

• 综述 •

三维基因组染色质构象捕获及其衍生技术

田昊*, 杨梓健*, 徐兴文*, 刘良玉

首都师范大学 生命科学学院 植物基因资源与低碳环境生物技术北京市重点实验室, 北京 100048

田昊, 杨梓健, 徐兴文, 等. 三维基因组染色质构象捕获及其衍生技术. 生物工程学报, 2020, 36(10): 2040–2050.

Tian H, Yang ZJ, Xu XW, et al. Three-dimensional chromosome conformation capture and its derived technologies. Chin J Biotech, 2020, 36(10): 2040–2050.

摘要: 线性染色质经过多重折叠凝缩到真核生物的细胞核中, 染色质的三维构象直接决定了真核生物的基因表达, 因此染色质可以在局部或远程空间上发生互作调控基因转录。折叠成环状构象的染色质可以借助染色质构象捕获 (Chromosome conformation capture, 3C) 技术来研究, 基于 3C 技术扩展的 4C/5C/Hi-C 从单个位点延伸到全基因组捕捉三维构象, 在此基础上, 染色质构象核心技术可以与免疫共沉淀、核酸分子杂交、单细胞、基因组测序等技术偶联而产生新的衍生技术和应用, 这极大地推动了染色质构象技术在基因时空特异性表达调控上的研究。文中将以 3C 和 Hi-C 等三维基因组核心技术为基础, 重点介绍染色质构象捕获及其衍生技术的原理和前沿应用。

关键词: 染色质构象捕获, 三维基因组, 高通量染色质构象捕获技术, 染色质免疫共沉淀, 配对末端标记测序技术分析染色质相互作用

Three-dimensional chromosome conformation capture and its derived technologies

Hao Tian*, Zijian Yang*, Xingwen Xu*, and Liangyu Liu

Key Laboratory of Plant Gene Resources and Biotechnology for Carbon Reduction and Environmental Improvement, College of Life Sciences, Capital Normal University, Beijing 100048, China

Abstract: Linear chromatin is compacted into eukaryotic nucleus through a complex and multi-layered architecture. Consequently, chromatin conformation in a local or long-distance manner is strongly correlated with gene expression. Chromosome conformation capture (3C) technology, together with its variants like 4C/5C/Hi-C, has been well developed to study chromatin looping and whole genome structure. In this review, we introduce new technologies including chromosome capture combined with immunoprecipitation, nuclei acid-based hybridization, single cell and genome sequencing, as well as their application.

Keywords: chromosome conformation capture, three-dimension genome, high-throughput chromatin conformation capture, chromatin immunoprecipitation, chromatin interaction analysis by paired-end tag sequencing

Received: March 10, 2020; **Accepted:** June 11, 2020

Supported by: National Natural Science Foundation of China (Nos. 31571258, 31800244).

Corresponding author: Liangyu Liu. Tel: +86-10-68901360; E-mail: liangyu.liu@cnu.edu.cn

*These authors contributed equally to this study.

国家自然科学基金 (Nos. 31571258, 31800244) 资助。

人类基因组的线性 DNA 分子全长可达 2 m, 它经过各种形式的压缩、折叠、盘绕凝缩到直径约 2 μm 大小的细胞核中。生物体的组织细胞结构与其功能相适应, 在分子层面上也是如此。在基因的转录调控中, 除了反式作用因子的参与, 染色质的三维构象 (Three dimension, 3D) 也是一个非常重要的调控层次, 即基因与远程调控元件之间形成的染色质空间结构会影响其基因表达和功能^[1]。

早在 2002 年, Job Dekker 等开发了基于 DNA 片段就近连接的 3C 技术 (Chromosome conformation capture, 3C), 并借助定量 PCR 的方式来检测染色质不同位点之间的互作频率。以 3C 技术为基础开发的环状染色质构象捕获技术 4C (Circular chromatin conformation capture)、染色质构象捕获碳拷贝技术 5C (Chromatin conformation capture

carbon copy)、高通量染色质构象捕获技术 Hi-C (High-throughput chromatin conformation capture) 等技术 (表 1), 为人们提供了从局部至全基因组范围内研究染色质远程相互作用的手段。

此外, 由于染色质的 3D 结构处在一个受多因素影响的动态变化的过程中, 并已证实这和多种疾病成因有关^[2], 因此基于 3C 和 Hi-C 等核心技术与免疫共沉淀技术 (Oligonucleotide capture OCT)、单细胞分离获取和测序技术 (Single-cell isolation and sequencing)、高通量测序 (High-throughput sequencing) 等技术结合 (表 1), 在动植物发育的各阶段全面解析染色质的空间相互作用如何影响基因的表达调控, 有助于人们在基因的线性调控认识之上, 进一步从 3D 的较多理解生命体内的基因调控网络。

表 1 染色质构象捕获及其衍生技术的特征

Table 1 The characteristic of Chromosome conformation capture and derived technology

