

## 三维基因组学研究进展

张富涵, 沈宗毅, 喻长远, 杨昭

北京化工大学 生命科学与技术学院, 北京 100029

张富涵, 沈宗毅, 喻长远, 等. 三维基因组学研究进展. 生物工程学报, 2020, 36(12): 2791–2812.

Zhang FH, Shen ZY, Yu CY, et al. Advances in three-dimensional genomics. Chin J Biotech, 2020, 36(12): 2791–2812.

**摘要:** 三维基因组学是一门研究基因组三维空间结构与功能的新兴学科, 主要研究基因组序列在细胞核内的三维空间构象, 及其对 DNA 复制、DNA 重组、基因表达调控等生物过程的生物学效应。自染色质构象捕获技术 (3C) 出现后, 三维基因组学相关研究领域飞速发展。借助于 3C 及其衍生技术、Hi-C 和 ChIA-PET 等技术, 科学家能对各类物种的三维基因组进行更为深入的研究, 从而揭示微生物、植物和动物基因组的三维空间构象、染色质的相互作用模式、转录调控以及不同生物学性状的形成机制; 挖掘与生命活动和疾病相关的关键基因和信号通路; 推动农业科学、生命科学和医学等领域的快速发展。文中就三维基因组学研究进展作一综述, 主要阐述三维基因组学的概念和研究技术的发展及其在农业科学、生命科学和医学等领域的应用, 尤其是肿瘤领域所取得的阶段性研究成果。

**关键词:** 三维基因组, 三维空间构象, 基因表达调控, 染色质构象捕获技术 3C, 肿瘤

## Advances in three-dimensional genomics

Fuhan Zhang, Zongyi Shen, Changyuan Yu, and Zhao Yang

College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

**Abstract:** Three-dimensional (3D) genomics is an emerging discipline that studies the 3D spatial structure and function of genomes, focusing on the 3D spatial conformation of genome sequences in the nucleus and its biological effects on biological processes such as DNA replication, DNA recombination and gene expression regulation. The invention of chromosome conformation capture (3C) technology speeds up the research on 3D genomics and its related fields. Furthermore, the development of 3C-based technologies, such as the genome-wide chromosome conformation capture (Hi-C) and chromatin interaction analysis using paired-end tag sequencing (ChIA-PET), help scientists get insight into the 3D genomes of various species. Aims of 3D genomics are to reveal the spatial genome organization, chromosomal interaction patterns, mechanisms underlying the transcriptional regulation and formation of biological traits of microorganism, plant, animal. Additionally, the identification of key genes and signaling pathways associated with

**Received:** April 9, 2020; **Accepted:** June 22, 2020

**Supported by:** Fundamental Research Funds for the Central Universities (No. buctrc201910), Beijing-Tianjin-Hebei Basic Research Cooperation Special Project (No. 19JCZDJC65800(Z)), National Key Research and Development Program of China (No. 2017YFA0105900).

**Corresponding author:** Zhao Yang. Tel: +86-10-64421335; E-mail: yangzhao@mail.buct.edu.cn

中央高校基本科研业务费专项资金 (No. buctrc201910), 京津冀基础研究合作专项 (No. 19JCZDJC65800(Z)), 国家重点研发计划 (No. 2017YFA0105900) 资助。

biological processes and disease via chromosome 3C technology boosts the rapid development of agricultural science, life science and medical science. This paper reviews the research progress of 3D genomics, mainly in the concept of 3D genomics, the development of chromosome 3C technologies and their applications in agricultural science, life science and medical science, specifically in the field of tumor.

**Keywords:** three-dimensional genomics, three-dimensional spatial conformation, gene expression regulation, chromosome conformation capture technology, tumor

基因组 DNA 在细胞核中并不是简单的线性状态, 而是通过高度折叠、浓缩成具有特定高级空间构象的染色质形式存在, 并储存于细胞核内作为遗传信息的载体<sup>[1]</sup>。基因组学就是将一个生物体的所有基因进行集体表征和量化, 研究基因组的结构、功能、进化、定位等问题, 并分析它们之间的相互关系以及对生物产生的影响<sup>[2]</sup>。基因组学研究自启动以来已经经历了 30 年的发展。1990 年, 美国、英国、法国、德国、日本和中国科学家共同启动了“人类基因组计划”(Human genome project, HGP)<sup>[3]</sup>, 该计划定义了人类基因组中的重要基因及线性结构。2003 年, 美国国家人类基因组研究所启动了“DNA 元件百科全书计划”(Encyclopedia of DNA elements, ENCODE)<sup>[4]</sup>, 旨在解析人类基因组中的所有功能性元件, 研究结果表明, 80.4% 的人类基因组具有生物学活性, 包括基因、转录本、转录调控元件、DNA 甲基化模式、组蛋白修饰和染色质状态等。

随着基因组学的发展, 科学家们对基因组的研究从一维(基因序列)、二维(不同序列的相互作用)层面, 逐渐深入到三维(染色质的空间构象)、四维(序列随时间的变化)层面。三维基因组学是对基因组序列在细胞核内的三维空间构象, 及其在基因复制、重组、表达等生物过程中发挥的功能进行研究的新兴学科。2002 年, Dekker 等<sup>[5]</sup>提出染色质构象捕获技术(Chromatin conformation capture, 3C), 三维基因组的相关研究飞速发展。在随后几年, 基于 3C 的衍生测序技术如可实现一点到多点测序的环状染色质构象捕获技术(Circular chromatin conformation capture

technology, 4C)<sup>[6]</sup>、可用于多点到多点测序的碳拷贝染色质构象捕获技术(Chromatin conformation capture carbon copy technology, 5C)<sup>[7]</sup>的出现, 使得科学家们对染色质三维结构及相互作用的认知愈加清晰。2009 年, Lander 与 Dekker<sup>[8]</sup>研究组共同提出了高通量染色体构象捕获(High-throughput chromosome conformation capture, Hi-C)技术, 使得科学家们对三维基因组学的认知取得突破性进展。在此基础上, Hi-C 衍生技术如原位 Hi-C<sup>[9]</sup>、DLO Hi-C (Digestion-ligation-only Hi-C)<sup>[10]</sup>、ChIA-PET (Chromatin interaction analysis using paired end tag sequencing)<sup>[11]</sup>等技术层出不穷。这些新兴技术的出现, 使得科学家能够鉴定不同基因、转录因子与调控元件之间的互作模式, 挖掘关键基因或关键转录因子, 进一步解读基因的功能<sup>[12]</sup>, 为多组学协同开展染色质的互作分析(基因互作)、动植物性状的分析、疾病发生机制的揭示等提供理论和技术支持。2015 年由 Dekker 等<sup>[13]</sup>牵头, 美国国家卫生研究院资助的一个全新的全球合作项目——“4D 核体计划”开始实施, 在未来的几年时间里科学家们将从空间(三维)和时间(四维)角度来研究细胞核结构形成原理, 探索细胞核对基因表达、细胞功能以及发育和疾病发生的影响。“4D 核体计划”的发展, 必能加深对生命科学的探索, 加强对生命现象和规律的认知, 为生物学的发展带来新的契机。本文就三维基因组学研究进展作一综述, 包括三维基因组学及其研究技术的发展、三维基因组学在不同领域的应用, 尤其是肿瘤领域所取得的阶段性研究成果。

## 1 三维基因组

真核生物的基因组通过线性 DNA 多层次地折叠成为染色质,以特定的三维空间构象存在于细胞核内。染色质的三维结构影响着基因的表达调控、DNA 的复制及重组。Bickmore 等<sup>[14]</sup>对染色体分裂间期的细胞核进行分析,发现细胞核内存在较长片段的核染色质区间和染色质疆域(Chromosome territory, CT),同时存在短片段的增强子-启动子连接区域,这些染色质三维结构对细胞的基因表达和调控具有重要影响。借助 Hi-C 等技术,科研人员们发现了拓扑关联结构域(Topologically associated domains, TADs)的存在, TADs 作为基因组折叠的基本单位,稳定存在于各类物种的细胞中,并在一定范围内影响基因的表达<sup>[15]</sup>。细胞核内的 TADs 存在于相对较大的结构单元——染色质区室(Chromatin compartment)中,染色质区室是基因组表观状态的体现,与染色质活性密切相关。在 TADs 的内部还存在着更为精细的折叠结构,被称为染色质环(Chromatin loop, CL),通常由启动子和增强子(远端)相互作用形成,是直接调控基因表达的基础功能单元<sup>[16]</sup>。

### 1.1 染色体疆域

早在 20 世纪初,细胞学家对一些动植物细胞进行研究时,就发现染色质在细胞核内并不是随机分布的,不同染色质占据不同的空间<sup>[17]</sup>。Cremer 等<sup>[18]</sup>发现在活细胞的间期,细胞核内的染色质组织占据了一块特定的不重合区域,并称此区域为染色质疆域(CT,图 1)。CT 在细胞核的定位与基因密度相关,低基因密度的染色质倾向于朝核外围定位,而高基因密度的染色质则占据细胞核中更中心的位置<sup>[19]</sup>,还有研究发现在不同的细胞复制时期和不同的基因类型中,CT 所占据的位置不同<sup>[20]</sup>。每条染色质被限制在特定的核空间内,不同的染色质之间只有在 CT 边界处才会重叠<sup>[21]</sup>,CT 重叠区域的存在可能是由染色质纤维被动混合引起的,也可能受到细胞的易位频率

或转录状态的影响<sup>[22]</sup>。CT 间的重叠程度对基因组的功能和稳定性都具有重要意义<sup>[23]</sup>。Mehta 等<sup>[24]</sup>研究了在 DNA 损伤修复过程中 CT 的变化,发现 DNA 损伤诱导了大规模的 CT 区域重新定位,其中包括部分 CT 从核内部向外部转移,但这是一个可逆的过程,在修复结束后,CT 会重新占据与未受损细胞类似的位置,这些研究进一步证明了 CT 在细胞核中存在的稳定性。

