转座酶和高通量测序的染色质分析技术(ATAC-seq)是一种新的表观遗传学技术，是近年来分析全基因组染色质可及性最有效的方法之一。它是在全基因组水平上探测Tn5转座酶到DNA序列来定位染色质的可及性^[26]。

在ATAC-seq方案中，包括细胞核制备、转座和扩增3个步骤。首先，将待检查的组织或细胞制备成均一的单细胞悬液，在裂解缓冲液中进行孵育，分离细胞核。其次，将携带已知DNA序列标签的Tn5转座酶人工添加到500~50 000个分离的细胞核中，Tn5转座酶同时切割DNA并插入DNA序列标签。由于空间位阻，大多数的DNA标签序列被整合到染色质开放区域。将带有转座酶标记的DNA序列进行扩增，以产生用于测序的DNA文库，进行高通量测序(图5)。ATAC-seq基本上依赖于酶反应构建的文库，即转座酶与样品染色质的反应程度，这可能会受到酶的用量、细胞核的数量以及染色质折叠程度和结构的影响。因此，在测序前需要对ATAC-seq文库的质量进行控制，以确保文库浓度达到测序标准^[11, 27]。

由于其简单、快速和可重复性，ATAC-seq已迅速成为全基因组范围内染色质可及性分析的重要方法。与其他检测技术相比，ATAC-seq具有许多优点。第一，转座酶效率高。通过简单的酶反应就可以实现DNA的断裂、末端修复和接头连接反应，通常在不到2 h内就能够完成对10 000~20 000个细胞的处理^[14]。第二，灵敏度高。ATAC-seq适合细胞数量有限的研究样本，约500个细胞就能够进行实验。但是，ATAC-seq技术也存在一些局限。Tn5转座酶进行剪切时，由于每个DNA片段两端的接头是随机的，这导致一段序列两端接头相同的可能性为50%，而只有连接不同的接头序列才能够进行富集扩增和测序，从而会产生一半不能用于富集、扩增和测序的序列。其次，Tn5转座酶倾向于结合

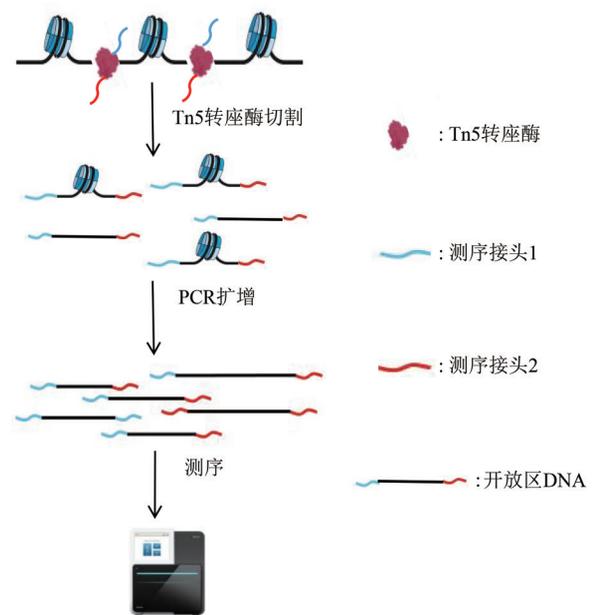


Fig. 5 Workflows of ATAC-seq technology

图5 ATAC-seq技术流程图

和切割转录因子结合区，这导致部分转录因子结合位点信息丢失。此外，由于线粒体DNA的存在，ATAC-seq获得的数据不可避免地包含一些线粒体读数^[28]。通过改善裂解条件(Omni-ATAC)、流式细胞术纯化细胞核、或在实验之前应用CRISPR技术切割线粒体核糖体DNA，这一限制因素得到显著解决^[10]。