Groups	Technologies	Features	Advantages and disadvantages
Core technologies of chromatin conformation capture	3C	One to One	Low cost, detect a few sites; need to predict interacting regions
	4C	One to Many	A few primers, higher throughput; need to characterize the interest target region
	5C	Many to Many	More interaction information; complicate to design primers
	3D-DSL Hi-C	All to All	Genome-wide detection, no bias, high throughput; high background
Combined with ChIP	ChIP-loop	ChIP+3C	Compare to 3C, lower background and higher specificity, capture specific protein mediated chromatin organization
	ChIA-PET	ChIP+Hi-C	Study genome-wide interactions mediated by specific proteins, with low background signal
	Hi-ChIP	ChIP+Hi-C+Tn5	Construct sequencing library by Tn5 with reduced input
Combined with OCT	Capture-C	3C with OCT	Capture conformation formed by target regions based on oligonucleotide probes
	Capture-Hi-C	Hi-C with OCT	Capture three-dimensional structure formed by target sequence in Hi-C library based on oligonucleotide probes, lower background signal
Combined with single cell technology	Single cell Hi-C	Application of Hi-C technology at the single cell level	Detect chromatin dynamic changes in single-cell level, background may be high
Combined with genome assembly	Hi-C associated genomic assembly	Using Hi-C data features to assist genome assembly	Can effectively improve the quality of genome assembly

1 染色质构象捕获技术的基本原理

1.1 3C 技术

3C 技术由 Job Dekker 于 2002 年开发并首先应用于酵母染色质互作研究,此后,3C 技术被迅速推广应用到动植物的染色质三维研究中。3C 主要用于检测特定的 DNA 位点与相邻位点之间的相互作用^[3]。该技术首先通过细胞交联 (Cross-link) 使得空间上相邻的染色质片段发生共价连接,然后用限制性核酸内切酶消化交联物,在极低的 DNA 浓度条件下借助 T4 DNA 连接酶促使空间上接近的 DNA 片段优先发生连接反应,这大大降低了随机发生的片段连接。最后,3C 技术借助定量 PCR 方法检测新连接片段的相对丰度,从而推测目标染色质片段是否存在物理相互作用^[4](图 1)。2004 年在酿酒酵母的研究中应用 3C 技术检测到 *SENI* 基因在 7 kb 区间内发生染色质相互作用^[5]。在较大的 DNA 区间内更容易检测 3C 相互作用,如 Skok 等在人胸腺细胞中确定了相距约 600 kb 的两个染色质片段之间存在远距离相互作用^[6]。由于 3C 需要较长的操作流程,因此对照的设计对于 3C 实验的成功是必需的。首先,考虑到不同 PCR 引物扩增效率引起的差异,需要建立互作嵌合连接产物的 DNA 对照模板,通常是 BAC 文库或克隆质粒的酶切连接产物。其次,当比较两个以上不同样品时,需要选取已知互作频率的位点用于归一化处理,综上,3C 技术需要严谨的操作、中间环节的质量控制和对照的设置^[4]。

1.2 4C 技术

在应用 3C 技术时,互作片段的检测受限在一定的区间内。为了在全基因组范围内无偏好地筛选与目的片段互作的候选片段,4C 技术被开发出来^[7]。4C 技术的关键是使交联的两个 DNA 分子连接成环,通过目的 DNA 片段的特异性引物进行反向 PCR,这样只需设计一对引物就可以研究一个特定位点与所有互作位点的互作频率^[8]

(图 1)。若 3C 概括为研究“一对一”的技术,那 4C 可以概括为研究“一对多”的技术。Zhao 等利用 4C 技术构建了人 *H19* 基因在全基因组范围内相互作用的调节网络^[9]。Simonis 等利用 4C 技术发现 *LMO3* 异位可能导致急性淋巴白血病细胞基因组发生染色质易位和倒位^[10]。为了进一步提高 4C 技术的准确性和灵敏度,Huang 等使用超声处理随机打断染色质的方法来代替特定位点的酶切反应,并结合新一代测序技术进行高通量分析,以减少假阳性的实验结果,增加检测精确度^[11]。

1.3 5C 技术

5C 技术的出现解决了 3C 技术研究多基因间互作时通量低的问题。在构建 3C 文库的基础上,5C 借助连接介导扩增的 LMA 方法 (Ligation-mediated amplification) 产生 5C 文库,并以此作为模板,在引物末端加上测序通用接头。若相互作用的片段能够形成重组的 DNA 连接产物,则连接的片段可以被扩增,最后利用微阵列芯片或测序分析多个位点之间的互作(图 1)。5C 技术可以概括为研究“多对多”的技术,Dostie 等利用 5C 发现了 K562 细胞中两个珠蛋白基因非编码调控区之间存在染色质环相互作用^[12]。2015 年 Sanyal 等通过 5C 检测人类基因组近端启动子与远端增强子之间的互作,发现远程相互作用更加普遍^[13]。然而,5C 技术也存在技术上的不足,比如设计成百上千的引物耗时费力,而且会遗漏部分染色质远程相互作用的信息。Kim 等对 5C 技术作了改进,他们在细胞核中进行限制性酶消化和连接反应实现了原位 3C。此外,与常规 5C 中仅在酶切位点设计单一方向引物不同,该方法设计了正反双向引物,从而有效地提高了检测的灵敏度^[14]。