### 1.2 染色体区室

当把染色质进行放大时,可以观察到在染色质内部仍然存在着相互间隔的“区块”,相邻“区块”间的互作模式各不相同,线性距离较远的“区块”也可以发生相互作用。2009 年, Lieberman-Aiden 等<sup>[8]</sup>首次利用 Hi-C 技术揭示了这些“区块”的真面目,提出了染色质三维空间的另一个重要特征,即染色质区室,并将这些区块划分为 A Compartment 和 B Compartment 两类(图 1)。A Compartment 为开放染色质区室,与常染色质、基因富集区域以及转录活跃区域相关;而 B Compartment 为封闭染色质区室,经常出现在异染色质、基因沙漠和低转录区域,基因表达量比 A Compartment 低,该特征与表观遗传学特征存在高度相关性。A Compartment 区域富集了更多与转录激活相关的组蛋白标签如 H3K36me3,而 B Compartment 则富集了更多抑制性组蛋白标签如 H3K27me3。2014 年, Rao 等<sup>[25]</sup>利用原位 Hi-C 技术,通过观察这些区室的远程互作模式将 Compartment A/B 进一步划分为 5 类亚区室(Subcompartment),即 A1、A2 和 B1、B2、B3。在细胞核空间中, A Compartment 更靠近核内部, B Compartment 则主要定位于核纤层附近,该模式的划分也与细胞学上的常染色质和异染色质分布情况相对应<sup>[26]</sup>。

近年来,科学家们对染色质区室的形成机制进行了探究。Steensel 等发现核纤维层对异染色质

存在吸引力<sup>[27]</sup>, Menon 等发现同类区室更容易相互靠近, 且活性染色质具有更强的移动能力<sup>[28]</sup>。2019年6月, 麻省理工学院医学工程与科学研究所及慕尼黑大学医学生物学系的研究组合作<sup>[29]</sup>, 揭示了染色质区室形成的驱动机制。该研究指出, 区室化的形成并不依赖于常染色质间的相互作用, 更多的是依赖于异染色质间的相互作用, 且染色质分离更有利于区室化的形成; 异染色质与核纤层之间的相互作用对典型细胞核结构的建立发挥着关键作用, 这些发现为细胞核形态学相关研究提供了大量参考依据。虽然染色质区室在细胞

核中的位置相对固定, 但在生长发育、疾病发生过程中存在着大量 A Compartment 和 B Compartment 间相互转化的情况, 说明染色质区室具有较高的可塑性, 并且与细胞特异性基因的表达有一定相关性。

### 1.3 拓扑关联结构域

2012年, Dekker 研究组<sup>[30]</sup>在小鼠失活的 X 染色体中心发现了一系列离散的、大小在 200 kb–1 Mb 之间的 TADs 结构 (图 1)。同时 Dixon 等<sup>[31]</sup>在研究人类和小鼠胚胎干细胞的染色质三维结构时, 也发现了 TADs 的存在。TADs 作为细胞内染

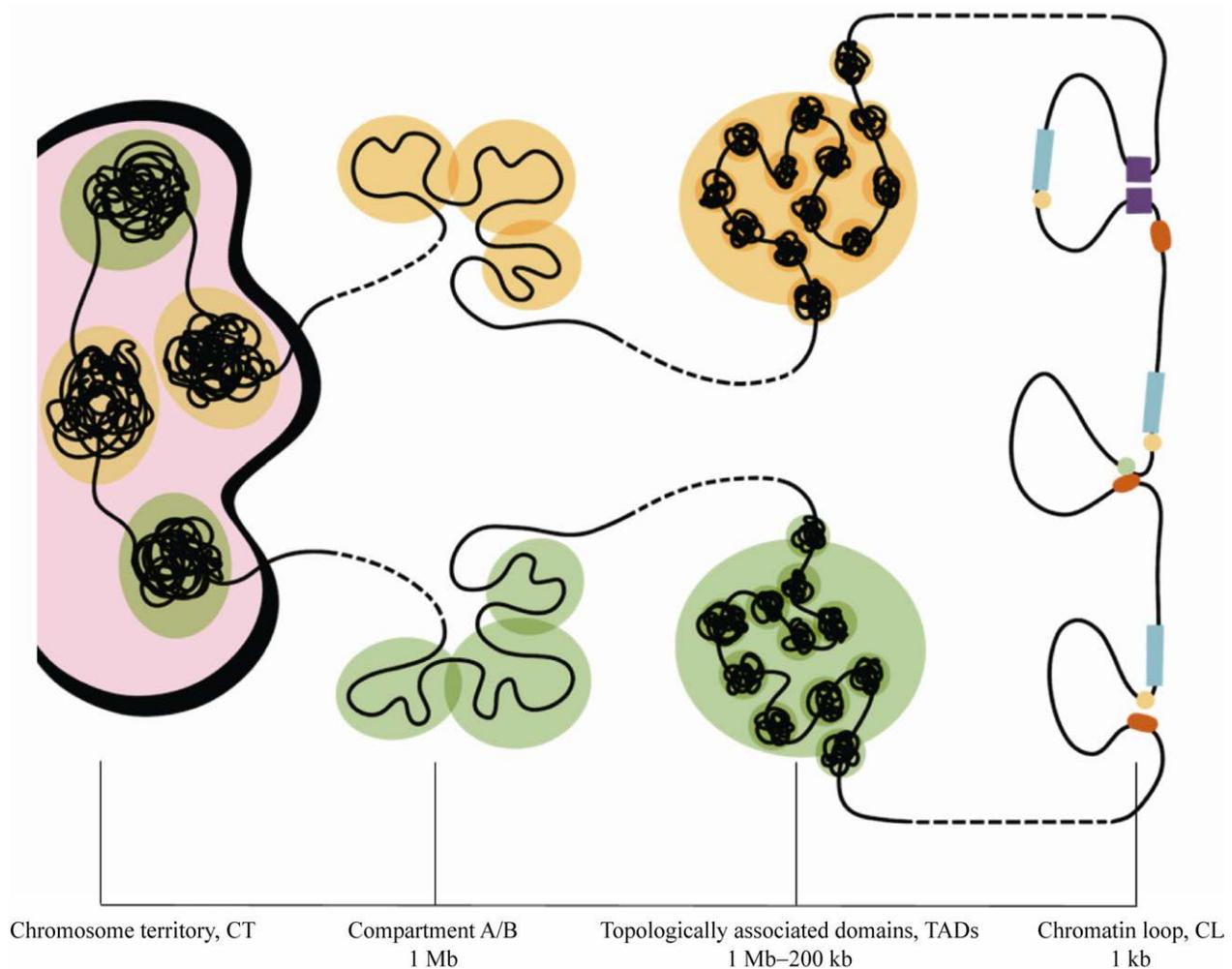


图1 三维基因组结构示意图

Fig. 1 3D schematic of genome structure.

色质折叠的二级结构单元,是一个高度自关联的连续区域,相邻区域间具有明显的边界,每一个 TADs 可形成一个独立的调控单元<sup>[32]</sup>。TADs 不仅存在于人和小鼠等哺乳动物的细胞中,而且在非哺乳动物如斑马鱼<sup>[33]</sup>、果蝇<sup>[34]</sup>、鸭子<sup>[35]</sup>、线虫<sup>[36]</sup>,以及植物如棉花<sup>[32]</sup>、水稻<sup>[37]</sup>、穿心莲<sup>[38]</sup>等生物中也发现了 TADs 结构的存在,不同细胞中 TADs 的位置相对稳定,定位也具有一定的保守性,即使是在分化过程中 TADs 也呈现出相对稳定的状态,只是相互作用的频率可能有所变化<sup>[31]</sup>。

TADs 作为调控局部基因表达的稳定结构单元,其结构特征和形成机制正在不断被揭示。当 Hi-C 互作谱图分辨率提高到 40 kb 以上时, TADs 在热图上呈现出大小不一且具有明显间隔的“三角形”<sup>[31]</sup>,这些“三角形”的边界被称为 TAD 边界(一般情况下植物中 TAD 边界因缺少绝缘蛋白,所以边界不明显)。TAD 边界富集了大量的 CTCF 和黏连蛋白复合物 (Cohesin complex),二者对 TADs 的定位和结构的稳定性发挥重要作用。CTCF 绑定在染色质上,促进 TADs 之间空间隔离的形成,并决定了黏连蛋白的定位,由黏连蛋白指导 TADs 内部的染色质互作,从而维持了 TAD 边界的稳定<sup>[39]</sup>。2018 年,上海交通大学生物医学工程学院王宽诚、邵志峰教授和特别研究员 Czajkowsky 博士带领的团队<sup>[40]</sup>,发现 CTCF 及黏连蛋白定义 TAD 边界这一特征在果蝇中似乎并不适用。通过超高深度 Hi-C 测序重新解析果蝇基因组的 TADs 结构,发现整个基因组全部被 TADs 覆盖,包括相对开放的活性染色质区域。同时在果蝇的染色质中绝大多数的 TAD 边界不是由与人同源的 CTCF/黏连蛋白定义,而是由果蝇特异性的绝缘子蛋白复合物 BEAF-32/CP190 或 BEAF-32/Chromator 所定义。BEAF-32 是果蝇的绝缘子蛋白之一,CP190 和 Chromator 恰好可与 BEAF-32 结合并介导远距离相互作用,功能与哺乳动物细胞的黏连蛋白类似。这表明在果蝇的染

色质中存在着与哺乳动物细胞中 CTCF/黏连蛋白功能类似但非同源的蛋白,它们取代了 CTCF/黏连蛋白的功能,发挥定义 TAD 边界的作用。除了 CTCF 和黏连蛋白, TAD 边界还存在着大量的组蛋白修饰、甲基化修饰位点(如 H3K4me3、H3K36me3 等)、转录起始位点 (Transcription start site, TSS)、看家基因、tRNA 等标记因子,这些标记因子通常与转录活性和表观遗传学密切相关,影响着 TADs 的形成。TAD 边界与复制域边界存在着大量的重合,这说明 TADs 还可能受复制时间调节<sup>[41]</sup>。此外,还有研究表明 DNA 超螺旋结构对 TADs 的建立也存在着重要影响<sup>[42]</sup>。

因为 TADs 参与调控 DNA 复制、转录和表观遗传修饰,因此, TAD 边界的破坏会对基因表达产生较大规模的影响,甚至导致疾病的发生。通过对多个样本 TADs 进行对比,在全基因组水平找出发生变化的 TADs 数量及 TAD 边界的变化,借助染色质免疫沉淀 (Chromatin immunoprecipitation, ChIP) 测序、全基因组 DNA 甲基化测序 (Whole genome bisulfite sequencing, WGBS) 等技术分析 TADs 的变化是否与表观遗传修饰相关,结合 RNA-seq 技术对相关基因的表达量进行统计,有助于解释不同样本间空间结构的差异与表观遗传修饰及转录调控之间的关系,对三维基因组与疾病发生关系的研究有着重要意义。