总体而言，ATAC-Seq技术简单快捷，只需要更短的准备时间和更少的细胞数量就可以高质量地分析染色质的可及性，识别基因组中活跃的调控序列。ATAC-Seq已用于确定给定细胞环境中的基本可及染色质区域，以及确定两个细胞状态之间的差异可及区域。它迅速应用于干细胞、早期胚胎和各种肿瘤细胞基因表达的动态研究，为揭示染色体可及性、胚胎发育和肿瘤发生的表观遗传机制和潜在的疾病生物标志提供了有意义的见解。

1.5 单细胞染色质可及性

随着高通量测序技术的发展和染色质可及性研究方案的不断完善，在单细胞水平上分析染色质可及性成为可能。一般来说，分析单细胞开放染色质图谱主要有3种方法，包括scDNase-seq、scMNase-seq和scATAC-seq。ScDase-seq和scMNase-seq使用传统的文库制备程序，利用DNase或MNase在开放的染色质区域消化DNA，

并将测序接头连接到消化的短DNA片段上进行测序, 通过苯酚-氯仿提纯DNA。这些实验增加了繁琐的操作步骤。而scATAC-seq只需将Tn5标记的染色质与流式细胞术相结合以分离单个细胞核或细胞。由于其简单性和敏感性, scATAC-seq成为分析染色质可及性最广泛的使用方法之一^[29]。

基于微流控的scATAC-seq和单细胞组合索引ATAC-seq (sciATAC-seq) 是目前采用最广泛的两种技术。前者是将标有特定标签的磁珠, 混有Tn5转座酶的细胞核悬浮液通过微流控系统形成孤立液滴, 每个液滴都包含一个细胞核和一个微珠, 从而实现单细胞的分离建库。由于其重复性和相对易用性, 已经成为单细胞数据生成的一种强有力的方法。sciATAC-Seq技术的核心是细胞或细胞核的重复汇集和分配, 并在每一步进行DNA片段的标记。在简单的sciATAC-Seq方案中, 细胞核被分配到一个96孔板中, 每孔含有一个唯一的索引, 用来标记这个孔中的转座酶, 然后将每孔细胞核汇集在一起并以低浓度重新分配到新的96孔板中, 再次进行标记, 随后对这些孔中的反应进行索引PCR, 为每个细胞生成唯一的标签组合^[30-31]。单细胞ATAC-seq能够检测出混杂样品测序所无法得到的异质性信息, 深度挖掘细胞异质性表观遗传调控机理。

单细胞方法的最新进展正在推动着在同一单细胞上同时进行多组学整合分析。多种方法已经被公布用于同时进行scATAC-seq和转录组分析。Taavitsainen等^[32]通过单细胞ATAC和RNA测序数据揭示了前列腺癌复发相关的预先存在和持续存在的细胞亚群。Pierce等^[33]将CRISPR技术与scATAC-seq (Spear-ATAC) 相结合, 分析了414个sgRNA敲除群体的104 592个细胞的图谱, 揭示了表观遗传对癌细胞调节的时间动态和转录因子结合谱之间的关联。这些分析使人们能够确定直接调控这些基因的假定致癌转录因子, 为个性化治疗靶点开辟了新的途径^[34]。总之, 单细胞多组学联合分析在疾病研究和理解基因组功能领域上都具有广阔的应用前景。

2 染色质可及性的应用及研究意义

基因的表达调控可以通过改变染色质的拓扑结构或染色质修饰来实现。当组蛋白和DNA之间的结合力增加时, 染色质浓缩成闭合构象并阻止转录因子进入DNA, 导致基因抑制。相反, 当组蛋白

与DNA之间的结合力降低时, 染色质解聚形成“开放”构象, 使转录因子可接近DNA, 从而导致转录激活。随着组织形态和功能的发展, 细胞的表观遗传状态和基因表达谱发生动态变化, 并随着环境的变化而不断变化, 调节染色质的可及性, 并调节与形态发生和谱系指定相关的基因表达。染色质这种动态的重塑变化与胚胎发育、细胞衰老、肿瘤发生、免疫、细胞命运决定过程密切相关^[35-37]。

