

评述与展望

Review and Progress

高通量染色体构象捕获技术的原理及应用

房瑞新^{1,2} 刘欣² 董依萌¹ 张然然¹ 王洪亮¹ 王天骄¹ 刘华淼¹ 刘汇涛¹ 王磊¹ 李洋^{1*} 邢秀梅¹

1 中国农科院特产研究所, 长春, 130112; 2 东北林业大学, 哈尔滨, 150040

* 通信作者, liyang01@caas.cn

摘要 染色体三维结构的研究已成为当前生物学研究的热点, 而高通量染色体构象捕获技术(high-throughput chromosome conformation capture, Hi-C)以其准确率高、应用范围广、操作简单成为研究基因表达、调控和染色体三维空间结构的关键技术, 在此基础上催发了染色体领域(chromosome territory, CT)、染色体区室(chromatin compartment A/B)、拓扑结构域(topological associated domain, TAD)等结构的发现。Hi-C技术在辅助基因组组装、探究基因功能、判断致病基因等方面都有广泛的应用。本文综述了Hi-C技术的发展过程及其在三维基因组研究中的重要作用, 对照了该技术与其他测序技术的异同, 详细介绍其应用前景、操作原理以及在推动染色质三维结构探索方面发挥的作用, 结合当前先进的创新技术与科研经验, 提供了关于Hi-C技术的全面简介。

关键词 Hi-C技术; 三维基因组学; 染色质结构

Principle and Application of High-throughput Chromosome Conformation Capture Technology

Fang Ruixin^{1,2} Liu Xin² Dong Yimeng¹ Zhang Ranran¹ Wang Hongliang¹ Wang Tianjiao¹ Liu Huamiao¹ Liu Huitao¹ Wang Lei¹ Li Yang^{1*} Xing Xiumei¹

1 Institute of Special Animal and Plant Sciences of CAAS, Changchun, 130112; 2 Northeast Forestry University, Harbin, 150040

* Corresponding author, liyang01@caas.cn

DOI: 10.13417/j.gab.040.003458

Abstract The research of the three-dimensional structure of the chromosome has become a hot spot in biology research. The high-throughput chromosome conformation capture (Hi-C) has become the key technology in the study of gene expression, regulation, and three-dimensional structure of chromosome with its high accuracy, wide application, and simple operation. On this basis, it has prompted the discovery of chromosome territory (CT), chromosome compartment (Compartment A/B), and topological structure domain (TAD). It has been widely used in assisting genome assembly, exploring gene function, and judging pathogenic genes. This paper introduces the development process of Hi-C technology and its important role in the study of the three-dimensional genome compared with other sequencing technology. The application prospect, principle, operation, and development of Hi-C technology are introduced in detail. Combined with the advanced innovative technology and scientific research experience, this paper aims to provide a comprehensive and concise introduction of Hi-C technology.

Keywords Hi-C technology; Three-dimensional genome; Chromatin structure

基金项目: 本研究由中国农业科学院基本科研业务费专项(1610342020027; 1610342020025; 1610342020021)共同资助

引用格式: Fang R.X., Liu X., Dong Y.M., Zhang R.R., Wang H.L., Wang T.J., Liu H.M., Liu H.T., Wang L., Li Y., and Xing X.M., 2021, Principle and application of high-throughput chromosome conformation capture technology, *Jiyinzhexue yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 40(11-12 combined issue): 3458-3467. (房瑞新, 刘欣, 董依萌, 张然然, 王洪亮, 王天骄, 刘华淼, 刘汇涛, 王磊, 李洋, 邢秀梅, 2021, 高通量染色体构象捕获技术的原理及应用, *基因组学与应用生物学*, 40(第11-12期合刊): 3458-3467.)

染色质是生物遗传信息的载体,是解决生物遗传进化问题的关键。2001年,人类基因组计划(Collins et al., 2003)的完成意味着生命科学的研究进入新的时代,越来越多的遗传信息被破译。科学家们努力探索遗传物质调控生命活动的机制,由于技术上的限制,只能对染色体形态进行初步观察,但从一维线性基因组的角度不足以精确锁定染色体发生变异的位置或揭示基因组的调控模式(Feuk et al., 2006),因而还有诸多问题亟待解决。如果将人类体细胞中23对染色体上的全部DNA首尾相连,其长度可达2 m,但是人类体细胞的细胞核直径只有6 μm,且细胞核内还含有蛋白质等其它物质。在如此小的空间内,DNA必然要经过高度浓缩和折叠(Rao et al., 2014)。为明确染色质在三维空间的结构及其在调控基因表达中的作用,2002年,Dekker等首次将染色体构象捕获技术(chromosome conformation capture, 3C)应用到酵母细胞的研究中,发现不同功能结构域的存在。3C可以检测任意两个基因座之间相互