#### 1.4 染色质环

随着染色质互作谱图精度的不断提升,生物学家们观察到了比 TADs 更为精细的结构单元。Rao 等<sup>[25]</sup>在 1 kb 的分辨率下,发现了目前直接调控基因表达最精细的结构和功能单元:一种简单染色质纤维折叠形成的环状结构——染色质环 (CL, 图 1)。该研究首次列出了上万个人类基因组上的 CL。这些 CL 通常连接着基因的启动子和增强子,与基因激活相关,具有跨细胞类型和跨物种的保守性;研究还发现 65% CL 的出现伴随着结构域的出现,且有 39% CL 的位置位于结构

域的边界。该研究首次报道染色质相互作用下这种现象的发生,并将此结构域定义为“环状结构域”。CL之间是不重合的,86%以上CL中都存在着CTCF以及黏连蛋白的亚基RAD21和SMC3,说明与TAD边界的形成类似,CTCF与黏连蛋白同样参与了CL的形成。Sanborn等<sup>[43]</sup>发现每个CL是由两对CTCF/黏连蛋白介导形成的,每对CTCF/黏连蛋白为一个DNA结合亚基,首先两个亚基中的黏连蛋白合到染色质上,并以反向对称的方式沿DNA序列滑动,对染色质进行挤压形成CL,这种挤压直到在链上检测到与之对应的CTCF才会停止,Eric Lander把这种现象命名为“环挤压模型”(Loop extrusion model)。随后Fudenberg等<sup>[44]</sup>的研究再次证明了上述观点,并提出在挤压过程中可能会释放WAPL和PDS5。

CL的形成与启动子、增强子、CTCF结合位点以及长距离互作密切相关。除启动子与增强子的相互作用可以形成CL外,科学家们还在小鼠细胞、人类淋巴母细胞和宫颈癌细胞中,发现启动子与启动子的相互作用以及增强子与增强子间的相互作用也可以形成CL,以上发现充分解释了

为什么在染色质的某些基因位点上,存在着高频率的远距离相互作用<sup>[12,45]</sup>。由于增强子总是通过CL长距离地控制非临近基因,所以对增强子靶基因的定义也是探究CL影响基因表达机制的研究重点。

## 2 三维基因组研究技术

2002年,Dekker等<sup>[5]</sup>在进行酵母细胞染色质研究时,提出了经典的染色质构象捕获技术,即3C技术(表1)。该技术可以捕获两个线性距离较远的基因位点之间的染色质相互作用,并分析染色质的物理特性。该技术的提出标志着三维基因组的研究进入到了一个“新时代”。因该技术存在一定的局限性,为了满足科学探索过程中对于更高的分辨率与通量的需求,3C技术的衍生技术以及其他测序技术在近十几年飞速发展,研究方式从研究两个单一位点间的相互作用,逐步发展到研究多位点与单一位点相互作用,再到研究多位点与多位点间的相互作用;检测范围也从区域靶向捕获逐渐发展到全基因组。3C及衍生技术的快速发展,极大地推动了其在不同领域的应用。

表1 染色体构象捕获技术(3C)与其他3C衍生技术概述

Table 1 Overview of chromosome conformation capture technologies (3C) and 3C-based technologies

Assay	Full name	Features (interactions between DNA sequences)	Application
3C	Chromosome conformation capture	1 versus 1	3C is the basis of all chromatin conformation capture techniques, which recognize known DNA interactions in a small number of cells and can only verify interactions between two specific genes
4C	Circular chromatin conformation capture technology	1 versus many	Identify known interactions between a specific fragment and the interaction of multiple genes and detect them in the form of a ring. Just need design a pair of primers
5C	Chromatin conformation capture carbon copy technology	Many versus many	Identifying genes that interact with multiple known multiple fragments by adding the same sequence to both multiplex PCR primers, a “tag” was added, high throughput, but there is a random positive connection caused by false positive
Hi-C	High-throughput chromosome conformation capture	All versus all	Capturing genome-wide interactions within the nucleus by using high-throughput sequencing. There is no specific primers, random connections and background noise in the whole process
CHIA-PET	Chromatin interaction analysis using paired end protein specific tag sequencing	Many versus many+	Antibodies are used to capture genome-wide interactions of target proteins. It has higher resolution than Hi-c, and can reconstruct the 3D structure of the genome together with Hi-c

## 2.1 染色质构象捕获技术 (3C)

3C 技术用于样本分析的第一步是构建 DNA 的三维结构模板, 首先使用甲醛固定样本, 使原位交联发生; 第二步用限制性内切酶 (如 *Hind* III、*Bgl* II、*Sac* I、*Bam*H I、*Eco*R I 或 *Aci* I、*Dpn* II) 进行酶切; 第三步将空间位置接近的 DNA 末端进行连接; 最后使用两个目标基因组位点的引物进行 PCR 来识别相互作用。根据不同实验目的, 在扩增阶段可用半定量或定量 PCR 扩增具有相互作用的片段, Würtele 等<sup>[46]</sup>于 2006 年基于开放式 3C 技术对小鼠 ES 细胞中 *HoxB1* 相关基因座进行全基因组扫描, 并将小鼠胚胎干细胞中诱导 *HoxB1* 基因表达前后空间环境划分为 3 个部分: 距 *HoxB1* 位点约 110 kb 的近距离覆盖区域为第一部分, 从该区域延伸至 800 kb 的区域为第二部分, 第三部分由远端染色体内和染色体间的相互作用组成。经典的 3C 技术用于研究两个特定位点之间的相互作用, 以判断特异性扩增片段的有无, 因此仅能“一对一”地研究目标基因组中某一特定的 DNA 序列在染色质中与另一特定的 DNA 序列的相互作用, 在实际应用中存在一定的局限性。

## 2.2 染色质构象捕获技术的衍生技术

### 2.2.1 环状染色质构象捕获技术 (4C)

2006 年, imonis 等<sup>[6]</sup>将微阵列与 3C 技术相结合, 发明了环状染色质构象捕获技术, 即 4C 技术 (表 1)。4C 技术与 3C 技术相比省略了 PCR 步骤, 直接用限制酶切割染色质的目标 DNA 并环化, 然后利用反向 PCR 技术对目标染色质位点接触的所有 DNA 序列进行扩增, 最后利用微阵列分析或下一代测序技术 (Next generation sequencing, NGS)<sup>[47]</sup>分析某一特定 DNA 序列与其他接触位点 DNA 序列的相互作用。该技术可以在全基因组范围内检测与靶向基因座接触的 DNA 基因座, 许多线性距离较远的染色质间的相互作用也可通过形成染色质接触而被检测出来。

### 2.2.2 染色质构象捕获碳拷贝技术 (5C)

为了能够并行捕获多位点对多位点的相互作用, 一种名为碳拷贝染色质构象捕获技术的高通量 3C 测序技术被开发出来, 即 5C 技术<sup>[7]</sup> (表 1)。该技术通过常规 3C 技术产生 3C 文库后, 通过 LMA (Ligation-mediated amplification) 进行扩增, 在多重 PCR 反应中与单链寡核苷酸探针相连接, 进一步创建形成 5C 文库, 最后使用多路复用引物和 NGS, 同时测定多个 DNA 序列间的相互作用<sup>[48]</sup>。该技术既可以在微阵列上进行, 也可在高通量测序中开展, 还可建立目标基因组区域的接触频率矩阵。法国的 Moisan 等<sup>[49]</sup>利用 5C 技术绘制了常染色体隐性遗传性耳聋致病基因 (*DFNB1*) 位点的染色质组织, 对比了与健康个体的正常细胞染色质结构差异, 寻找并量化与耳聋相关基因 (*GJB2*) 的启动子相互作用的染色质区域, 发现了一种新型的 *GJB2* 顺式作用元件, 为该疾病发生机制的探究提供了新思路。

## 2.3 高通量染色质构象捕获技术 (Hi-C)

为了实现高通量的染色质相互作用分析, Dekker 等<sup>[8]</sup>发明了高通量染色质构象捕获技术, 即 Hi-C 技术 (表 1)。该技术是 3C 技术的一个高通量版本, 能够检测所有目标基因组位点的所有相互作用。它的研究对象为整个细胞, 通过捕获细胞内染色质全部 DNA 的相互作用模式, 从而研究整个染色质 DNA 在空间位置上的关系, 还可以进一步得到分辨率较高的染色质三维结构信息。Hi-C 技术改变了 3C 技术中创立模板的过程, 在连接目标 DNA 末端之前, 用生物素标记的脱氧核苷酸填充 DNA 的限制性末端。在 DNA 末端连接之后, 将其纯化并剪切, 然后富集被亲和素标记的连接头部分进行分析。最后得到整个基因组 DNA 片段之间的相互作用频率矩阵, 其分辨率与限制性位点的密度以及测序深度相关。虽然 Hi-C 技术能够捕获全基因组范围内的染色质相互作用, 但因其没有特异性, 所以如仅针对特定的

基因组位点和基因座进行研究则过于昂贵,一般该技术都会与其他测序技术联合应用来达到研究特定相互作用的目的。

Hi-C 技术在疾病风险预防、辅助动植物基因组组装<sup>[32,50]</sup>、基因组分子网络搭建、潜在药物靶点开发等领域得到了广泛的应用,包括解析了类风湿性关节炎、I 型糖尿病和 Crohn 氏病<sup>[51]</sup>、肢体畸形<sup>[52]</sup>、前列腺癌<sup>[53]</sup>、乳腺癌<sup>[54]</sup>等多种疾病的三维基因组变化与疾病发生的关系。

#### 2.4 配对末端标签测序分析染色质相互作用 (ChIA-PET)

为了研究转录因子与转录调控的结合,研究者们开发了 ChIP<sup>[55]</sup>芯片分析和 ChIP 测序技术 (ChIP-seq)<sup>[56]</sup>,以及全基因组双末端标记测序系统 (Paired-end-tag, PET)<sup>[11]</sup>。随后 Wei 等<sup>[57]</sup>结合 ChIP 和 PET 两种技术的优势,开发出了 ChIP-PET 用于基因组层面蛋白结合位点的研究 (表 1)。他们分析了近 50 万个 PET 序列的样品,获得了 65 572 个单独的 ChIP 实验 DNA 片段,并建立了重叠 PET 簇,用于特异性标记 *p53* 结合位点,但以上技术无法检测到远端转录因子结合的靶基因。