Ranzoni等^[38]应用单细胞RNA测序(scRNA-seq)和单细胞转座酶可及染色质测序(scATAC-seq)技术, 对来自胎肝和骨髓的8 000多个人类免疫表型血细胞进行了测序, 鉴定了造血干细胞(HSCs)/多能祖细胞(MPPs)下游的3个高度增殖寡能祖细胞。Muto等^[39]使用单核ATAC(snATAC-seq)和RNA(snRNA-seq)测序技术分析成人肾脏细胞的状态和功能。分析表明, 大多数差异可及染色质区域定位于启动子, 其中很大一部分与差异表达基因密切相关。这种配对分析方法提高了检测肾脏内独特细胞状态的能力, 并重新定义了成人肾脏细胞的异质性。

Liang等^[40]利用ATAC-seq分析了87个供者的人类神经前体细胞及其分化神经元后代的染色质可及性数量性状位点(QTL)。通过整合特定细胞类型的染色质可及性QTL和与脑相关的全基因组关联数据, 精细绘制和识别了非编码神经精神障碍风险位点潜在的调控元件。

此外, 染色质的动态结构还和许多其他生物学功能有关。比如, 环境污染等外界因素导致DNA甲基化或组蛋白修饰发生改变, 引起一系列的疾病状况。这说明, 染色质结构的改变与人类健康密切相关。在全基因组层面对染色质可及性有了全局的认识后, 就可以破译基因转录调控中的有效调控元件, 为深入认识疾病的致病机制提供新的思路。

3 总结与展望

在上述4种方法中, 除FAIRE-seq使用超声波物理破碎基因组DNA外, MNase-seq、DNase-seq和ATAC-seq都是基于不同的酶对DNA片段进行切割; MNase-seq直接检测核小体DNA序列, 绘制出核小体占据图, 从而间接反映出染色质的可及性区域, 另外3种方法则直接检测染色质开放区域。每一种分析方法各有其优点和局限性(表1)。虽然各技术都在不断地优化和改进, 但仍然对实验材料的质和量有一定的要求, 否则会导致一些临床样

本难以进行表观遗传学分析。

尽管不同实验室在此基础上都开发出了单细胞 scDNase-seq、scMNase-seq、scATAC-seq 分析技术，能以更少的核起始量获得相同的检测结果，解决了群体细胞异质性这一难题，但对分群的细胞进行不同的标记仍然是一大挑战。因此，未来的研究

需要对各种单细胞表观基因组分析技术染色质靶位点的标记进行改进，以减少不同单细胞之间的技术差异。这不仅有助于在整个细胞水平上理解细胞异质性，而且还有助于在单个位点上理解细胞异质性^[41]。

Table 1 Comparison of advantages and disadvantages of four chromatin accessibility analysis methods

表1 4种染色质可及性分析方法优缺点比较

研究方法	获取方式	优点	缺点	参考文献
DNase-seq	DNase I切割不受蛋白质保护的DNA序列	具有单碱基对的酶切位点分辨率	繁琐的酶滴定优化步骤；对DNA序列切割存在固有偏好	[8-9, 11]
MNase-seq	MNase优先消化无核小体的DNA	酶切特性使其具有较高的分辨率	对富含A、T的序列存在切割倾向；需要精确控制酶消化水平；无法分辨核小体与其他DNA-蛋白质复合物	[11, 16-17]
FAIRE-seq	超声波打断染色质	操作简单、无偏性分析；重复性好	依赖于超声对DNA的最佳剪切；信噪比很大程度上取决于正确的交联效率	[12, 25, 42]
ATAC-seq	Tn5转座酶同时切割开放区DNA并插入测序接头	细胞需求量少；快速而方便的操作	容易引入线粒体、叶绿体等细胞器DNA的污染	[2, 27-28]