与 5C 技术类似,3D-DSL 技术结合了 3C 和 DNA 选择性连接 (DNA selection and ligation, DSL)^[15]的方法。它首先利用生物素标记富集了 3C 产物,然后利用预先设计的含接头探针对靶位点片段进行特异扩增,最后通过芯片或高通量测

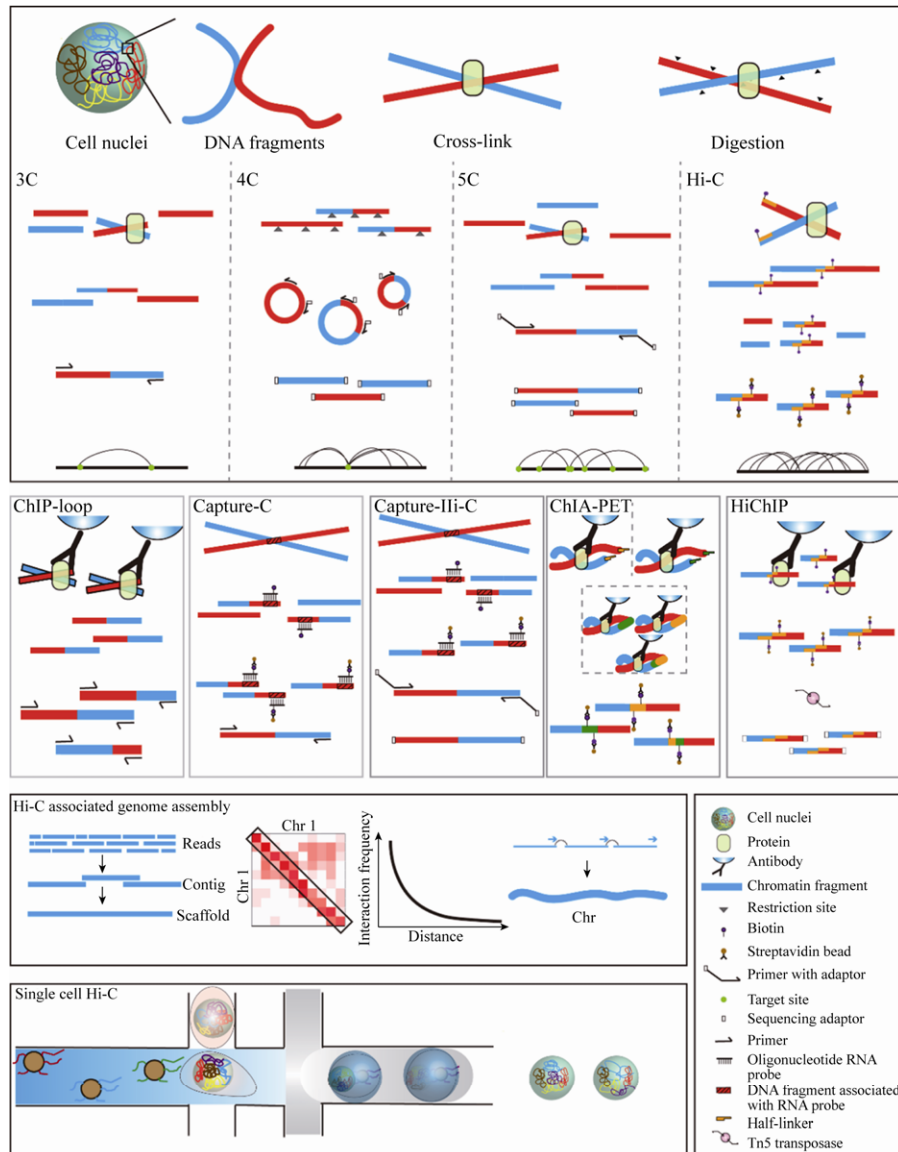


图 1 染色质构象捕获及其衍生技术

Fig. 1 Chromosome conformation capture and its derived technologies. 3C, 4C, 5C and Hi-C all need to cross-link cells, extract nuclei before experiments and digest chromatin by the restriction enzyme. 3C: During ligating the digested chromatin fragments in a big volume, the fragments that are physically close bias to be ligated first. Then the interaction frequency of two fragments is measured by qPCR. It's so called as "one to one". 4C: Chromatin fragments are cyclized based on two steps of digestion-ligation, and only a pair of primers is used to detect the interactions between the target site and other chromatin sites, namely "one-to-many". 5C: Amplifying DNA chimeric fragments with T3 and T7 adaptors followed by deep sequencing, namely "many to many". Hi-C: a biotin labeled nucleotide is incorporated to the ligation junction site and deep-sequencing is carried after purification by streptavidin beads, namely "all to all". ChIP-loop: Combining ChIP with 3C, the chromatin conformation mediated by specific protein is captured. ChIA-PET: Combining ChIP with Hi-C, and capturing the whole chromatin organization medicated by specific protein. HiChIP: Construct sequencing library by Tn5 transposase which could reduce starting material. Capture-C: Capture a few events of chromatin interaction among target DNA sites probed by biotin-labeled oligonucleotides. Capture Hi-C: Combination of Hi-C with Capture-C to measure whole genome interacting regions with target DNA fragments. Hi-C associated genomic assembly: Using the basic feature of Hi-C data that the interaction frequency decreases while increasing the distance to improve the quality of genome assembly. Hi-C single cell: Using single-cell sorting technology to barcode the isolated cell and combining with Hi-C high-throughput sequencing to capture chromosome conformation at single-cell level.