为了进一步研究远端的基因组区域在调控生命进程中发挥的作用,2009 年新加坡基因组研究院 (Genome Institute of Singapore, GIS) 的研究人员开发出了一种整合了染色质邻近式连接 (Chromatin proximity ligation)、ChIP、PET 以及 NGS 等技术,用来研究基因组范围内染色质远程交互的新兴技术——“配对末端标签测序分析染色质相互作用”,即 ChIA-PET 技术。该技术首先用超声波将甲醛交联的 DNA-蛋白质复合物片段化,然后利用 ChIP 富集所需的 DNA-蛋白复合物片段,在片段末端加上包含 *Mme* I 位点的生物素化寡核苷酸 linker,再经过连接、消化、固定化等 PET 序列步骤,最后,利用 NGS 技术对目标 DNA 进行测序。ChIA-PET 是一种无偏、全基因组、高通量的测序方法,在功能研究中,与 Hi-C 相比,

ChIA-PET 在更高的分辨率下与相关蛋白质关联性更好,且因使用超声打断 DNA-蛋白质复合物,测序产生的噪音也比较小。

#### 2.5 染色质开放性测序技术 (ATAC-seq)

在 2013 年,美国斯坦福大学 Greenleaf 教授研发了一种可利用 DNA 转座酶结合高通量测序的技术<sup>[58]</sup>,这种研究染色质可接近性的全新方法,被称为染色质开放性测序技术 (Assay for transposase accessible chromatin with high-throughput sequencing),即 ATAC-seq。该技术是一种创新的表观遗传学研究技术,能够测得开放区域染色质的 DNA 序列。该技术的原理是利用转座酶 Tn5 获取开放性染色质,然后对 Tn5 酶捕获到的 DNA 序列标签进行 PCR,再通过高通量测序及生物信息学分析来挖掘相关基因信息。传统的开放染色质方法有 DNase-seq<sup>[59]</sup>及 FAIRE-seq<sup>[60]</sup>,但因其存在需要细胞量大、检测背景高、测序信噪比低、数据解读困难等问题,所以限制了上述技术的应用与发展。与传统方法相比,ATAC-seq 所需细胞量少,操作简单,测序信号更好,可以在全基因组范围内检测染色质的开放状态。目前 ATAC-seq 已经成为研究染色质开放性的首选技术方法。Kelso 等<sup>[61]</sup>利用 ATAC-seq 技术,证明 *ARID1A* 及 *ARID1B* 可通过调控核小体重塑,发挥抑癌作用。

### 3 三维基因组学的应用领域

3C 技术及其衍生技术 (4C、5C、Hi-C)、Hi-C 衍生技术 (原位 Hi-C)、HiChIP<sup>[62]</sup>、BL-Hi-C (Bridge Linker-Hi-C)<sup>[63]</sup>、DLO Hi-C (Digestion-ligation-only Hi-C) 和 ChIA-PET 等相关测序技术的发展,推动了三维基因组学在农业科学、生命科学以及医学等众多领域的应用。通过多组学高通量测序联合分析获得的大量交互数据,可用于基因组三维结构的解析,从而实现染色质空间构象、信号传导通路、转录因子调控机制、基因表达机制、机体生长发育机制等问题的深入研究,

并为进一步探索未知领域的奥秘,奠定了坚实的基础。

### 3.1 农业科学领域

农业科学是研究农业发展自然规律和经济规律的科学。农业生产对象的多样性和生产条件的复杂性,决定了农业科学的范围广泛和门类繁多,农业科学主要包括农业环境科学、作物生产科学、畜牧科学、农业工程科学和农业经济科学 5 个门类。三维基因组在很大程度上影响动植物的复杂性状的形成,通过三维基因组研究技术辅助完成动植物基因组组装,能够解析染色体水平的全基因组序列,从而寻找影响动植物优良性状的关键基因、阐明动植物重要经济性状,了解染色质修饰以及基因表达的调控机制,探究生物进化过程,对农业生物技术的发展具有重要意义。

西北农林科技大学动物医学院及中国农业科学院的科学家们综合全基因组测序 (Whole Genome Sequencing, WGS)、RNA-seq、Hi-C 等技术,通过解析北京鸭的三维基因组空间构象<sup>[35]</sup>,成功定位出与北京鸭体型和羽色相关的主要基因。该研究发现由于一个 6.6 kb 的大片段序列插入到 *MITF* 基因中,导致负责黑色素合成的所有基因被抑制甚至沉默,从而使北京鸭具有白色的羽毛。同时,由于远程增强子的自然突变,导致胰岛素样生长因子 II mRNA 结合蛋白 1 (*IGF2BP1*) 基因在北京鸭体内可以持续表达。*IGF2BP1* 基因在动物胚胎期起到促进生长的作用,该基因的持续表达在很大程度上提升了北京鸭的采食量及饲料利用率,从而使其体格变大。该发现对提升鸭子的经济学效益具有重要意义。

东北师范大学的宫磊、刘宝研究组<sup>[37]</sup>利用 Hi-C 技术对水稻染色质的三维结构进行了全基因组解析,发现水稻中的 TADs 结构均匀分布在水稻的 12 条染色质中, TAD 边界强度在 A/B Compartment 中无明显差异, Compartment 的边界也是划分 TADs 的边界, TAD 边界的基因表达水平

高于 TADs 内部的基因表达水平。进一步解析发现,水稻的近距离交互热点 (Frequently interacting regions, FIREs) 多位于 A Compartment 中,该区域富集了大量活跃表观修饰基因。该研究为探索水稻和其他谷类作物的分子机制提供强有力的数据信息和理论参考。

### 3.2 生命科学领域

生命科学是研究生物 (包括微生物、动物、植物等) 的结构、功能、发生和发展规律的科学。高等生物的 DNA 复制、重组、基因转录主要在细胞核的染色质三维空间中完成,细胞核内染色质会被折叠成复杂的三维结构,并在生命活动中发挥重要的动态调控作用。

#### 3.2.1 胚胎发育的三维基因组学研究

胚胎发育通常是指从受精卵起到胚胎出离卵膜的发育过程。在配子、合子和早期胚胎发育过程中,基因组的三维空间构象及其功能的解析是一个亟待解决的生物学问题。在胚胎发育早期,表观遗传重编程会影响基因调控的精确性<sup>[64]</sup>。探究重编程过程中三维基因组的变化,以及表观遗传学信息在早期胚胎发育中遵循什么样的遗传机制和规律,是解决上述问题的关键。

Xue 等<sup>[65]</sup>通过研究果蝇早期胚胎的染色质构象,发现在合子基因组激活 (Zygotic genome activation, ZGA) 之前,果蝇的早期胚胎中几乎没有 TADs 结构,但是在 ZGA 之后,细胞会重新建立 TADs 结构。还有研究人员利用原位 Hi-C 技术,分析斑马鱼发育过程中染色质三维结构变化<sup>[33]</sup>,发现在斑马鱼的 TADs 和 Compartment 中具有与哺乳动物相似的基因组特征,如 A Compartment 富含 H3K4me3、H3K4me1 和 H3K27ac 且具有更高的基因密度; TAD 边界具有保守性,该研究还证明环挤压模型也可能是斑马鱼 TADs 形成的原因。斑马鱼的 TADs 在 ZGA 之前开始出现,但是在 ZGA 后明显减少,这点与果蝇 TADs 的变化不一致,说明斑马鱼 TADs 的形成不依赖于转录起始。

中国科学院北京基因研究所刘江和清华大学生命科学学院颀伟研究组深入研究了哺乳动物胚胎发育过程中染色质三维结构的变化<sup>[66]</sup>。他们开发出了一种可适用于极少量细胞的 Hi-C 技术 (Small scale *in situ* Hi-C, sisHi-C), 发现在小鼠早期胚胎发育过程中, 精子保留了经典染色质高级结构 TADs 和 Compartment 结构, 而处于中期 II 期的成熟卵子则缺乏 TADs 结构。在精子和卵子结合为合子后, 其染色质呈现为一种松散的状态, 染色质三维结构在 2 细胞时期开始逐渐建立, 到内细胞团 (Inner cell mass, ICM) 时期基本建成和胚胎干细胞相似的三维结构, 这个过程非常缓慢, 主要表现为 TADs 结构的逐渐稳定以及 A/B Compartment 的逐渐隔离。此外, 胚胎中 A Compartment 富集了比 B Compartment 更多的未甲基化的 CpGs, A Compartment 甲基化水平的降低程度比 B Compartment 高, 这也表明 A Compartment 具有更加松散开放的结构状态<sup>[67]</sup>。

### 3.2.2 细胞衰老的三维基因组学研究

细胞衰老是在多种作用下 (如 DNA 双链断裂或癌基因激活等) 一种不可逆的细胞变化, 在形态学上表现为细胞结构的退行性变化, 在生理学上的表现为功能衰退与代谢低下<sup>[68]</sup>。细胞衰老过程中, 也伴有染色质三维结构的变化, 细胞核中会形成衰老相关异染色质聚集 (Senescence-associated hetero-chromatin foci, SAHF), 染色质重塑是调控基因时序性表达的重要环节<sup>[69]</sup>。Chandra 等<sup>[70]</sup>利用 Hi-C 技术对衰老细胞和胚胎干细胞 (Embryonic stem cell, ESCs) 的染色质空间构象进行探究, 发现与 ESCs 相比, 衰老细胞的 TADs 局部相互作用丧失, 从而导致其物理紧实度发生变化。虽然衰老细胞的 TAD 边界位置在很大程度上保持不变, 但他们的质量受染色质局部和远端相互作用影响, 有很大一部分 TADs 的绝缘强度降低。另一项研究也发现, 当衰老发生时, 染色质重塑是由于 CTCF 簇的形成, 导致 CL 的重组, HMGB2