参 考 文 献

- [1] Tolsma T O, Hansen J C. Post-translational modifications and chromatin dynamics. *Essays Biochem*, 2019, **63**(1):89-96
- [2] 吴杰, 全建平, 叶勇, 等. 染色质转座酶可及性测序研究进展. *遗传*, 2020, **42**(4):333-346
Wu J, Quan J P, Ye Y, *et al.* *Hereditas* (Beijing), 2020, **42**(4): 333-346
- [3] 宋乔, 王亚琦, 侯玉丽, 等. 衰老细胞染色质开放变化的初步分析. *生命科学研究*, 2020, **24**(2): 109-117
Song Q, Wang Y Q, Hou Y L, *et al.* *Life Science Research*, 2020, **24**(2):109-117
- [4] Wang Y, Liu Q, Tang F, *et al.* Epigenetic regulation and risk factors during the development of human gametes and early embryos. *Annu Rev Genomics Hum Genet*, 2019, **20**:21-40
- [5] Kwon S Y, Grisan V, Jang B, *et al.* Genome-wide mapping targets of the metazoan chromatin remodeling factor NURF reveals nucleosome remodeling at enhancers, core promoters and gene insulators. *PLoS Genet*, 2016, **12**(4):e1005969
- [6] Chen A, Chen D, Chen Y. Advances of DNase-seq for mapping active gene regulatory elements across the genome in animals. *Gene*, 2018, **667**:83-94
- [7] Cho J S, Blitz I L, Cho K W Y. DNase-seq to study chromatin accessibility in early xenopus tropicalis embryos. *Cold Spring Harb Protoc*, 2019, **2019**(4):pdb.prot098335
- [8] 丛瑞达. 基于DNase-Seq的转录因子DNA蛋白结合位点识别方法研究[D]. 哈尔滨: 哈尔滨工程大学控制科学与工程系, 2019
Cong R D. Abinding Sites Identification of Transcription Factors based on DNase high Throughput Sequencing [D]. Harbin: Department of Control Science and Engineering, Harbin Engineering University, 2019
- [9] Moyano T C, Gutierrez R A, Alvarez J M. Genomic footprinting analyses from DNase-seq data to construct gene regulatory networks. *Methods Mol Biol*, 2021, **2328**:25-46
- [10] Minnoye L, Marinov G K, Krausgruber T, *et al.* Chromatin accessibility profiling methods. *Nat Rev Methods Primers*, 2021, **1**(1):1-24
- [11] Tsompana M, Buck M J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*, 2014, **7**(1):33
- [12] Klein D C, Hainer S J. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res*, 2020, **28**(1):69-85
- [13] Song L, Crawford G E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, 2010, **2010**(2):pdb.prot5384
- [14] Klemm S L, Shipony Z, Greenleaf W J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*, 2019, **20**(4): 207-220
- [15] Hoeijmakers W A M, Bartfai R. Characterization of the nucleosome landscape by micrococcal nuclease-sequencing (MNase-seq). *Methods Mol Biol*, 2018, **1689**:83-101
- [16] Chereji R V, Bryson T D, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol*, 2019, **20**(1):198
- [17] 邓玮杭, 李鑫辉. MNase-seq与核小体定占位研究. *遗传*, 2020, **42**(12):1143-1146
Deng WH, Li XH. *Hereditas* (Beijing), 2020, **42**(12):1143-1146