表1 3C技术与其它技术的对比
Table 1 Comparison between 3C technology and other technologies

技术名称 Technical classification	全称 Full name	原理 Theory	通量 Throughput	特点 Features
3C	染色体构象捕获技术 Chromosome conformation capture	甲醛交联, DNA 邻近连接, DNA 纯化, 正向的 PCR 测序 Formaldehyde crosslinking, Proximity dependent DNA ligation, DNA purification, Forward PCR	一对一 One to one	邻近连接 Proximity ligation
4C	环状染色质构象捕获技术 Circular chromosome conformation capture	甲醛交联, DNA 邻近连接, 反向 PCR, 芯片分析或通过添加适配器排序 Formaldehyde crosslinking, Proximity dependent DNA ligation, Reverse PCR, Chip analysis or sequenced by adding adapter	一对多 One to many	反向 PCR Reverse PCR
5C	3C 碳拷贝技术 Chromosome conformation capture carbon copy	甲醛交联, DNA 邻近连接, DNA 纯化, 多重 PCR Formaldehyde crosslinking, Proximity dependent DNA ligation, DNA purification, Multiplex PCR	多对多 Many to many	多个序列相互作用 Interact with multiple sequences
Hi-C	高通量染色体构象捕获技术 High-throughput chromosome conformation capture	甲醛交联, 细胞溶解, 限制酶消化, 生物素标记, 邻近连接, 反向交联, 超声处理, 添加适配器, PCR 和测序 Formaldehyde crosslinking, Cell lysis, Cut with restriction enzyme, Biotin labeling, Proximity ligation, Reverse crosslinking, Sonication, Add adapter, PCR and sequencing	所有对 All to all	无差别的全基因组测序 Undifferentiated whole-genome sequencing
ChIA-PET	基于双末端测序的染色质交互分析 Chromatin interaction analysis by paired-end tag	甲醛交联, 细胞溶解, 超声处理, ChIP, 特定连接子的邻近连接, MMEI 或 ECOP15I 消化, 添加适配器, PCR 和测序 Formaldehyde crosslinking, Cell lysis, Sonication, ChIP, Proximity dependent specific linker ligation, MMEI or ECOP15I digestion, Add adopter, PCR and sequencing	多对多 All to all	噪声评估 Noise assessment

作用频率,但是具有定量而非定性测定的局限性(Dekker et al., 2013),成为Hi-C技术的基础。

Lieberman-Aiden等(2009)以整个细胞核为研究对象,通过大规模平行测序法并结合生物信息学,研究DNA在空间中任意两点之间的互作关系。将染色体的构象总结为类似“分型球”的模型,成为从技术上突破对三维基因组学认知障碍的开端,确定了Hi-C技术具有绘制基因组三维立体构象的能力,使三维基因组学研究达到认识染色体结构组成和基因表达调控的高度。

1 Hi-C 概述

Hi-C技术属于3C的高通量版本但并不是唯一的衍生版本,除此之外还有4C(circularized chromosome conformation capture)技术(Marieke et al., 2006)、5C(carbon-copy chromosome conformation capture)技术(Dostie and Dekker, 2007)以及ChIA-PET(chromatin interaction analysis using paired end tag sequencing)(Li et al., 2010; 2017)等技术。如果说3C用于测定两个特定的基因座之间的相互作用频率称为“一对一”的话,4C的“一对多”则是能够评估某一特定的

位点与其它任意基因位点之间相互作用的方法,而5C是3C技术的另一种变体,它支持引物的多样性,能够测定多个位点与多个位点之间的相互作用,即“多对多”。除此之外,将3C与染色质免疫沉淀(chromatin immunoprecipitation, ChIP)相结合形成ChIA-PET,可以测定与特定的蛋白质相结合的全基因组染色质相互作用。与其它的3C衍生技术相比,ChIA-PET需要构建的文库复杂度更低,能够以较少的读取次数识别交互作用。Hi-C技术的出现极大地促进了三维基因组学的发展,表1清晰地展示了相关技术的对比。

科学家将Hi-C技术的基本原理总结为“proximity-dependent ligation”,即邻近连接(图1)。将细胞或新鲜组织样本与甲醛交联后,用限制酶进行消化,以生物素标记产生的DNA末端,将之置于稀溶液中,以有利于邻近DNA片段末端在同一复合物中的随机再连接(Dekker, 2006)。这样得到的DNA样品含有最初在空间上较紧密片段组成的连接产物。通过剪切DNA片段构建Hi-C文库,进行PCR和Illumina Hiseq测序。

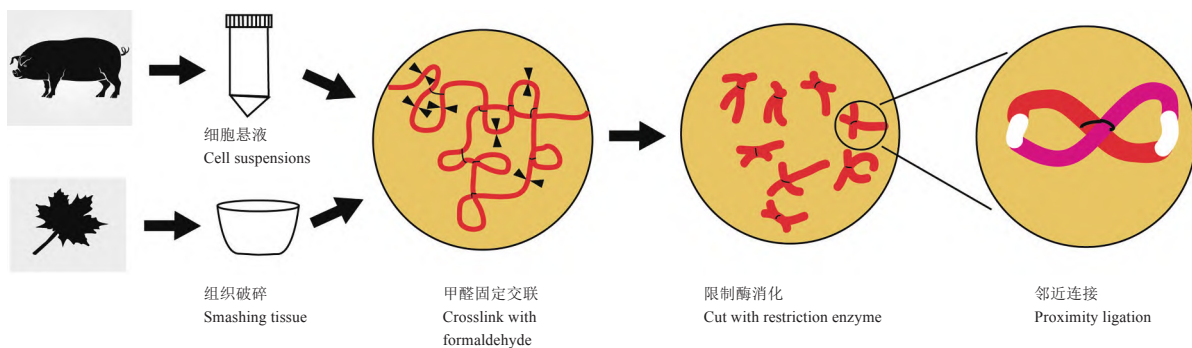


图1 Hi-C实验流程

Figure 1 The experimental process of Hi-C

Hi-C的实验阶段对条件要求严格,限制酶消化、生物素标记和连接条件都是不可忽视的。就培养中的细胞而言,细胞培养基的组成、甲醛交联的时间长短、黏附细胞的收集以及细胞的渗透环境都关乎结果的准确性(Belton et al., 2012)。科学家们追求更高质量的Hi-C结果,因此在原来技术的基础上不断进行改进。制备新鲜甲醛并在使用前进行一定倍数的稀释有助于提高重现性,防止甲醛失控聚合(Gondor et al., 2008)。获取交联细胞时,从培养皿中直接手工