蛋白在此过程中发挥着重要作用<sup>[71]</sup>。

### 3.2.3 多能干细胞的三维基因组学研究

多能干细胞 (Pluripotent stem cell, PSC) 是当前干细胞研究的热点, PSC 可以分化成体内所有的细胞, 进而形成身体的组织和器官, PSC 在器官再生、修复和疾病治疗方面极具应用价值。基因组结构在细胞分化和体细胞重编程过程中会发生细胞特异性重配<sup>[72]</sup>。路德维希癌症研究所任兵研究组和美国加州大学 Neil Chi 研究组在合作研究心肌细胞分化过程中染色质结构变化时<sup>[73]</sup>, 意外地发现在 PSC 中存在特异性 TADs, 并证明这些特异性 TADs 的形成与一种在人体内高度表达的内源性逆转录病毒 (HERV-H) 有关, 而这些特异性 TADs 仅在类人猿中出现, 说明 HERV-H 在灵长类进化过程中引入了新的染色质结构, 并可能与基因组进化出新的基因调控逻辑息息相关。

## 3.3 医学领域

疾病严重威胁着人类的健康, 虽然人类在医学领域的研究发展迅速, 但很多新发传染病 (如寨卡病毒病、登革热、新型冠状病毒肺炎等), 以及对人类健康造成重大威胁的疾病 (如染色体疾病、神经精神疾病、癌症等) 的研究都还处于初级阶段。越来越多证据表明, 疾病的发生往往伴随着基因组三维空间构象与功能的改变, 三维基因组学为科学家对癌症发生机制的探究提供新的方向。利用三维基因组学, 研究疾病发生前后染色质三维结构如 CT、A/B Compartment、TADs、CL 的变化, 以及不同调控元件与目标基因相互作用的改变, 可以帮助分析疾病的发生及遗传机制, 挖掘影响疾病发生的关键基因, 寻找潜在的特异性标志物, 筛选治疗靶点, 从而为疾病的治疗提供新的思路及解决方案。

### 3.3.1 染色体疾病的三维基因组学研究

染色体疾病通常指染色体数目或结构异常引起的疾病, 其中唐氏综合征 (21-三体综合征) 及猫叫综合征 (5q-综合征) 是最为大众所熟知的两

类染色体异常疾病。这类疾病的实质是染色体上的基因、基因群的增减或变位影响了众多基因的表达和作用,破坏了基因的平衡状态,因而妨碍了人体相关器官的分化发育,造成机体形态和功能的异常。如 21-三体综合征是因为多了一条 21 号染色体,此类患者的主要表现为智力缺陷、生长发育障碍、多发畸形等,还会诱发先天性心脏病、胃肠道功能缺陷和急性淋巴白血病等。Kemeny 等<sup>[74]</sup>通过检测 21-三体综合征患者三体细胞的三维基因组结构,发现三体细胞核的染色体内 CT 的径向定位发生改变,但染色体的大小、基因密度及染色体的径向排列并未发生太大变化。三体细胞多出来的 HSA21 的 CT 会导致 HSA1 和 HSA3 的 CT 向核空间外围转移,使 HSA1 和 HSA17 的 CT 具有更高浓缩程度及功能改变。三体细胞的染色体或基因在细胞核内的重新定位,可能会改变共调控基因或独立调控基因的状态,并形成新的顺式和反式相互作用。

### 3.3.2 神经精神疾病的三维基因组学研究

神经精神疾病是由于神经系统障碍导致大脑功能性活动紊乱,从而造成人的认知、情感、行为和意志等活动异常。大量与人类疾病相关的基因位于特定的调控区域,这些区域的调控元件可以通过调节基因表达水平,改变或破坏转录网络从而导致复杂疾病的发生,其中许多复杂的神经精神疾病的发生就与调控元件的变化有关。三维基因组的改变是导致调控元件变化的关键因素, Song 等<sup>[75]</sup>利用 Hi-C、ATAC-seq 和 RNA-seq 测序技术,综合分析不同的神经细胞类型:诱导多能干细胞 (Induced pluripotent stem cells, iPSC) 诱导的兴奋性神经元和下运动神经元, iPSC 衍生的海马齿状回状神经元和原代星形胶质细胞,发现了数十万个启动子以及远端启动子的相互作用区域 (Promoter-interacting regions PIRs) 的远程相互作用。利用 CRISPR 技术验证了几个关键启动子作用区域,证明 *CDK5RAP3*、*STRAP* 和 *DRD2*

在转录上受到物理连接的增强子调控,揭示了在疾病发生过程中调控元件与目标基因相互作用的改变。

### 3.3.3 癌症的三维基因组学研究

癌症是严重威胁人类健康的重大疾病之一, 2018 年美国癌症学会官方期刊 *A Cancer Journal for Clinicians* 发表的数据显示, 2018 年约有 1 810 万癌症新发病例和 960 万死亡病例, 其中肺癌 (11.6%)、女性乳腺癌 (11.6%)、结直肠癌 (10.2%)、前列腺癌 (7.1%) 在所有癌症类型中发病率最高; 肺癌 (18.4%)、结直肠癌 (9.2%)、胃癌 (8.2%)、肝癌 (8.2%) 的死亡率最高<sup>[76]</sup>。国家癌症中心的统计报告显示, 全国恶性肿瘤新发病例数 380.4 万例, 我国高发的恶性肿瘤为肺癌、胃癌、结直肠癌、肝癌和女性乳腺癌; 死亡率排名前五的癌症类型分别是肺癌和支气管癌、胃癌、肝癌、食道癌和结肠直肠癌, 约占所有癌症死亡的四分之三<sup>[77]</sup>。

目前科学家们对各类癌症的研究仍处于初级阶段, 对其发病机制了解甚少, 随着 3C 及其衍生技术和 NGS 的发展, 科学家们可以通过更高的通量、更高的效率分析癌细胞的染色质三维结构变化, 进一步探究癌症发生机制。Chen 等<sup>[78]</sup>探究了 33 种类型癌症 8 928 个样本, 发现绝大多数癌症中有增强子被激活的现象。进一步检测增强子与癌基因之间的相互作用, 发现其相互作用是通过特定的 TADs 形成的, 原本线性距离非常远的调控因子可以通过染色质折叠成三维结构, 达到空间上的近距离接触, 从而导致癌症发生。目前, 三维基因组在癌症细胞特异性研究上取得了一定的研究成果, 包括对于肺癌、乳腺癌、前列腺癌、结直肠癌、胃癌、肝癌、骨髓瘤等恶性肿瘤的发病肿瘤发生机制的探究。

#### (1) 肺癌

肺癌是世界上发病率和死亡率最高的癌症。肺癌根据病理形态可以分为小细胞肺癌 (Small cell

lung cancer, SCLC) 和非小细胞肺癌 (Non-small cell lung cancer, NSCLC)<sup>[79]</sup>。目前 SCLC 的治疗手段主要是化学治疗; NSCLC 患者根据不同分期, 分别采取手术切除 (I 期和 II 期)、放化疗 (III 期) 或者免疫治疗 (IV 期)<sup>[80]</sup>。目前多数用于治疗晚期 NSCLC 的药物只能适用于特定基因 (如 *EGFR*、*HER2*、*ALK* 等) 突变的肺癌患者, 并容易产生耐药性。因此, 三维基因组学探究肺癌发病机制, 寻找新型的药物靶点对于肺癌尤其是晚期肺癌患者的治疗有着重要意义。

四川大学华西医院李为民教授研究组和谢丹教授研究组利用 ATAC-seq、全基因组测序、转录组测序多组学测序技术, 对 50 例原发性 NSCLC 患者包括 34 例肺腺癌 (Lung adenocarcinoma, LUAD) 患者、13 例肺鳞癌 (Lung squamous carcinoma, LUSC) 患者和 4 例良性孤立肺结节 (Benign solitary pulmonary nodules, BSPN) 患者的肿瘤样本进行测序分析, 揭示了肺癌复杂的基因调控网络<sup>[81]</sup>。样本聚类分析结果显示, 不同病理类型的肺癌染色质开放区域特征各不相同, 如在肺鳞癌的染色质开放区内 *LCE1E*、*LCE3D*、*SPRR3* 等上皮细胞角质化过程相关基因表达上调, 在早期肺腺癌的染色质开放区内 *ITGAV*、*ITGA6* 等与预后相关的基因特异性表达。除此之外, 研究还发现了一些广泛开放区域 (长度在 1 kb 到 180 kb 之间), 在此区域内 *EGFR*、*JUN*、*ERBB3*、*WNT9A* 等肺癌驱动基因的表达量都发生了大范围的波动, 这些广泛开放区域的 ATAC-seq 峰值有望作为识别 NSCLC 关键基因的标志物。通过多组学分析, 发现在 99.7% 的样本中, 拷贝数变异 (Copy number variations, CNV) 片段携带基因表达相关的调控元件; 同时鉴定出了包括 *GSTM1*、*RPTN*、*GSTT1*、*ROS* 在内的 21 个与基因组开放区域及转录调控有关的数量性状位点 (Quantitative trait locus, QTLs)。初步推测这些位于非编码区域的调控位点, 可以作为肺癌诊断以

及药物设计的靶标, 指导肺癌的精准化治疗。

## (2) 乳腺癌

乳腺癌发生在乳腺上皮组织, 是全球女性中最常见的癌症<sup>[82]</sup>。我国乳腺癌发病率的增长速度远高于世界平均水平<sup>[77]</sup>。早期 (I 期或 II 期) 乳腺癌患者最常采用的治疗方式是保乳手术配合辅助性放疗 (49%), III 期患者则多采用乳房切除治疗 (68%)<sup>[80]</sup>。乳腺癌的早期检测与更好的预后效果密切相关, 因此阐明乳腺癌的发生机制以及筛选乳腺癌相关标志物, 对乳腺癌的精准治疗至关重要。