序检测染色质间的互作,大大提高信噪比^[16]。借助 3D-DSL 技术,Harismendy 等发现冠状动脉疾病 (Coronary artery disease, CAD) 相关增强子区域与染色体 9p21 等区域存在远程相互作用^[17]。

1.4 Hi-C 技术

Hi-C 技术通过结合生物素富集和高通量测序方法,研究全基因组范围内染色质内或染色质间空间位置上的互作关系,从而获得高分辨率的染色质相互作用图谱。Hi-C 技术是基于 3C 技术建立的,该技术在酶切后通过末端补平加入生物素标记,再扩大反应体系进行邻位连接,然后利用链霉亲和素偶联的磁珠富集带有生物素标记的连接片段,建库并进行高通量测序,获得全基因组范围内的互作信息 (图 1)。由于可以提供全基因组范围内的所有染色质位点之间的高精度互作信息,因此 Hi-C 技术被广泛应用于挖掘基因调控元件、揭示细胞时空特异性染色质构象变化以及绘制基因组三维图谱等研究工作中。2009 年 Job Dekker 研究组利用 Hi-C 技术绘制出 1 Mb 分辨率人类三维基因组图谱^[18]。Rudan 等运用 Hi-C 技术发现 CTCF (CCCTC-binding factor) 蛋白与粘连蛋白 (Cohesin) 在细胞核中共定位,保守的 CTCF 在不同物种中均在染色质的拓扑相关结构域 TAD (Topologically associated domains) 边界进行富集,但是 CTCF 与不同 DNA 序列结合的差异可能导致内部结构域发生变化,进而造成功能不同^[19]。Chandra 等使用 Hi-C 绘制了细胞衰老过程中基因组的 3D 结构变化图谱,发现在衰老细胞中 TAD 内部相互作用减少,而 TAD 之间的相互作用增多^[20]。Taberlay 等利用 Hi-C 发现癌症细胞中 TAD 形成了额外的域边界 (Domain boundaries),因此导致 TAD 数目增多平均大小变小,癌细胞特有的域边界中大多出现在拷贝数发生变异的区域,此外还发现癌细胞染色质构象重构受远程表观遗传修饰的调节^[21]。

植物的多倍体化增加了基因转录调控的复杂

性,Hi-C 技术的应用有助于对多倍体化和三维基因组结构动力学的理解。Wang 等利用 Hi-C 技术绘制了二倍体和四倍体棉花的 3D 基因组结构,并发现了特异的 A/B 区室 (A/B compartment) 和 TAD,这项研究增进了对多倍体植物染色质结构的了解,并为 3D 基因组进化与转录调控之间的关系提供了新的思路^[22]。然而,在植物中是否广泛存在 TAD 目前还存在争论^[23]。

为了提高 Hi-C 的精确性,人们开发了一系列 Hi-C 优化技术,如原位 Hi-C (*in situ* Hi-C) 和基于酶切酶连的 Hi-C (Digestion-Ligation-Only Hi-C, DLO Hi-C) 等技术。与 2009 年开发的在大体系下进行的 Hi-C 技术 (也被称为第一代 Dilution Hi-C) 相比^[24-18],2014 年开发的在小体系下进行的 *in situ* Hi-C 通过在完整的核内原位完成交联、酶切、连接步骤,降低了假阳性结果,在提高信噪比的同时缩短了实验周期。DLO Hi-C 采用了边切边连的策略,先用 *Hind* III 切割交联的染色质,然后在切口末端加上接头序列形成 *Mme* I 的识别位点,这样可以保证 *Hind* III 尽可能酶切彻底,同时有效地防止了 *Hind* III 切点的自连,两个接头序列发生连接形成的嵌合片段被 *Mme* I 切割成 80 bp 大小,此长度的片段通过凝胶电泳切胶富集,可以有效地降低 Hi-C 的背景信号。与 Hi-C 技术相比,DLO Hi-C 成本低、周期短、获得有效数据比例高。DLO Hi-C 与 Dilution Hi-C 相比,在较低测序深度的情况下,可以找到更多的染色质互作位点^[25]。Rao 等在 2014 年通过 *in situ* Hi-C 在 GM12878 细胞中绘制了分辨率达 1 kb 的基因组三维图谱^[26]。