刮取可以使得DNA保持完整,刮取的同时还可以在显微镜下评估交联板以记录细胞外观(Makhija et al., 2016)。交联过程在细胞悬液中进行可以提高回收率(Gollosi et al., 2018)。Nagano等(2015)发现在细胞核中连接平末端可以消除限制性片段长度偏差,扩大染色体内部和染色体之间接触的范围,还可以增强对结构特征的检测。核内连接在更大的距离范围内可以更一致地捕捉染色质相互作用,并显著降低实验噪声和偏倚。张香媛等(2017)通过优化Hi-

C实验中的交联、酶切、限制性内切酶的失活等关键步骤,制备了更高质量的Hi-C文库,使Hi-C流程更加高效、稳健、可靠。

2 Hi-C与三维基因组

Hi-C技术的发展促进了人们对基因组、转录组的研究,通过设置不同的分辨率和通量,可以解释细胞核内染色质的折叠组装、结构划分以及组件作用规律。Hi-C的数据以 N^2 矩阵的形式展示了 N 个染色质位点与 $N-1$ 个染色质位点之间的互作关系,其中蕴含着庞大的数据量,为更好地利用文库中的信息,

科学家们将基因组划分为等长的窗口,成为bin,bin窗口的长度即为分辨率,窗口越小则基因组区域划分得越细致。随着分辨率的提高,现有的研究已经证明组成染色质的层级结构单元由大到小依次为:当分辨率为1 Mbp时可构建3D空间立体模型,识别出CT结构;当分辨率为100 kbp时,能够区分Compartment A/B;当分辨率为20~40 kbp时,可以观察到TAD结构;当分辨率为1~10 kbp时,分析的对象是Loop结构。如图2所示的染色质结构在各个层级上发生的改变和修饰都会影响基因组的整体表达和调控。

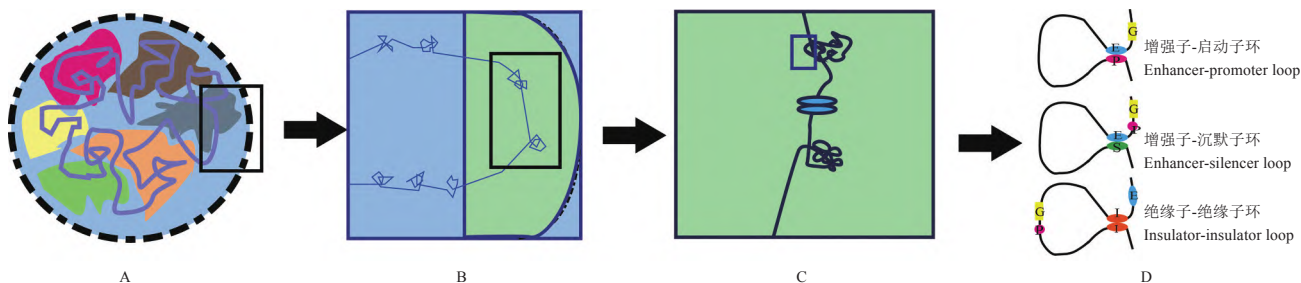


图2 染色质的层级结构(Ning et al., 2019)

注: A: 染色体领域; B: 染色体区室; C: 拓扑关联结构域; D: 染色质环

Figure 2 The hierarchical structure of chromatin (Ning et al., 2019)

Note: A: Chromosome territory; B: Chromosome compartment; C: Topological associated domain; D: Chromatin loop

2.1 染色质领域

Boveri(1909)在研究蛔虫细胞时发明了“染色体领域(CT)”一词。细胞分裂间期,每一条染色体都会在细胞核中占据一片特定的区域。而以Hi-C技术为代表的测序技术的发展为染色质结构的发现提供了平台,其中CT是基因组各级结构的最高水平,对X染色体的沉默(Deng et al., 2015; Chen et al., 2016)、DNA 的损伤修复(Mehta et al., 2013)和各种细胞分化过程(Borden and Manuelidis, 1988; Martou and Boni, 2000; Solovei et al., 2009)都很重要。随着对活细胞研究的不断深入,有学者利用Hi-C在Mbp分辨率下观察到标记物显著振荡运动和形状转换过程,而且富含活性基因的CT结构变化更活跃,证实了CT在细胞核内并不是一成不变的(Pliss et al., 2009)。不同染色体在细胞核中占据不同CT,较小的染色体通常位于细胞核的内部,较大的染色体则位于细胞核的外围。然而,有DNA含量相近但基因密度差异很大的CT,在核内的位置差别很大,例如人类18号染色体和19号染色体的DNA含量相似(分别

为85 Mbp和67 Mbp),但基因贫乏的18号染色体通常出现在细胞核外围,而基因丰富的19号染色体则位于细胞核内部(Croft, 1999)。说明基因含量是决定CT定位的关键因素。此外,同源CT之间的关系值得深入研究,目前已经观察到,对于基因密集的较小染色体,CT的同源关联更为频繁,这一发现可以用它的位置更偏向于细胞核内部来解释 (Cremer and Cremer, 2001)。