多项研究证实染色质与基因定位对乳腺癌的发生有着重要影响<sup>[83-84]</sup>。Barutcu 等<sup>[85]</sup>利用正常乳腺上皮细胞系 (MCF-10A) 和乳腺癌细胞系 (MCF-7) 探究正常细胞与癌细胞之间三维基因组结构差异。研究发现, MCF-7 癌细胞在 16 号至 22 号染色体区域内的相互作用频率低于 MCF-10A 上皮细胞。进一步鉴定 A/B Compartment 之间的转化, 发现在 16 号至 22 号染色体区域, MCF-10A 中部分 B Compartment 在 MCF-7 中转化为了 A Compartment (以下称为“区域 1”), 表明与 MCF-10A 相比 MCF-7 的染色体在该区域内更具开放性。通过差异表达分析, 研究确定了 2 437 个 MCF-7 上调基因和 2 427 个 MCF-7 下调基因 ( $\text{Log}_2$  fold change > 1,  $P < 0.01$ ), 发现 MCF-7 上调基因在 A Compartment 富集; 深入分析“区域 1”中富集的基因, 发现这些基因均与已知的致癌途径 (WNT 信号通路等) 相关。在 MCF-7 中有 85% 左右的 TAD 边界与 MCF-10A 的 TAD 边界一致, 但是在 MCF-7 中有部分 TADs 被“分解”为更小的几个亚 TADs, 同时在 MCF-7 的 TAD 边界处富集了多种癌症相关蛋白 (如 GABP、ELF1、PML、C-MYC、MAX 等)。Seth Frietze 与 Victor Jin<sup>[86]</sup> 研究组合作利用雌二醇 (Estradiol, E2) 刺激雌激素受体  $\alpha$  (Estrogen receptor, ER) 阳性乳腺癌细胞, 发现在 E2 刺激 1 h 后, 细胞的高级染色质结

构发生了重新分区 (A/B Compartment 转化等), 继续延长 E2 的刺激时间, 发现随着时间的推移, 活性染色质区域更容易发生变化。该研究确定了与染色质开放区域相关的具有时间相关性的高动态区室子集, 并发现在这些高动态区室中, 具有他莫昔芬耐药的乳腺癌细胞表现出更高的变化, 同时还增强了与 ER 的结合, 降低了与 CTCF 的结合。进一步扩展了对 E2 作用于 ER 阳性乳腺癌的分子机制的理解。

### (3) 前列腺癌

前列腺癌 (Prostate cancer, PCa) 作为男性最为常见的癌症, 其发病率在全球排第三, 在男性中致死率仅次于肺癌排名第二<sup>[76]</sup>。在我国, 前列腺癌的发病率和死亡率逐年上升<sup>[77]</sup>。PCa 的发生与家族病史、年龄、种族以及各类环境因素有关, PCa 患者的前列腺细胞转录组的异常变化, 导致细胞增殖失控<sup>[87]</sup>, 探究染色质三维结构变化对基因转录的影响, 对前列腺癌发生机制的解读具有重要意义。

Taberlay 等<sup>[53]</sup>通过整合多组学数据, 对正常前列腺上皮细胞 (PrEC) 和两种前列腺癌细胞系 (PC3 和 LNCaP) 的染色质三维结构进行差异分析, 发现癌细胞中的 TADs 变小且数量增多。TAD 边界是维持 TADs 结构稳定划分 TADs 区域的重要结构, 研究还发现在前列腺癌细胞特有的 TAD 边界处富集了大量的 CTCF 以及启动子相关的组蛋白修饰 (H3K4me3), 但是与增强子相关的 H3K4me1 和 H3K27ac 未在边界处富集。此外, 约 70% 的 CNV 是 LNCaP 和 PC3 共有的, 并且位于癌细胞特有的 TAD 边界处, 如在 17p13.1 处癌细胞特有的基因删除与新的 TAD 边界的形成相关, 这个位置存在肿瘤抑制基因 *p53*, 并且几乎在所有的前列腺癌细胞中都发生了缺失, 证明 CNV 与癌细胞特有的 TAD 边界形成有关。进一步分析癌细胞与正常细胞的差异相互作用, 发现癌症特异性染色质互作与基因的表达高度相关,

如癌细胞中 *KCNMA1* 基因座的接触频率增加 (PC3 与 PrEC 倍数变化=7.11) 导致 *KCNMA1* 表达增加 10 倍以上 ( $\log_2$  fold change=3.38), 表明染色质相互作用的改变会导致基因表达的失调, 并促进癌细胞中癌基因的表达增加。Rhie 等<sup>[88]</sup>继续探究了染色质三维结构对前列腺癌转录调控的影响。与 Taberlay 的发现类似, 该研究在正常前列腺细胞 (RWPE1) 中鉴定出约 520 个大 TADs, 在前列腺癌细胞 (C42B 和 22Rv1) 中鉴定出约 850 个小 TADs。部分基因的表达量在这些“变小”的 TADs 区域内升高, 推测变小的 TADs 将基因与抑制基因表达的调控元件隔离开, 从而使癌细胞具有更高的转录活性。前列腺癌细胞里的特异性增强子-启动子环调控的基因在癌细胞中高表达, 根据此特征, 研究发现了一个癌症特异性转录因子蛋白 FOXA1, C42B 增强子-启动子环中 31% 的癌症特异性增强子与 FOXA1 结合, 是癌症基因转录的主要调节剂。FOXA1 可与调控细胞周期的蛋白 (如 CDK4、CDC23、MYC) 以及雄激素受体 (AR)、雄激素响应基因 (如 *GRHL2*) 相结合。癌细胞特异性增强子-启动子环可通过 FOXA1 上调与其结合的基因的表达。PCa 的 AR 表达具有异质性, 是因为在 AR 基因座附近存在癌症特异性小 TADs, 使得增强子与启动子的相互作用更频繁, 上调了 AR 的表达。此前 Rhie 等<sup>[89]</sup>还成功预测了一组 PCa 相关的非编码区 SNP, 并发现癌症特异性增强子-启动子环可能影响 *KCNM3* 及 *KRT78* 基因表达, 从而导致前列腺癌的发生。

### (4) 结直肠癌

结直肠癌是全球第四大致命癌症, 每年约有 90 万人死于结直肠癌。随着人们生活水平的提高和生活方式的改变, 结肠癌的发病率正逐年上升<sup>[90]</sup>。结直肠癌的分子特征主要包括: ①基因组不稳定, 表现为染色体数目和结构的改变; ②表观遗传学异常; ③基因表达紊乱: 如包含 CpG 岛的启动子导致的过度甲基化, 从而引起抑癌基因沉

默；以及基因组微卫星标记不稳定，导致错配修复 (Mismatch repair, MMR) 功能缺失<sup>[91]</sup>。

Jäger 等<sup>[92]</sup>利用 Capture Hi-C 方法，捕获结直肠癌细胞系 LS174T、LoVo、Colo205 中增强子-启动子间的相互作用，同时结合 RNA-seq 检测数据分析了 14 个结直肠癌风险位点的相互作用，鉴定了 SNP 位点与重要的远距离作用元件间的顺式和反式相互作用。研究在染色体 8q24 上鉴定了 rs6983267 和 MYC 之间的相互作用，并将 CCAT1 确定为上游相互作用因子；在染色体 11q23 上发现与 AB231705 编码区域有交互作用的 *C11orf53*、*C11orf92* 和 *C11orf9* 基因；还证明 MECOM 与 TERC 的强启动子之间的相互作用共同调控癌基因的表达。Seaman<sup>[93]</sup>通过整合 Hi-C 和 RNA-seq 数据，分析了在结直肠癌细胞系 HT-29 中，染色质三维结构变化对核结构和基因表达的影响。发现在染色体 8q 上的 HSR 与许多其他基因组区域相互作用，该现象使得 HSR 更容易被转录；研究还发现在 HT-29 中出现了更多的染色体易位，并在髓性白血病细胞系 K562 的染色体易位情况进行验证分析，证明了染色体易位现象是在癌细胞中普遍存在的。北京大学第三医院汤富酬、乔杰、付卫研究组<sup>[94]</sup>合作从多组学水平解析了人类结直肠癌在发生和转移过程中，基因组 CNV、DNA 甲基化异常及基因表达水平变化的特点及相互关系。该研究利用了一种全新的单细胞三重组学测序方法 (Single-cell triple omics sequencing technique, scTrio-seq)，对患有结直肠癌并伴随转移的 12 例患者基因组进行分析，发现受肿瘤微环境动态影响，患者间癌细胞基因组 DNA 甲基化存在个体差异，同时发现结直肠癌细胞间的共性变化，如癌细胞的 DNA 甲基化水平普遍低于癌旁的正常细胞，基因启动子区域的 DNA 甲基化与相应基因表达呈负相关。在肿瘤细胞转移过程中，基因组 DNA 甲基化水平相对稳定，只有局部区域 (4 号、5 号、8 号、13 号、18 号和 X 染色体 6 条染色体)

的 DNA 甲基化水平发生较大改变，这些区域可能就是影响肿瘤发生的重要区域。

#### (5) 胃癌

胃癌是我国临床最常见的消化道恶性肿瘤，除了临床常见腺癌外，还包括腺鳞癌、鳞癌、类癌等类型，根据组织起源不同还可以分为肠型和胃型<sup>[95]</sup>。由于常见胃癌的发病机制及疾病进展尚不十分明确，临床诊断易发生漏诊或误诊等情况，从而错过最佳治疗时间<sup>[96]</sup>。因此明确胃癌发生机制，寻找合适的胃癌标志物，对胃癌的早诊早治极其重要。

基因重排导致的基因融合是癌症发生的重要标志，Yao 等<sup>[97]</sup>分析了 15 种不同类型胃癌患者样本的开放染色质区域的染色质三维结构，确定了 7 个高频基因重排位点和 136 个融合基因，并从中筛选出了与胃癌复发显著相关的癌症特性异融合基因 *CLDN18-ARHGAP26*。研究发现负责编码紧密连接相关蛋白的基因 (*CLDN18*) 和编码 RHOA 抑制剂的基因 (*ARHGAP26*) 在发生融合后，会破坏胃粘膜上皮细胞的屏障作用，促使胃腔的 H<sup>+</sup> 侵入胃粘膜，从而阻止伤口愈合，促进早期癌症的发展和晚期癌症的进展。四川大学华西医院胡建昆团队等<sup>[98]</sup>进一步研究 *CLDN18-ARHGAP26* 融合基因对细胞耐药性的影响，发现具有 *CLDN18-ARHGAP26* 融合基因的胃印戒细胞癌患者生存率相对较低，并证明以 5'-氟尿嘧啶/奥沙利铂为基础的化疗方式对此类患者无效。多种研究表明该融合基因的发现可作为临床诊断及预后和化疗耐药性的重要指标，为胃癌的临床诊治带来新希望。