2 染色质构象捕获偶联免疫共沉淀的衍生技术

2.1 染色质免疫共沉淀环技术

染色质免疫共沉淀环 ChIP-loop (Chromatin immunoprecipitation-loop) 技术也被称为 ChIP-3C。

作为研究蛋白质介导的染色质互作的一种方法, ChIP-loop 结合了染色质免疫共沉淀 (Chromatin immunoprecipitation, ChIP) 和 3C 技术。ChIP 是鉴定反式作用因子结合基因组中 DNA 位点的常用方法, 但是它无法确定这些 DNA 位点之间是否存在物理性的相互作用, 而 3C 具有分析局部或远程染色质相互作用的优势, 因此 ChIP-loop 融合两种技术的优点, 识别受特定蛋白质介导的染色质间相互作用, 降低了非特异性背景信号^[27]。交联的染色质经过酶切后, 利用靶蛋白的特异性抗体进行免疫沉淀, 然后再进行产物的连接反应^[28] (图 1)。Horike 等采用 ChIP-loop 发现了由 MECP2 介导形成于 *DLX5* 和 *DLX6* 基因间的染色质环可以抑制相关基因的转录过程^[29]。Kumar 等利用超声技术代替限制性消化染色质的方法揭示了 SATB1 将核基质结合区 (Matrix association regions, MAR) 束缚到核基质上从而促进 *MHC-I* 基因组成独特的高级染色质环结构, 进而调控基因表达^[30]。为了更进一步证实特定蛋白直接参与了染色质环的形成, ChIP-loop 往往需要在敲除该蛋白表达的突变体中检测染色质环的形成是否正常^[31]。

2.2 配对末端标记测序技术分析染色质相互作用

配对末端标记测序技术分析染色质相互作用 (Chromatin interaction analysis by paired-end tag sequencing, ChIA-PET) 主要结合 ChIP 和 Hi-C, 在全基因组范围内捕获靶蛋白特异介导的 DNA 互作片段。ChIA-PET 技术流程类似于 ChIP-loop, 用靶蛋白的特异性抗体免疫共沉淀 DNA-蛋白质复合物后, 进行建库测序^[32] (图 1)。ChIA-PET 可以无偏好性地检测全基因组范围内染色质之间的相互作用, 它是 ChIP-loop 技术全基因组范围的升级扩展版本。Tang 等利用 ChIA-PET 绘制了人类不同细胞系中由 CTCF 和 RNA 聚合酶 II 介导的全基因组水平的染色质相互作用图谱^[33]。Fullwood 等利用 ChIA-PET 揭示了雌激素受体 α

通过远程相互作用调控下游基因启动子, 并可形成染色质环整合相关基因调控共转录^[34]。Peng 等利用 ChIA-PET 技术捕获了 H3K4me1 等 4 种组蛋白修饰及 RNA 聚合酶 II 介导的染色质构象, 在玉米中构建了启动子和远端调控元件之间的高分辨率染色质相互作用图谱。研究结果表明调节元件之间形成了染色质环, 近端启动子区发生染色质互作的基因更倾向于共表达^[35]。Zhao 等利用 ChIA-PET 技术绘制了水稻中由 H3K4me3 组蛋白修饰和 RNA 聚合酶 II 介导的启动子-启动子相互作用, 以及由 H3K9me2 组蛋白修饰介导的异染色质相互作用, 从而划分出相应的活跃和非活跃转录潜能的独立染色质互作模块^[36]。ChIA-PET 通常是在一个文库中研究单个蛋白介导的染色质相互作用, 而 Fullwood 等开发了 Multiplex ChIA-PET, 在 ChIA-PET 建库过程中引入 6 个半连接子 (Half-linker), 可以同时分析多个转录因子并获得更多染色质相互作用信息, 与 ChIA-PET 相比具有起始细胞量低、耗时少、费用低的优势^[37]。

2.3 利用染色质免疫共沉淀的原位 Hi-C 技术

利用染色质免疫共沉淀的原位 Hi-C (*in situ* Hi-C followed by chromatin immunoprecipitation, HiChIP) 是一种结合原位 Hi-C、免疫共沉淀和转座酶建库来解析染色质构象的方法。HiChIP 在免疫共沉淀后得到较低起始量 DNA 情况下, 利用转座酶 Tn5 完成微量 DNA 建库。与 ChIA-PET 相比, HiChIP 细胞起始量降低到 1%, 互作序列读取量高达 10 倍以上, 与原位 Hi-C 相比其信噪比更高^[38] (图 1)。HiChIP 可以用于针对特定蛋白介导的染色质互作构象分析, 转录因子作用机制研究, 表观修饰对基因调控的机制研究。2017 年科学家们利用 HiChIP 对人细胞 H3K27ac、H3K27me3 和 RNA 聚合酶 II 介导的染色质相互作用进行了检测, 揭示了细胞核中新的区室化特异互作^[39]。

3 染色质构象捕获偶联核酸分子杂交的衍生技术

3.1 Capture-C

Capture-C 结合了 OCT 和 3C 技术,它通过将 3C 文库与携带生物素的 RNA 捕获探针杂交并进行富集,以构建 Capture-C 文库(图 1)。Capture-C 可在单个实验中提供数百个基因位点的无偏好、高分辨率的顺式互作图谱,除可以阐明互作的顺式作用元件调控基因时空表达的机制,它还可以进一步挖掘引发复杂疾病的 DNA 变异。Hughes 等利用 Capture-C 技术获得了数百个基因启动子与调控元件之间高精度互作图谱^[40]。Capture-C 技术的主要局限是需要大量的细胞进行实验。但许多组织和稀有细胞群无法提供实验所需细胞的量,因此 Oudelaar 等开发了一种低起始量 Capture-C 方法(Low-input Capture-C),这种方法通过超声处理、酶促反应后末端修复并添加连接子,可以利用约 20 000 个细胞以较大分辨率生成约 10 000 个高质量相互作用图谱^[41]。