2.2 染色体区室

染色体区室是位于染色质领域的下一层级结构,将Hi-C技术和荧光标记原位杂交技术(fluorescent in situ hybridization, FISH)进行结合对同一染色体上的不同位点进行标记后,观察到在线性距离上更接近的两个位点在空间上并不是最接近的,这个结果证实了整个基因组可以被划分为两个不同的区域,后经完善称为Compartment A/B (图3A)。两种分布在染色体中是交叉存在的,每个区室内部的相互作用频率会比区室之间的频率高。两个区室所含碱基的密度不同导致性质也有很大的区别,Com-

partment A 中所含的常染色体较多,(G+C)的含量相对较高,属于开放区域。而 Compartment B 则相反,包含异染色质区域更多,基因的表达量相对较低且相对封闭(Solovei et al., 2016)。实验过程中发现从 Compartment A 转换到 Compartment B 的基因,其表达量会降低,从 Compartment B 向 Compartment A 转换的基因则相反,仅仅有一少部分的基因会随着这种转换发生相应的变化,大部分的基因表达并不会受到影响,这表明 Compartment A/B 在决定基因表达的细胞类型特异模式方面具有协同作用,但不具有决定性作用(Dixon et al., 2015)。Rao (2014)利用 Hi-C 技术与核细胞分析的方法构建人和小鼠的基因图谱时,在较高的分辨率下重现染色体区室,发现 Compartment A/B 还能够细分为更加精细的结构,在 25 kbp 的分辨率下,Compartment A 还能够分为 A1 和 A2 两个亚组 (Subcompartment), Compartment B 可划分为 B1、B2 和 B3 亚组。

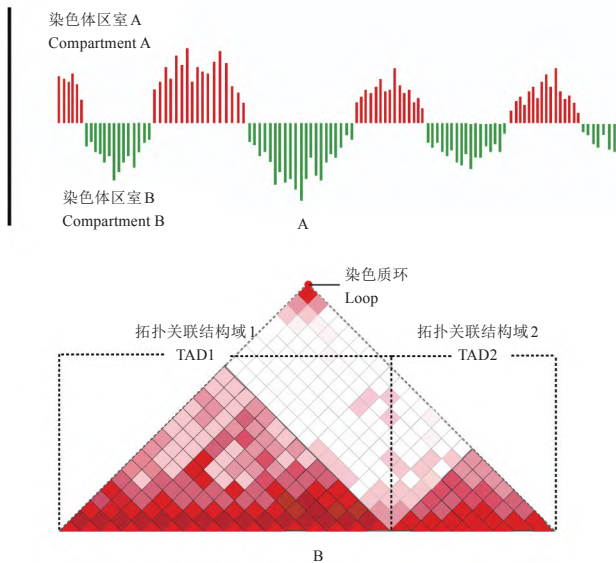


图3 染色质区室和拓扑关联结构域在热图中的示意图
注: A: Compartment A/B 的示意图; B: TAD 和 Loop 结构在热图中的示意图

Figure3 Schematic diagram of chromatin compartment and topological domain in heatmap

Note: A: Schematic diagram of Compartment A/B; B: TAD and Loop structures shown in the heat map

2.3 拓扑关联结构域

Compartment A/B 的下一级别就是拓扑关联结构域 (topological associated domain, TAD; 图 3B)。

TAD 内部染色体之间的相互作用频率远高于相邻 TAD 之间的相互作用频率,在热图中呈现为强互作的三角形区域。大量实验证明,TAD 通过将增强子与启动子之间的接触限制在同一个 TAD 中实现基因调控和功能调节(Mifsud et al., 2015),而增强子-启动子对之间的物理接触似乎足以达到激活基因的目的(Deng et al., 2012)。Dixon 在研究人和小鼠细胞复制机制的过程中,将 Hi-C 与 3C、5C 和 FISH 的数据结果进行比较后发现 TAD 的边界上富集了阻遏子 CTCF (CCCTC binding factor)、看家基因 (housekeeping genes)、短散在元件 (short interspersed repetitive element, SINE)、粘连蛋白复合物、H3K4me3 和 H3K36me3 组蛋白修饰位点以及转录起始位点 (transcription start sites, TSS) 等元件,这些因子在建立基因组的 TAD 方面发挥了作用。目前关注最多的两种蛋白为 CTCF 与粘连蛋白复合物 (cohesin complex),二者均定位于 TAD 的边界部分,在间期染色质中发挥着重要的作用。为了更深入地了解 TAD,克服现有技术缺陷,科学家们不断开发更高效的方法,SpectralTAD (Cresswell et al., 2020) 就是其中之一。它不仅能够自动校对参数,选择合适的测序深度和分辨率,还能够提高 Hi-C 数据的峰度,是研究 TAD 的有力工具。