#### (6) 肝癌

肝癌是全球最常见的癌症之一，其死亡率在全球癌症中排第三，中国作为肝癌大国，全球每年新发肝癌病患中约有一半以上来自中国。肝细胞癌 (Hepatocellular carcinoma, HCC) 患者占到 90% 左右，除此之外，还包括肝内胆管细胞癌

(Intra-hepatic cholangiocarcinoma, ICC) 及混合型肝癌<sup>[99]</sup>。因肝癌具有高度的异质性, 其发生机制非常复杂, 目前认为乙肝病毒 (Hepatitis B virus, HBV)、丙肝病毒 (Hepatitis C virus, HCV) 感染、长期酗酒、黄曲霉毒素感染导致的肝硬化、非酒精脂肪性肝炎等为肝癌的重要致病因素<sup>[100]</sup>。在中国, 原发性肝癌多与慢性 HBV 感染有关<sup>[101]</sup>, 但被病毒入侵的宿主细胞发生了什么变化, 病毒与宿主细胞间的相互作用尚不明确。

巴斯德研究所 Romain Koszuly 团队与 Christine Neuveut 团队合作利用 HBV 和 5 型腺病毒 (Adenovirus type 5, Ad5) 两种 DNA 病毒感染人原代肝细胞 (Primary human hepatocytes, PHH), 探究病毒侵染前后宿主三维基因组的变化及染色质相互作用情况<sup>[102]</sup>。研究发现被 HBV 侵染前后的 PHH 染色体三维结构基本一致, 而被 Ad5 侵染后 PHH 的 A/B Compartment、基因表达和组蛋白修饰均发生显著变化。HBV 基因组优先与宿主活跃的染色质区域结合, 包括 CpG 岛区域, 进一步研究发现这些区域通常甲基化程度很低, 调控因子 CxxC Finger Protein-1 (CFP1) 可以结合到非甲基化的 CpG 岛上, 从而富集 H3K4me3 修饰使染色质处于活跃状态, 病毒 DNA 的甲基化及 H3K4me3 的富集也影响着病毒的转录。Ad5 则更倾向与转录起始位点及增强子区域发生相互作用。此研究表明, DNA 病毒可以通过不同的途径入侵宿主, 并与宿主基因组的特定区域靶向结合, 同时产生自我复制必需的转录因子, 从而使细胞基因表达失调。

HepG2 作为生物医学研究中使用最为广泛的肝癌细胞系之一, 其三维基因组空间构象特征尚不明确。2019 年, Urban 团队首次报道了肝癌细胞系 HepG2 的整合及单倍型的全基因组特征<sup>[103]</sup>。该研究分析了 20 个处于中期的 HepG2 细胞, 发现所有细胞均含有 49–52 条染色体, 并显示出了多种染色体结构和数目异常, 包括 1 号染色体和

21 号染色体短臂间的易位, 12、16、17 号染色体三体型, 20 号染色体四体型, 同时在 16 号和 17 号染色体上发现了非典型染色体重组和不同数量的标记染色体, 表明不同的 HepG2 细胞系之间也存在核型差异; 在高分辨率 (10 kb) 下测定染色质区域的拷贝数 (Copy numbers, CN), 有 92.9% 的 HepG2 基因组 CN 大于 1, 在多个样本中均发现致癌基因 *VEGFA* (6p21.1) 的 CN 增加 (CN=3); 有 377 个单核苷酸变异 (Single-nucleotide variants, SNV) 和 255 个插入及缺失均为 PPA (Private protein-altering), HepG2 的 PPA 变体还包括致癌基因和肿瘤抑制因子 (*NRAS*、*STK11/LKB1*、*PREX2*)。负责编码 polo 样蛋白激酶的 *PLK2* 基因, 是 *p53* 的转录靶标, 在许多人类癌症中均发现了 *PLK2* 的表达下调<sup>[100]</sup>, 在 HepG2 癌细胞中也发现了类似的现象, 认为是染色体 CpG islands (CGIs) 内 *PLK2* 等位基因的特异性甲基化导致了 *PLK2* 基因的表达下调。

#### (7) 骨髓瘤

多发性骨髓瘤 (Multiple myeloma, MM) 是由浆细胞恶性增殖引起的一种恶性血液疾病。MM 的发生常引起一系列溶骨性改变, 其主要特征是在骨髓中多个位点出现单克隆、长寿命的浆细胞异常增生, 临床表现为骨骼疼痛、高钙血症、病理性骨折等, 因癌症特异性症状不明显, 导致误诊率高、预后效果差<sup>[104]</sup>。

北京大学李程研究组<sup>[105]</sup>通过整合 Hi-C、WGS 和 RNA-seq 的数据来比较 MM 细胞 (RPMI-8226 和 U266) 和正常 B 细胞 (GM12878) 中 CNV 与三维基因组的关系。研究发现与正常细胞相比, MM 细胞中 TADs 的数量增加了 25%, TADs 的平均长度减小, 同时发生 CNV 的断点处与癌症特异性的 TAD 边界重合, 推测 CNV 能够帮助形成新的 TADs 并影响基因表达。研究还发现在癌症基因组上存在大量染色体互作热点, 通过对比 WGS 数据, 发现这些热点中有一部分反映了染

染色体易位的发生,同时该位置包含了许多与骨髓瘤相关的基因,如 RPMI-8226 细胞中的 *WWOX*、*MAF*、*ADORA2B* 等基因和 U266 细胞中的 *TNIK*、*FBXW7*、*TRIM2* 等基因。在探究 MM 细胞中三维基因组变化对基因表达量影响过程中,发现在染色体 2q11.2-q12.1 处存在一个细胞因子受体基因簇,而且该基因簇在 GM12878 细胞中为 A Compartment,但在 RPMI-8226 和 U266 中该区域中一部分转化为 B Compartment,并且与几种白细胞介素 *IL1R1*、*IL1R2*、*IL18R1* 和细胞因子 *MAP4K4* 基因表达下调有关。以上研究表明基因表达水平与表观遗传的改变均与染色质三维结构的变化有关。

#### (8) 胶质瘤

胶质瘤是来源于神经系统胶质细胞和神经元的肿瘤统称,由大脑和脊髓胶质细胞癌变产生的胶质瘤称为脑胶质瘤,是最为常见的颅内恶性肿瘤。通过分析胶质母细胞瘤<sup>[106]</sup>和弥漫性神经胶质瘤<sup>[107]</sup>的综合基因组,研究人员发现在胶质瘤中特异性表达负责编码异柠檬酸脱氢酶的 *IDH* 基因,突变的 *IDH* 能将  $\alpha$ -酮戊二酸转化为 2-羟戊二酸 (2-HG),2-HG 可以抑制前者的靶点,导致这些靶点表达异常引发癌症。

Flavahan 等<sup>[108]</sup>发现在 *IDH* 基因突变的胶质瘤细胞 CTCF 结合位点呈现出高度甲基化,减少了与甲基化敏感性绝缘子蛋白的结合,从而破坏了 TADs 结构,使 TADs 之间的绝缘性丧失,并激活异常基因(胶质瘤癌基因 *PDGFRA*) 的表达,促进胶质瘤的发生。中南大学肿瘤研究所陶永光研究组揭示了染色质重塑蛋白 LSH 促胶质瘤的新机制<sup>[109]</sup>。

天津医科大学于士柱研究组<sup>[110]</sup>发现了新型染色质结构调节因子 *SNDA* 促进胶质瘤增殖和侵袭的新机制。金黄色葡萄球菌核酸酶结构域 1 (*SND1*) 在很多肿瘤中都存在表达量升高的现象,在胶质瘤中 *SND1* 可以通过重塑染色质构象,

从而实现长距离诱导 *RhoA* 启动子的表达,促进 *RhoA* 的转录。转录生成的 *RhoA* 可以顺序调节 *CCND1*、*CCNE1*、*CDK4* 和 *CDKN1B* 的表达,并加速神经胶质瘤细胞增殖中的 G1/S 相变,促进肿瘤的发生。*SND1* 作为一种新型的染色质结构修饰剂,有望成为今后神经胶质瘤分类治疗的预后指标。

#### (9) 急性淋巴细胞白血病

白血病是一种造血干细胞恶性克隆性疾病。根据白血病的分化程度、自然病程的长短可分为急、慢性白血病。临床上常将白血病分为淋巴细胞白血病、髓细胞白血病、混合细胞白血病等。急性淋巴细胞白血病 (Acute lymphoblastic leukemia, ALL) 是一种起源于淋巴 B 系或 T 系细胞在骨髓内异常增生的恶性肿瘤性疾病,在 20 岁以下的白血病患者中,超过一半的患者被诊断为 ALL。目前 T 系细胞引发的急性淋巴细胞白血病 (T-ALL) 原发性耐药和复发性白血病患者的预后仍然不佳<sup>[80]</sup>。

Whitehead 研究所 Hnisz 等<sup>[111]</sup>绘制了 T-ALL 中绝缘区域的互作谱图,发现肿瘤细胞基因组存在特异性的基因微缺失,该缺失使得环状染色质结构被破坏,导致 T-ALL 的原癌基因被激活。在许多类型的癌症中都发现了染色体内邻域边界的突变。因此,推测癌基因激活可以通过破坏恶性细胞中绝缘区的遗传改变而发生。纽约大学医学院的研究人员<sup>[112]</sup>探索了原代 T-All 细胞与正常的 T 细胞中染色质三维结构变化,并展示了 TADs 如何通过改变启动子-增强子结构、摆脱转录程序的控制并最终导致疾病的发生。他们发现 TADs 活性与 CTCF 富集程度、超增强子活性和白血病相关基因表达相关。此外,他们还发现了一个围绕 *myc* 基因位点的特异性“TADs 融合”事件,该融合很可能通过影响超级增强子和启动子之间的染色质相互作用来驱动 *myc* 的过表达。这种 TADs 的融合伴随着 TAD 边界的 CTCF 完全缺失和 T-ALL 样本中 *myc* 表达的增加。此发现为染色质