3.2 Capture Hi-C

将 OCT 与 Hi-C 相结合衍生出 Capture Hi-C 技术,与 Capture-C 相比,Capture Hi-C 可以无偏好、高精度地实现全基因组中基因与调控元件之间的互作。Capture Hi-C 通过在 Hi-C 建库后,用带有生物素标记的 RNA 探针与 Hi-C 文库杂交,再进行富集测序(图 1)。2015 年利用 Capture Hi-C 揭示了直肠癌风险基因 *CRC* 与长链非编码 RNA (Long non-coding RNA, lncRNA)、已知癌基因和非编码区间存在相互作用^[42]。Martin 等利用 Capture Hi-C 发现 4 种自身免疫疾病变异位点会与功能靶点互作,并且这些疾病相关的单核苷酸多态性位点(Single nucleotide polymorphism, SNP)倾向与远距离的靶基因互作,而不是与其最近的基因^[43]。

4 染色质构象捕获偶联单细胞测序的衍生技术

3C、Hi-C 及其衍生技术往往需要数以百万计的细胞材料,但是如胚胎细胞及干细胞等珍贵稀有样品,通常难以获得足量的细胞以完成三维基因组检测。同时荧光原位杂交(Fluorescence in situ hybridization, FISH)结果表明即便表型甚至基因组一致的细胞之间也存在着不同的三维染色质构象^[1]。因此,检测单个细胞水平上的三维基因组结构有着重要的意义。而单细胞 Hi-C 的难点在于多个方面,如难以从组织分离获取单细胞;将单细胞技术与 3C 或 Hi-C 技术偶联;提高平行处理多个单细胞的通量;获取高质量数据等。应用流式细胞分选(Fluorescence-activated cell sorter, FACS)技术,激光捕获显微切割(Laser capture microdissection, LCM)等方法可以获取单细胞,利用微孔矩阵和微流控等方法可一次性区分 10^3 – 10^6 个细胞^[44](图 1)。

2013 年 Nagano 等首次用毛细玻璃管在显微镜下通过吸取获得了单个细胞核实现了单细胞 Hi-C,但难以提高分离获取单细胞的通量^[45]。2018 年 Tan 等开发的二倍体染色质构象捕获(Diploid chromatin conformation capture, Dip-C)技术,利用多末端标记扩增(Multiplex end-tagging amplification, META)方法进行全基因组扩增,具有灵敏度更高、假阳性低的优点,最终构建了分辨率 20 kb 的人类二倍体单细胞三维图谱^[46]。2019 年 Zheng 等利用微流控技术开发了 ChIA-Drop 技术(Multiplex chromatin-interaction analysis via DROplet-based and barcode-linked sequencing),利用微液滴及条形码标记的测序技术以单分子精度检测多重染色质相互作用,研究发现果蝇 *Drosophila melanogaster* S2 细胞利用 RNA 聚合酶 II 捕获的启动子区域具有方向偏好性,该证据

可能支持单向染色质外环外挤 (One-sided extrusion) 模型^[47], 并且 S2 细胞中介导 TADs 形成的复合体具有高度异质性的特征, 该结果与在单个细胞中使用超分辨率成像方法观测的结果类似^[48]。未来, 利用微液滴及条形码标记的 ChIA-Drop 等技术将弥补单细胞 Hi-C 在进行大规模细胞建库时的不足, 辅助单细胞 Hi-C 分析。通过结合单细胞转录组测序以及单细胞转座酶介导的染色质可接近性 (Assay for transposase-accessible chromatin, ATAC) 等方法, Hi-C 可以为我们展示单细胞水平上染色质空间结构与基因转录之间更加直接的调控关系^[49]。

5 染色质构象捕获辅助基因组组装的衍生应用

完整的基因组信息是开展基因功能研究的重要基础。基因组组装是指将测序获得的序列片段进行拼接获得全基因组序列信息的过程。基因组组装分为 3 个步骤: 首先测序得到短读序列 (Reads), 根据它们的重叠区域拼装成较长的连续序列 (Contig); 然后再将连续序列按照顺序拼接成更长的支架结构 (Scaffolds); 最后将支架合并组装到染色体上, 在拼装过程中支架之间可能存在诸多空白 (Gap) 区域, 即没有信息的序列^[50]。基因组中 gaps 的多少成为制约基因组组装质量的关键。组装过程中, 重复序列、同源多倍体等都可导致空白序列的产生。根据 Hi-C 技术中相互作用强度随距离衰减的原理, 可以帮助 scaffold 成功拼接到染色体上, 有效地辅助基因组的组装^[51] (图 1)。

Hi-C 技术在目前基因组组装中应用非常广泛, 借助 Hi-C 技术, 2017 年将测序产生的埃及伊蚊 *Aedes aegypti* 大片段 scaffold 进行组装^[52], 将大麦 *Hordeum vulgare* 基因组 95% 的超级 scaffold (Super-scaffold) 成功锚定在染色体上^[53]。2018 年利用 Hi-C 技术完成了亚洲棉 *Gossypium arboreum* 基因组的重新组装, 将基因组 Contig