2.4 染色质环

由于染色质在细胞核内是高度浓缩折叠的,使得很多在线性距离上距离较远的位点之间互作概率增加。提高 Hi-C 的分辨率会出现呈径向对称的焦点峰,每一个峰代表了两个比相邻基因座之间互作更强烈的基因座组合,称为染色质环 (chromatin loop; 图 3B),由启动子与增强子的长距离互作产生。环 (loop) 是染色质保持结构稳定的关键,也是影响转录和复制顺利进行的重要因素(Ardakany et al., 2020)。大部分 Loop 都很短 (< 2 Mbp),并且在细胞类型之间以及物种之间(如人和老鼠)具有很强的保守性。启动子增强子环 (promoter-enhancer loop) 是最常见的,与基因激活有关,但并不是唯一的染色质环,增强子与增强子之间的相互作用、启动子与启动子之间的相互作用也能够形成复杂的 Loop 结构。Loop 最显著的特性是边界中存在一对以聚收敛的方向出现的 CTCF 基序。关于 Loop 的形成,目前还存在很多的争议,大多数的科学家相信“环挤压假说(Haarhuis et al., 2017)”,为了从不同的角度更深入地研究 Loop,

众多针对染色质环的Hi-C辅助工具出现:Peakachu (Salameh et al., 2020)能够在3000万染色质间Hi-C读长中探测到高分辨率Loop,对于研究染色体空间结构有重要的意义。

3 Hi-C技术的主要应用

3.1 染色体三维结构构建

Hi-C技术最典型的应用就是研究染色体的三维立体结构,构建其三维模型,加深对生物基因组的理解和研究(图4)。Dekker (2006)最先利用Hi-C技术研究酵母细胞减数分裂过程中着丝粒和同源染色体的空间结构,揭示了位于不同染色体上基因位点之间的相对空间分布,描述Ⅲ号染色体为“一个扭曲的环状”。之后,Hi-C技术得到更广泛的应用,越来越多的三维基因组模型被构建出来。Dixon等(2015)绘制了人的H1时期的胚胎干细胞以及中内胚层干细胞(mesendoderm, ME)、间充质干细胞(mesenchymal stem, MS)、神经祖细胞(neural progenitor, NP)、滋养层细胞(trophoblast-like, TB)的全基因组染色质相互作用图,发现广泛的染色质重组现象,证实TAD是染色体区室动态变化的单位,为染色质动力学研究提供了全局观点。华中农业大学实验室(Xie et al., 2015)参考Moissiard等(2012)发表过的拟南芥染色质全基因组互作数据,对拟南芥的生态型进行了测序并验证了结果的可重复性,证明Hi-C技术可以用于植物细胞的组装以及染色质相互作用的测定,测定结果高度准确。至今,科学家们已经实现了对人、小

鼠、大鼠、哺乳动物、微生物类(真菌和原核生物)、水稻、玉米、大豆多种动植物的Hi-C测序和分析,促进了对染色质三维空间结构信息的更深的了解。

3.2 辅助基因组组装

染色质间的互作强度并没有其内部的互作强度高,而且同一条染色体上近距离基因之间的相互作用会比远距离的基因作用更强。利用这一原理,科学家们利用Hi-C辅助基因组组装(Lachesis),显著提高基因组组装的准确率(Wang et al., 2020)。Burton等(2013)证实了Hi-C生成的数据集是一个丰富的远程信息源,可以用于将基因组序列聚类(cluster)、排序(order)和定向(orient)到染色体,他在此基础上成功地利用Hi-C技术辅助组装了人和小鼠的基因组。Xie等(2015)利用Hi-C技术组装了拟南芥的基因组,表明该技术在植物细胞中同样适用。Bian等(2020)在没有参考基因组的前提下,利用Hi-C技术组装了一种濒危鱼类的基因组,对濒危物种的研究提供了借鉴意义。除此之外,埃及伊蚊(Dudchenko et al., 2017)和甘蔗(Zhang et al., 2018)等动植物基因组组装试验也都获得了成功,且准确率高,经济高效。

3.3 探究基因功能和表达差异

Hi-C技术除了能够辅助组装基因组之外,还能够分析特定基因的功能及其在不同条件下的表达差异。有科学家在分析秀丽隐杆线虫X染色体和常染色体在分子水平上的拓扑结构差异时,利用3C和Hi-C相结合的方法生成了全基因组染色质互作图。发现X染色体上的TAD边界比常染色体的边界更加明显和丰富,揭示了剂量补偿复合物(dosage compensation complex, DCC)通过与REX位点结合来介导TAD边界的形成,从而影响X染色体整体拓扑结构的过程。它不仅能凝集和分解有丝分裂和减数分裂期的染色体,而且是调控基因表达的重要元件(Crane et al, 2015)。

3.4 疾病中的应用

很多疾病的发生都与染色质的空间结构存在着密切的联系,利用Hi-C技术可以准确地判断基因致病机制或发现致病基因,辅助制定治疗方案。了解基因是如何在细胞核中缺失、扩增和失调的,使治疗策略更有针对性。