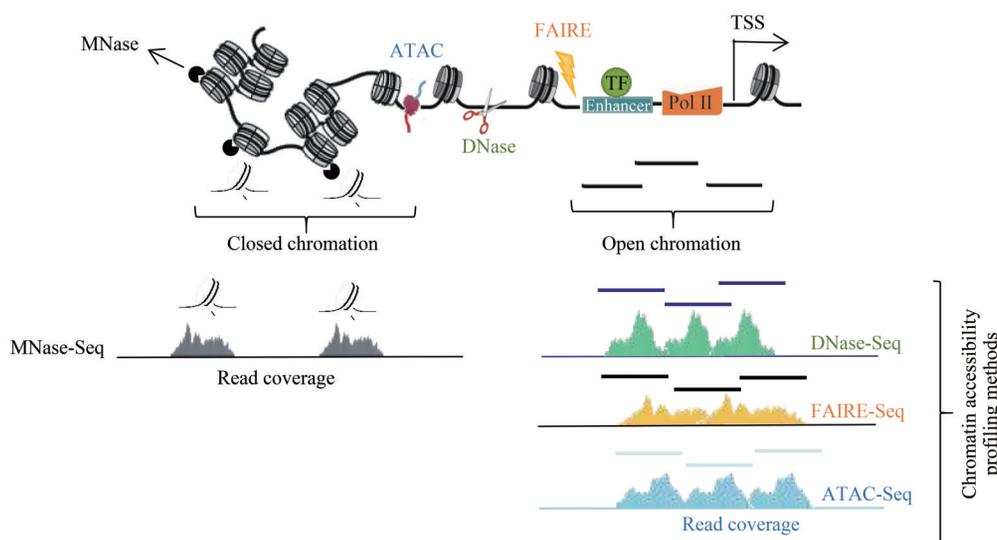
- [18] Pajoro A, Muino J M, Angenent G C, *et al.* Profiling nucleosome occupancy by MNase-seq: experimental protocol and computational analysis. *Methods Mol Biol*, 2018, **1675**:167-181
- [19] Ramani V, Qiu R, Shendure J. High sensitivity profiling of chromatin structure by MNase-SSP. *Cell Rep*, 2019, **26**(9):2465-2476
- [20] Omidbakhshfard M A, Winck F V, Arvidsson S, *et al.* A step-by-step protocol for formaldehyde-assisted isolation of regulatory elements from *Arabidopsis thaliana*. *J Integr Plant Biol*, 2014, **56**(6):527-538
- [21] Rodriguez-Gil A, Riedlinger T, Ritter O, *et al.* Formaldehyde-assisted isolation of regulatory elements to measure chromatin accessibility in mammalian cells. *J Vis Exp*, 2018(134):57272
- [22] Bianco S, Rodrigue S, Murphy B D, *et al.* Global mapping of open chromatin regulatory elements by formaldehyde-assisted isolation of regulatory elements followed by sequencing (FAIRE-seq). *Methods Mol Biol*, 2015, **1334**:261-272
- [23] Baum S, Reimer-Michalski E M, Jaskiewicz M R, *et al.* Formaldehyde-assisted isolation of regulatory DNA elements from *Arabidopsis* leaves. *Nat Protoc*, 2020, **15**(3):713-733
- [24] McKay D J. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to identify functional regulatory DNA in insect genomes. *Methods Mol Biol*, 2019, **1858**:89-97
- [25] Simon J M, Giresi P G, Davis I J, *et al.* A detailed protocol for formaldehyde-assisted isolation of regulatory elements (FAIRE). *Curr Protoc Mol Biol*, 2013, **102**(1):Unit21.26
- [26] 欧阳也, 秦玉婷, 姚超, 等. 利用 ATAC-seq 技术在人免疫细胞中检测染色质开放性的方法建立. *现代免疫学*, 2020, **40**(2):93-99
OuYang Y, Qin Y T, Yao C, *et al.* *Current Immunology*, 2020, **40**(2):93-99
- [27] Reske J J, Wilson M R, Chandler R L. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin*, 2020, **13**(1):22
- [28] Sun Y, Miao N, Sun T. Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas*, 2019, **156**(1):29
- [29] Xu W, Wen Y, Liang Y, *et al.* A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility. *Nat Protoc*, 2021, **16**(8):4084-4107
- [30] Marand A P, Schmitz R J. Single-cell analysis of cis-regulatory elements. *Curr Opin Plant Biol*, 2021, **65**:102094
- [31] Baek S, Lee I. Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. *Comput Struct Biotechnol J*, 2020, **18**:1429-1439
- [32] Taavitsainen S, Engedal N, Cao S, *et al.* Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse. *Nat Commun*, 2021, **12**(1):5307
- [33] Pierce S E, Granja J M, Greenleaf W J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat Commun*, 2021, **12**(1):2969
- [34] Pierce S E, Kim S H, Greenleaf W J. Finding needles in a haystack: dissecting tumor heterogeneity with single-cell transcriptomic and chromatin accessibility profiling. *Curr Opin Genet Dev*, 2021, **66**:36-40
- [35] Lee J H, Kim E W, Croteau D L, *et al.* Heterochromatin: an epigenetic point of view in aging. *Exp Mol Med*, 2020, **52**(9):1466-1474
- [36] 江小霞, 王常勇. 表观遗传学与组织工程. *生命科学*, 2020, **32**(3):299-307
Jiang X X, Wang C Y. *Chinese Bulletin of Life Sciences*, 2020, **32**(3):299-307
- [37] Nemeč S, Kilian K A. Materials control of the epigenetics underlying cell plasticity. *Nat Rev Mater*, 2020, **6**(1):69-83
- [38] Ranzoni A M, Tangherloni A, Berest I, *et al.* Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis. *Cell Stem Cell*, 2021, **28**(3):472-487
- [39] Muto Y, Wilson P C, Ledru N, *et al.* Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun*, 2021, **12**(1):2190
- [40] Liang D, Elwell A L, Aygun N, *et al.* Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat Neurosci*, 2021, **24**(7):941-953
- [41] Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet*, 2021, **22**(4):235-250
- [42] Segorbe D, Wilkinson D, Mizeranschi A, *et al.* An optimized FAIRE procedure for low cell numbers in yeast. *Yeast*, 2018, **35**(8):507-512