N50 提升到 1.1 Mb^[54]。2019 年结合 Hi-C 和纳米孔测序技术高质量完成了红斑石斑鱼 *Epinephelus akaara* 97% 基因组的组装^[55]。另外, Hi-C 技术也已成功应用于具有同源多倍体物种的基因组组装, 如 2018 年基于 Hi-C 数据成功组装甘蔗野生种“割手密” *Saccharum spontaneum* 基因组^[56]。在宏基因组研究中, Burton 等利用 Hi-C 技术成功地将真菌、细菌和古菌物种的基因组进行了区分和组装, 能够高效高质地从微生物群落中鉴定和拼装出多个单物种的基因组^[57]。

6 展望

近年来 3C 及其衍生技术迅速发展并被广泛应用, 人们对染色体异常相关疾病的了解也越来越深入。研究人员已经解析了乳腺癌和前列腺癌等细胞的染色质三维结构^[21-57]。Barutcu 等利用 Hi-C 发现了乳腺癌细胞 (MCF-7) 基因组中常染色质之间的相互作用频率降低, 特别是端粒和亚端粒区域^[58]。除了癌症外, 非肿瘤相关疾病也可以通过改变染色质三维结构来影响相关基因表达。Lupianez 等发现敲除、倒置或重复改变跨越 TAD 边界的 *WNT6* 等基因会引起人肢体畸形^[52]。此外, 基于 3C 及其各种衍生技术, 研究人员能够揭示染色质中如 A/B 区室、TAD、染色质环等不同的三维拓扑结构^[59-60]。此外, 随着单细胞分离技术的日渐成熟, 结合 Hi-C 等技术从单细胞的染色质构象动态变化上解析生命活动无疑是生命科学的一个热点领域。细胞内染色质构象的功能一直是基因组学研究的重点, 3C 及其衍生技术与应用有助于人类了解并攻克染色体异常疾病、细胞周期、细胞分化、动植物基因组进化等核心问题。

尽管染色质构象捕获及其衍生技术目前仍存在一些难题尚待解决, 比如难以检测染色质间的弱相互作用、实验周期长导致无法快速真实反应互动状况、信噪比及分辨率低等, 但 3C 及其衍生技术无疑为结构和功能基因组学的研究提供了

强大工具支撑。三维基因组学及其衍生技术在生命科学中发挥着越来越重要的作用。随着交叉学科的不断加强,也会产生更多探究染色体生物学的新技术新方法。同时,基于染色质构象捕获衍生开发的新技术,也必将为遗传学、细胞生物学、生物物理、生物信息等学科发展注入新的活力。

REFERENCES

- [1] Parmar JJ, Woringer M, Zimmer C. How the genome folds: the biophysics of four-dimensional chromatin organization. *Annu Rev Biophys*, 2019, 48: 231–253.
- [2] Li RF, Liu YT, Hou YP, et al. 3D genome and its disorganization in diseases. *Cell Biol Toxicol*, 2018, 34(5): 351–365.
- [3] Dekker J, Rippe K, Dekker M, et al. Capturing chromosome conformation. *Science*, 2002, 295(5558): 1306–1311.
- [4] Dekker J. The three ‘C’ s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 2006, 3(1): 17–21.
- [5] O’Sullivan JM, Tan-Wong SM, Morillon A, et al. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, 2004, 36(9): 1014–1018.
- [6] Skok JA, Gisler R, Novatchkova M, et al. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat Immunol*, 2007, 8(4): 378–387.
- [7] Simonis M, Klous P, Splinter E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 2006, 38(11): 1348–1354.
- [8] Göndör A, Rougier C, Ohlsson R. High-resolution circular chromosome conformation capture assay. *Nat Protoc*, 2008, 3(2): 303–313.
- [9] Zhao ZH, Tavoosidana G, Sjölander M, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 2006, 38(11): 1341–1347.
- [10] Simonis M, Klous P, Homminga I, et al. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods*, 2009, 6(11): 837–842.
- [11] Fullwood MJ, Ruan YJ, Huang PYH, et al. Protocol: Sonication-based circular chromosome conformation capture with next-generation sequencing analysis for the detection of chromatin interactions. *Sci Technol Food Ind*, 2010, 3(9): 656–662.
- [12] 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.
- [13] Sanyal A, Lajoie BR, Jain G, et al. The long-range interaction landscape of gene promoters. *Nature*, 2012, 489(7414): 109–113.
- [14] Kim JH, Titus KR, Gong WF, et al. 5C-ID: Increased resolution Chromosome-Conformation-Capture-Carbon-Copy with *in situ* 3C and double alternating primer design. *Methods*, 2018, 142: 39–46.
- [15] Kwon YS, Garcia-Bassets I, Hutt KR, et al. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor α -binding program on human gene promoters. *Proceedings of the National Academy of Sciences of the United States of America*. 2007, 104.
- [16] Cao J, Luo Z, Cheng Q, et al. Three-dimensional regulation of transcription. *Protein Cell*, 2015, 6(4): 241–253.
- [17] Harismendy O, Notani D, Song X, et al. 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature*, 2011, 470(7333): 264–268.
- [18] Lieberman-Aiden E, Van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, 326(5950): 289–293.
- [19] Rudan MV, Barrington C, Henderson S, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*, 2015, 10(8): 1297–1309.
- [20] Chandra T, Ewels PA, Schoenfelder S, et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep*, 2015, 10(4): 471–483.
- [21] Taberlay PC, Achinger-Kawecka J, Lun AT, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res*, 2016, 26(6):