在正常人群中,有5%的人类基因组在结构上是

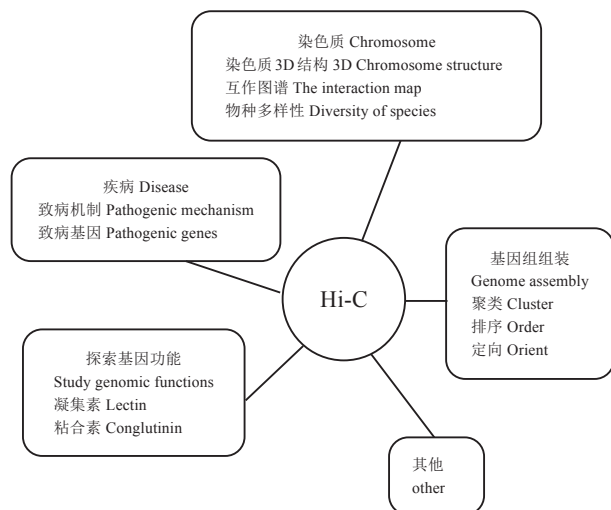


图4 Hi-C的应用速览

Figure 4 Application of Hi-C

可变的,形式包括基因组片段的缺失(deletion)、重复(duplication)、倒位(inversion)、插入(insertion)和易位(translocation)等,统称为基因组结构变异(structure variations, SVs),作为遗传疾病的主要原因之一备受关注。随着生物技术的迅速发展,越来越多的证据证明SVs通过作用于TAD导致错误表达和病变(Benko et al., 2011; Spielmann and Mundlos, 2013)。不同类型的SVs在Hi-C相互作用矩阵中表现为异常的信号增强或减弱。因而科学家们推测从Hi-C数据中观察到的强染色体内信号有助于发现癌症基因组中的染色体重排和染色质易位。将单细胞Hi-C技术与其它单细胞转录组和表观遗传学分析相结合,有助于肿瘤治疗策略的选择,Wu等(2020)整合分析了Hi-C数据和ChIP-seq结果后,观察到转录增强子与癌症基因相关,其中LcsMYC-1(肺癌特异性增强子RNA1)可能有助于原癌基因的表达。这些发现加深了人们对增强子的DNA、RNA和蛋白质特性的理解。

除此之外,Hi-C数据还可以用于重建染色体单倍型(图4),这一发现使针对染色质结构和基因表达的研究都能够在二倍体的基础上进行。Dixon等(2015)利用Hi-C数据集和全基因组测序产生了包含93.5% H1杂合子变体的染色体跨度单倍型。还观察到预测的单倍型和成对序列在“长插入”大小的数据集读取中高度一致,表明重建的单倍型质量很高。综上所述,Hi-C和其它衍生生物技术的高速发展必然会为科学家探索基因功能、攻克疾病和推动医学进步提供有力的支持。

4 展望

染色质与生物遗传进化息息相关,其活性与功能由染色质的三维空间构象决定,Hi-C的出现与发展为三维基因组学提供了技术支持,人们不断追求着更高的分辨率以便对染色质有更深入的了解。尽管对染色质和基因的研究已经取得了显著的成就,但是相比于整个生物基因组来说无异于九牛一毛,未来还有无限的空间等待着我们去探索。

Hi-C技术的广泛应用极大地促进了生物信息学的发展,染色质不同的层级结构也为人所知,CT的存在、活跃的Compartment A和闭塞的Compartment B、TAD内部的显著互作以及与启动子-增强子有着密切联系的Loop结构都是当下研究的热点。随着人类基因组计划的完成,三维基因组的大门已经敞

开,Hi-C以及其它高通量技术即将发挥更大的作用,我们需要更进一步地了解染色质三维结构、加快成果转化,不断改进技术的缺陷,提高精确度和准确度,为探索更深层次的生物技术提供科学基础,相信基因组学会在疾病的诊断、靶向药物的研发等方面都给人可喜的成果。

作者贡献

房瑞新完成论文初稿的写作;董依萌、刘汇涛和张然然协助论文的修改;刘华淼、王磊、王天骄和王洪亮协助资料的收集;刘欣、李洋和邢秀梅是项目的构思者及负责人,指导论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由中国农业科学院基本科研业务费专项(1610342020027; 1610342020025; 1610342020021)共同资助。

参考文献

- Ardakany A.R., Gezer H.T., and Ay F., 2020, Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation, *Genome Biology*, 21(1): 256-273.
- Belton J.M., McCord R.P., Gibcus J.H., Naumova N., Zhan Y., and Dekker J., 2012, Hi-C: a comprehensive technique to capture the conformation of genomes, *Methods*, 58 (3): 268-276.
- Benko S., Gordon C.T., Mallet D., Sreenivasan R., Thauvin-Robinet C., Brendehaug A., Thomas S., Bruland O., and David M., 2011, Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development, *Journal of Medical Genetics*, 48(12): 825-830.
- Bian L., Li F.H., Ge J.L., Wang P.F., Chang Q., Zhang S.N., Li J., Liu C.L., Liu K., Liu X.T., Li X.M., Chen H.J., Chen S.Q., Shao C.W., and Lin Z.S., 2020, Chromosome-level genome assembly of the greenfin horse-faced filefish (*Thamnaconus septentrionalis*) using Oxford Nanopore PromethION sequencing and Hi-C technology, *Molecular Ecology Resources*, 20(4): 1069-1079.
- Borden J., and Manuelidis L., 1988, Movement of the x chromosome in epilepsy, *Science*, 242(4886): 1687-1691.