Research Progress of Chromatin Accessibility Analysis*

XU Lan, REN Li-Cheng**

(Department of Biology, Hainan Medical University, Haikou 571199, China)

Graphical abstract



Abstract Chromatin accessibility refers to the level of physical compaction of chromatin, which is determined by the chromatin binding factors that hinder DNA contact, nucleosome occupancy and topological structure. The chromatin accessibility pattern will be changed dynamically with external stimuli and developmental cues. Analyzing the TF (transcription factor) binding sites in the regulatory regions within the accessible chromatin can provide insight into the lineage factors and gene regulatory networks of specific cell types. Combined with high-throughput sequencing technology, several biochemical methods have been developed to describe the accessibility of chromatin, including bulk and single-cell level analysis. Depending on the techniques, the using enzymatic cleavage (DNase/MNase), transposition (Tn5) or physical methods (FAIRE) to isolate the accessible chromatin and subsequently using the high-throughput sequencing provide a genome-wide panorama of chromatin organization. This review introduced the common techniques (DNase-seq, MNase-seq, FAIRE-seq, and ATAC-seq) for determining chromatin accessibility and nucleosome positioning. The advantages and disadvantages of these 4 chromatin accessibility analysis techniques were summarized and compared. Their principles and main experimental procedures were introduced in detail; the development and application of related technologies were briefly discussed. The ATAC-seq based single-cell chromatin accessibility analysis and the view of potential useful were specially introduced. Although the chromatin accessibility profile is very valuable for studying gene regulation, it only provides a partial view of this complex process. We envision that technological improvements including single-molecule, multi-omics and spatial methods will bring further insight into the secrets of genome regulation.

Key words chromatin accessibility, DNase-seq, MNase-seq, FAIRE-seq, ATAC-seq

DOI: 10.16476/j.pibb.2021.0313

* This work was supported by grants from the Natural Science Foundation of Hainan Province (820RC638) and The National Natural Science Foundation of China (31660318, 31260269).

** Corresponding author.

Tel: 86-898-66893739, E-mail: lcheng08@126.com

Received: October 14, 2021 Accepted: December 17, 2021