- 719–731.
- [22] Wang MJ, Wang PC, Lin M, et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants*, 2018, 4(2): 90–97.
- [23] Doğan ES, Liu C. Three-dimensional chromatin packing and positioning of plant genomes. *Nat Plants*, 2018, 4(8): 521–529.
- [24] Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 2014, 159(7): 1665–1680.
- [25] Lin D, Hong P, Zhang SH, et al. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat Genet*, 2018, 50(5): 754–763.
- [26] Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 2014, 159(7): 1665–1680.
- [27] Fullwood MJ, Ruan YJ. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem*, 2009, 107(1): 30–39.
- [28] Gavrillov A, Eivazova E, Pirozhkova I, et al. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification//Collas P, Ed. *Chromatin Immunoprecipitation Assays*. Totowa, NJ: Humana Press, 2009, 567: 171–188.
- [29] Horike S, Cai ST, Miyano M, et al. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 2005, 37(1): 31–40.
- [30] Kumar PP, Bischof O, Purbey PK, et al. Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*, 2007, 9(1): 45–56.
- [31] Splinter E, Heath H, Kooren J, et al. CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev*, 2006, 20(17): 2349–2354.
- [32] Fullwood MJ, Wei CL, Liu ET, et al. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res*, 2009, 19(4): 521–532.
- [33] Tang ZH, Luo OJ, Li XW, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, 2015, 163(7): 1611–1627.
- [34] Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature*, 2009, 462(7269): 58–64.
- [35] Peng Y, Xiong D, Zhao L, et al. Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun*, 2019, 10: 2632.
- [36] Zhao L, Wang SQ, Cao ZL, et al. Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat Commun*, 2019, 10: 3640.
- [37] Choy J, Fullwood MJ. Deciphering noncoding RNA and chromatin interactions: multiplex chromatin interaction analysis by paired-end tag sequencing (mChIA-PET). *Enhancer RNAs*. 2017, 1468: 63–89.
- [38] Mumbach MR, Rubin AJ, Flynn RA, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*, 2016, 13(11): 919–922.
- [39] Rowley MJ, Nichols MH, Lyu XW, et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol Cell*, 2017, 67(5): 837–852.e7.
- [40] Hughes JR, Roberts N, McGowan S, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*, 2014, 46(2): 205–212.
- [41] Oudelaar AM, Davies JOJ, Downes DJ, et al. Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Res*, 2017, 45(22): e184.
- [42] Jäger R, Migliorini G, Henrion M, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun*, 2015, 6: 6178.
- [43] Martin P, McGovern A, Orozco G, et al. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun*, 2015, 6: 10069.
- [44] Zhang K. Stratifying tissue heterogeneity with scalable single-cell assays. *Nat Methods*, 2017, 14(3): 238–239.
- [45] Nagano T, Lubling Y, Stevens TJ, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 2013, 502(7469): 59–64.

- [46] Tan LZ, Xing D, Chang CH, et al. Three-dimensional genome structures of single diploid human cells. *Science*, 2018, 361(6405): 924–928.
- [47] Zheng M, Tian SZ, Capurso D, et al. Multiplex chromatin interactions with single-molecule precision. *Nature*, 2019, 566(7745): 558–562.
- [48] Bintu B, Mateo LJ, Su JH, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 2018, 62(6413).
- [49] Deal RB, Henikoff S. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat Protoc*, 2011, 6(1): 56–68.
- [50] Gnerre S, Maccallum I, Przybylski D, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA*, 2011, 108(4): 1513–1518.
- [51] Kong SY, Zhang YB. Deciphering Hi-C: from 3D genome to function. *Cell Biol Toxicol*, 2019, 35(1): 15–32.
- [52] Mascher M, Gundlach H, Himmelbach A, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 2017, 544(7651): 427–433.
- [53] Du XM, Huang G, He SP, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet*, 2018, 50(6): 796–802.
- [54] Ge H, Lin KB, Shen MQ, et al. *de novo* assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Mol Ecol Resour*, 2019, 19(6): 1461–1469.
- [55] Zhang JS, Zhang XT, Tang HB, et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat Genet*, 2018, 50(11): 1565–1573.
- [56] Burton JN, Liachko I, Dunham MJ, et al. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)*, 2014, 4(7): 1339–1346.
- [57] Barutcu AR, Lajoie BR, McCord RP, et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol*, 2015, 16: 214.
- [58] Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 2015, 161(5): 1012–1025.
- [59] Naumova N, Dekker J. Integrating one-dimensional and three-dimensional maps of genomes. *J Cell Sci*, 2010, 123(12): 1979–1988.
- [60] Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 2012, 485(7398): 376–380.

(本文责编 郝丽芳)