- Boveri T., 1909, Die blastomerenkerne von ascaris megaloccephala und die theorie der chromosomen individualität, *Archiv für Zellforschung*, 3: 181.
- Burton J.N., Adey A., Patwardhan R.P., Qiu R.L., Kitzman J.O., and Shendure J., 2013, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions, *Nature Biotechnology*, 31(12): 1119-1125.
- Chen C.K., Blanco M., Jackson C., Aznauryan E., Ollikainen N., Surka C., and Guttman M., 2016, Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing, *Science*, 354(6311): 468-472.
- Collins F.S., Morgan M., and Patrinos A., 2003, The human genome project: lessons from large-scale biology, *Science*, 300 (5617): 286-290.
- Crane E., Bian Q., Mccord R.P., Lajoie B.R., Wheeler B.S., Ralston E.J., Uzawa S., Dekker J., and Meyer B.J., 2015, Condensin-driven remodelling of X chromosome topology during dosage compensation, *Nature*, 523 (7559): 240-244.
- Cremer T., and Cremer C., 2006, Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories, *European Journal of Histochem*, 50(3): 161-176.
- Cremer T., and Cremer C., 2001, Chromosome territories, nuclear architecture and gene regulation in mammalian cells, *Nature Reviews Genetics*, 2(4): 292-301.
- Cresswell K.G., Stansfield J.C., and Dozmorov M.G., 2020, Correction to: SpectralTAD: an R package for defining a hierarchy of topologically associated domains using spectral clustering, *BMC Bioinformatics*, 21(1): 319.
- Croft J.A., Bridge J.M., Boyle S., Perry P., Trague P., and Bickmore W.A., 1999, Differences in the localization and morphology of chromosomes in the human nucleus, *Journal of Cell Biology*, 145(6): 1119-1131.
- Dekker J., Rippe K., and Dekker M., 2002, Capturing chromosome conformation, *Science*, 295(5558): 1306-1311.
- Dekker J., 2006, The three "C" s of chromosome conformation capture: controls, controls, controls, *Nat Methods*, 3 (1): 17-21.
- Dekker J., Marti-Renom M.A., and Mirny L.A., 2013, Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data, *Nature Reviews Genetics*, 14(6): 390-403.
- Deng X., Ma W.X., Ramani V., Hill A., Yang F., Ay F., Distech C.M., Berletch J.B., Blau C.A., Shendure J., Duan Z.J., and Noble W.S., 2015, Bipartite structure of the inactive mouse X chromosome, *Genome Biology*, 16(1): 152.
- Deng W., Lee J., Wang H., Miller J., Reik A., Gregory P.D., Dean A., and Blobel G.A., 2012, Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor, *Cell*, 149(6): 1233-1244.
- Dixon J.R., Jung I., Selvaraj S., Shen Y., Antosiewicz-Bourget J.E., Lee A.Y., Ye Z., Kim A., Rajagopal N., Xie W., Diao Y.R., Liang J., Zhao H.M., Lobanenko V.V., Eckert J.R., Thomson J.A., and Ren B., 2015, Chromatin architecture reorganization during stem cell differentiation, *Nature*, 518(7539): 331-336.
- Dostie J., and Dekker J., 2007, Mapping networks of physical interactions between genomic elements using 5C technology, *Nature Protocols*, 2(4): 988-1002.
- Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K., Hoeger M., Durand N.C., Shamim M.S., Machol I., Lander E.S., Aiden A.P., and Aiden E.L., 2017, De novo assembly of the aedes aegypti genome using Hi-C yields chromosome-length scaffolds, *Science*, 356(6333): 92-95.
- Feuk L., Carson A.R., and Scherer S.W., 2006, Structural variation in the human genome. *Nature Reviews Genetics*, 7(2): 85-97.
- Gollosi R., Sanders J., and Mccord R.P., 2018, Iteratively improving Hi-C experiments one step at a time, *Methods*, 142: 47-58.
- Gondor A., Rougier C., and Ohlsson R., 2008, High-resolution circular chromosome conformation capture assay, *Nature Protocols*, 3 (2): 303-313.
- Haarhuis J.H.I., Weide R.H., Blomen V.A., Yáñez-Cuna J.O., Amendola M., van Ruiten M.S., Krijger P.H.L., Teunissen H., Medema R.H., van Steensel B., Brummelkamp T.R., Thijn R., Wit D.E., and Rowland B.D., 2017, The cohesin release factor wapl restricts chromatin loop extension, *Cell*, 169(4): 693-707.
- Haberer G., and Zhang X.Q., 2017, A chromosome conformation capture ordered sequence of the barley genome, *Nature*, 544(7651): 427-433.
- Hughes J.R., Roberts N., Gowan M.S., Hay D., Giannoulitou E., Lynch M., De Gobbi M., Taylor S., Gibbons R., and Higgs D.R., 2014, Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment, *Nature Genetic*, 46(2): 205-212.
- Li G., Fullwood M.J., Han X., Mulawadi F.H., Velkov S., Vega V., Ariyaratne P.N., Bin Mohamed Y., Ooi H.S., Tennakoon C., Wei C.L., Ruan Y.J., and Sung W.K., 2010,

- ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing, *Genome Biology*, 11(2):1-13.
- Li X., Luo O.J., Wang P., Zheng M., Wang D., Piecuch E., Zhu J.J., Tian S.Z., Tang Z., and Li G., 2017, Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions, *Nature Protocols*, 12(5): 899-915.
- Lieberman-Aiden E., Berkum N. V., Williams L., Imakaev M., Ragozcy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M. O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A., Lander E.S., and Dekker J., 2009, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science*, 326(5950): 289-293.
- Makhija E., Johun D.S., and Shivashankar G.V., 2016, Nuclear deformability and telomere dynamics are regulated by cell geometric constraints, *PNAS*, 113 (1) : E32-E40.
- Marieke S., Petra K., Erik S., Yuri M., Rob W., Elzo D.W., Bas V.S., and Wouter D.L., 2006, Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C), *Nature Genetic*, 38(11): 1348-1354.
- Martou G., and Boni D.U., 2000, Nuclear topology of murine, cerebellar purkinje neurons: Changes as a function of development, *Experimental Cell Research*, 256(1): 131-139.
- Mehta I.S., Kulshreshtha M., Chakraborty S., Kolthur-Seetharam U., and Rao B.J., 2013, Chromosome territories reposition during DNA damage-repair response, *Biophysical Journal*, 14(12): 135-150.
- Mifsud B., Tavares-Cadete F., Young A.N., Sugar R., Schoenfelder S., Ferreira L., Wingett S.W., Andrews S., Grey W., Ewels P.A., Herman B., Happe S., Higgs A., Leproust E., Follows G.A., Fraser P., Luscombe N.M., and Osborne C.S., 2015, Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C, *Nature Genetic*, 47(6): 598-606.
- Nagano T., Várnai C., Schoenfelder S., Javierre B.M., Wingett S.W., and Fraser P., 2015, Comparison of Hi-C results using in-solution versus in-nucleus ligation, *Genome Biology*, 16(1): 175-188.
- Ning C.Y., He M.N., Tang Q.Z., Zhu Q., Li M.Z., and Li D.Y., 2019, Advances in mammalian three-dimensional genome by using Hi-C technology approach, *Yichuan (Hereditas)*, 41(3): 215-233.(宁椿游, 何梦楠, 唐茜子, 朱庆, 李明洲, 李地艳, 2019, 基于 Hi-C 技术哺乳动物三维基因组研究进展, *遗传*, 41(3): 215-233.)
- Pliss A., Malyavantham K., Bhattacharya S., Zeitz M., and Berezney R., 2009, Chromatin dynamics is correlated with replication timing, *Chromosoma*, 118(4): 459-470.
- Rao S.S., Huntley M. H., Durand N. C., Stamenova E. K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., and Aiden E.L., 2014, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, *Cell*, 159(7): 1665-1680.
- Salameh T.J., Wang X.T., Song F., Zhang B., Wright S.M., Khunsriraksakul C., Ruan Y.J., and Yue F., 2020, A supervised learning framework for chromatin loop detection in genome-wide contact maps, *Nature Communications*, 11(1): 773-782.
- Solovei I., Kreysing M., Lanctôt C., Kösem S., Peichl L., Cremer T., and Joffe B., 2009, Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution, *Cell*, 137(2): 356-368.
- Solovei I., Thanisch K., and Feodorova Y., 2016, How to rule the nucleus: divide et impera, *Current Opinion in Cell Biology*, 40(6): 47-59.
- Spielmann M., and Mundlos S., 2013, Structural variations, the regulatory landscape of the genome and their alteration in human disease, *Bioessays*, 35(6): 533-543.
- Wang J.W., Liu W.Z., Zhu D.Z., Hong P., Zhang S.Z., Xiao S.J., Tan Y., Chen X., Xu L., Zong X.J., Zhang L.S., Wei H.R., Yuan X.H., and Liu Q.Z., 2020, Chromosome-scale genome assembly of sweet cherry (*Prunus avium* L.) cv. Tieton obtained using long-read and Hi-C sequencing, *Horticulture Research*, 7(1): 122.
- Wu Y.C., Yang Y., Gu H.Y., Tao B.R., Zhang E., Wei J.H., Wang Z., Liu A.F., Sun R., Chen M.M., and Mao R.F., 2020, Multiomics analysis reveals the functional transcription and potential translation of enhancers, *International Journal of Cancer*, 147(8): 2210-2224.
- Xie T., Zheng J.F., Liu S., Peng C., Zhou Y.M., Yang Q.Y., and Zhang H.Y., 2015, De Novo plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana*, *Molecular Plant*, 8(3): 489-492.
- Zhang C. Y., He C., and Ye B. Y., 2017, Optimization and quality control of genome-wide Hi-C library preparation, *Yichuan (Hereditas)*, 39(9): 847-855.(张香媛, 何

超, 叶丙雨, 2017, 全基因组染色质相互作用 Hi-C 文库制备的优化及其质量控制, 遗传, 39(9): 847-855.)

Zhang J.S., Zhang X.T., Tang H.B., Zhang Q., Hua X.T., Ma X.K., Zhu F., Jones T., Zhu X.G., Bowers J., Wai C.M., Zheng C.F., Shi Y., Chen S., Xu X.M., Yue J.J., Nelson D.R., Huang L.X., Li Z., Xu H.M., Zhou D., and Wang Y.J., 2018, Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L, Nature Genetic, 50(11): 1565-1573.

Zhang J.X., Lei Y.Y., Wang B.T., Li S., Yu S., Wang Y., Li H., Liu Y.X., Ma Y., Dai H.Y., Wang J.H., and Zhang Z.H., 2020, The high-quality genome of diploid strawberry

(*Fragaria nilgerrensis*) provides new insights into anthocyanin accumulation, Plant Biotechnology Journal, 18(9): 1908-1924.

Zhou Q., Guo X.Y., Huang Y., Gao H.Y., Xu H., Liu S.S., Zheng W.W., Zhang T.S., Tian C.X., Zhu C.H., Lin H.R., and Chen S.L., 2020, De novo sequencing and chromosomal-scale genome assembly of leopard coral grouper, *Plectropomus leopardus*, Molecular Ecology Resources, 20(5): 1403-1413.

(责任编辑 庞晓鑫 罗厚枚)