结构的改变如何通过重组癌基因和抑癌基因结构影响疾病进展提供了新的思路。

清华大学陈阳研究员、张奇伟教授合作<sup>[113]</sup>研究了以全反式维甲酸 (All-trans-retinoic acid, ATRA) 诱导的人早幼粒白血病细胞 (HL-60) 的染色质三维构象, 发现在细胞分化过程中, 约 90% 的 TAD 边界未发生改变, 有不到 1% 的 Compartment 发生 A、B 间的转换, 但 TADs 内部的相互作用发生了特异性变化, 并促使了基因的差异表达。研究还鉴定出 ATRA 诱导分化后基因区-调控区的相互作用变化 (在原有基础上减少了 330 对互作, 新增了 422 对互作)。通过 ATAC-seq 和转录调控网络分析, 他们还发现 ATRA 诱导后细胞内转录因子编码基因 *GATA2* 和锌指蛋白编码基因 *ZBTB16* 相关的互作显著减少, 并且通过基序 (Motif) 分析表明 *GATA2* 可与自身的启动子结合从而调控自身的表达, 为白血病新型潜在药物靶点的挖掘提供了研究基础。

## 4 总结与展望

通过三维基因组学, 可以清晰地认识到染色质的空间构象变化机制、基因的转录调控机制、生物学性状的形成机制、信号通路的传导机制以及基因组的运行机制。结合 3C 技术和其他衍生测序技术应用于三维基因组学研究, 使得科学家们对动植物及微生物的三维基因组空间构象认知越来越清晰。虽然目前三维基因组学在生物学各个领域得到了广泛的应用, 在癌症发生机制的研究上也取得了阶段性成果。但仍然存在一些问题, 需要科学家们继续研究与探索。

(1) 新型测序技术的开发: Hi-C 技术作为三维基因组学研究的首选方法, 可以全面获取细胞内染色质相互作用, 但通过 Hi-C 技术获得的信息庞大冗杂, 后续的数据整合及不同细胞三维结构差异分析非常复杂; 同时高成本、低成功率也是限制 Hi-C 技术广泛应用的因素, 虽然 DLO Hi-C

可以简化 Hi-C 技术的步骤和成本; Capture-C 可以减少对样本细胞量的需求, 但是仍然需要有更多更具创新性的测序技术待开发, 以满足三维基因组学研究过程中不同的需求。

(2) 可视化分析工具的研发: 三维基因组学研究过程中涉及的数据规模较大, 经典的可视化工具如 Browser、DNA Database 等并不能达到展示三维基因组学数据特征的效果, 这一限制给不同类型的多组学数据整合的分析带来了一定的困难。因此开发可以进行多组学数据整合、展示空间构象与基因表达联系的可视化工具, 对三维基因组学的发展有极大的推动作用。

(3) 单细胞水平的三维基因组研究: 单细胞测序技术的发展, 给三维基因组结构在单细胞水平的研究提供了新的思路, 如何利用相关测序技术探究三维基因组结构及染色体互作的异质性, 讨论不同因素对三维基因组结构的影响, 以及进行个性化基因组研究, 也将是三维基因组学研究热点。

(4) 四维基因组学: 基因组在细胞核内的空间构象是随着时间不断变化的, 以保证细胞正常的生长发育。因此, 时间因素也是影响基因组结构的重要因素, 随着“4D 核体计划”的推进, 对不同时间细胞三维基因组结构的动态变化趋势研究也将成为可能。

综上, 三维基因组学的研究仍处于初级阶段, 如何将目前获得的阶段性数据转化为具有临床应用价值的研究成果, 并真正促进相关领域的发展, 是三维基因组学研究的重中之重, 如在逐渐了解如癌症等疾病发生过程中, 染色质内部和染色质间的结构变化后, 能否根据这些变化寻找到合适的治疗靶点, 开发靶向药物, 从而调控染色质的相互作用, 以达到治疗疾病的目的; 在了解动植物的复杂性状的基因调控网络与机制后, 如何利用这些优势三维结构特征, 提升农业生物技术水平, 提高人民生活质量, 这些科学问题还需要科学家们继续深入探索。

---

**REFERENCES**

- [1] Tanay A, Cavalli G. Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. *Curr Opin Genet Dev*, 2013, 23(2): 197–203.
- [2] Gordon S. Genomics and world health: report of the advisory committee on health research. Geneva: World Health Organization, 2002. x+248pp. Price Sw.fr.35/US\$31.50 (in developing countries Sw.fr.14). ISBN 92-4-154554-2. *Trans Roy Soc Trop Med Hyg*, 2002, 96(6): 669.
- [3] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 2004, 50(2): 931–945.
- [4] The ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA Elements) project. *Science*, 2004, 306(5696): 636–640.
- [5] Dekker J, Rippe K, Dekker M, et al. Capturing chromosome conformation. *Science*, 2002, 295(5558): 1306–1311.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Liu C. *In situ* Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. *Methods Mol Biol*, 2017, 1629: 155–166.
- [10] 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.
- [11] Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature*, 2009, 462(7269): 58–64.
- [12] Zhang YB, Wong CH, Birnbaum RY, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*, 2013, 504(7479): 306–310.
- [13] Dekker J, Belmont AS, Guttman M, et al. The 4D nucleome project. *Nature*, 2017, 549(7671): 219–226.
- [14] Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. *Cell*, 2013, 152(6): 1270–1284.
- [15] Dixon JR, Gorkin DU, Ren B. Chromatin domains: the unit of chromosome organization. *Mol Cell*, 2016, 62(5): 668–680.
- [16] Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell*, 2013, 49(5): 773–782.
- [17] J. B. F. Die stofflichen Grundlagen der Vererbung im organischen Reich. *Nature*, 1906, 75(1935): 98–99.
- [18] Cremer T, Kurz A, Zirbel R, et al. Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb Symp Quant Biol*, 1993, 58: 777–792.
- [19] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2001, 2(4): 292–301.
- [20] Parada LA, McQueen PG, Misteli T. Tissue-specific spatial organization of genomes. *Genome Biol*, 2004, 5(7): R44.
- [21] Zink D, Cremer T, Saffrich R, et al. Structure and dynamics of human interphase chromosome territories *in vivo*. *Hum Genet*, 1998, 102(2): 241–251.
- [22] Hlatky L, Sachs RK, Vazquez M, et al. Radiation-induced chromosome aberrations: insights gained from biophysical modeling. *Bioessays*, 2002, 24(8): 714–723.
- [23] Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol*, 2006, 4(5): e138.
- [24] Mehta IS, Kulashreshtha M, Chakraborty S, et al. Chromosome territories reposition during DNA damage-repair response. *Genome Biol*, 2013, 14(12): R135.
- [25] 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.
- [26] Solovei I, Thanisch K, Feodorova Y. How to rule the nucleus: *divide et impera*. *Curr Opin Cell Biol*, 2016, 40: 47–59.
- [27] van Steensel B, Belmont AS. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, 2017, 169(5): 780–791.
- [28] Ganai N, Sengupta S, Menon GI. Chromosome positioning from activity-based segregation. *Nucleic Acids Res*, 2014, 42(7): 4145–4159.
- [29] Falk M, Feodorova Y, Naumova N, et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, 2019, 570(7761): 395–399.
- [30] Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 2012, 485(7398): 381–385.
- [31] 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.
- [32] 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.
- [33] Kaaij LJT, van der Weide RH, Ketting RF, et al. Systemic loss and gain of chromatin architecture throughout zebrafish development. *Cell Rep*, 2018, 24(1): 1–10.e4.
- [34] Sexton T, Yaffe E, Kenigsberg E, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 2012, 148(3): 458–472.
- [35] Zhou ZK, Li M, Cheng H, et al. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat Commun*, 2018, 9: 2648.
- [36] Crane E, Bian Q, McCord RP, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, 2015, 523(7559): 240–244.
- [37] Dong QL, Li N, Li XC, et al. Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. *Plant J*, 2018, 94(6): 1141–1156.
- [38] Sun W, Leng L, Yin QG, et al. The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J*, 2019, 97(5): 841–857.
- [39] Zuin J, Dixon JR, van der Reijden MIJA, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA*, 2014, 111(3): 996–1001.
- [40] Wang Q, Sun Q, Czajkowsky DM, et al. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun*, 2018, 9(1): 188.
- [41] Pope BD, Ryba T, Dileep V, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature*, 2014, 515(7527): 402–405.
- [42] Benedetti F, Dorier J, Burnier Y, et al. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Res*, 2014, 42(5): 2848–2855.
- [43] Sanborn AL, Rao SS, Huang SC, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci USA*, 2015, 112(47): E6456–E6465.
- [44] Fudenberg G, Imakaev M, Lu C, et al. Formation of chromosomal domains by loop extrusion. *Cell Rep*, 2016, 15(9): 2038–2049.
- [45] Sanyal A, Lajoie BR, Jain G, et al. The long-range interaction landscape of gene promoters. *Nature*, 2012, 489(7414): 109–113.
- [46] Würtele H, Chartrand P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended chromosome conformation capture methodology. *Chromosome Res*, 2006, 14(5): 477–495.
- [47] Splinter E, de Wit E, Nora EP, et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev*, 2011, 25(13): 1371–1383.
- [48] Belton JM, Dekker J. Chromosome conformation capture carbon copy (5C) in budding yeast. *Cold Spring Harb Protoc*, 2015, 2015(6): 593–598.

- [49] Moisan S, Nabec AL, Quillévéré A, et al. Characterization of *GJB2* cis-regulatory elements in the *DFNB1* locus. *Hum Genet*, 2019, 138(11/12): 1275–1286.
- [50] Burton JN, Adey A, Patwardhan RP, et al. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol*, 2013, 31(12): 1119–1125.
- [51] Javierre BM, Burren OS, Wilder SP, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*, 2016, 167(5): 1369–1384.e19.
- [52] 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.
- [53] Taberlay PC, Achinger-Kawecka J, Lun ATL, 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.
- [54] Dryden NH, Broome LR, Dudbridge F, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res*, 2014, 24(11): 1854–1868.
- [55] Schulz S, Häussler S. Chromatin immunoprecipitation for ChIP-chip and ChIP-seq. *Methods Mol Biol*, 2014, 1149: 591–605.
- [56] Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 2007, 316(5830): 1497–1502.
- [57] Wei CL, Wu Q, Vega VB, et al. A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 2006, 124(1): 207–219.
- [58] Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 2013, 10(12): 1213–1218.
- [59] Song LY, Crawford GE. 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.
- [60] Giresi PG, Kim J, McDaniell RM, et al. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*, 2007, 17(6): 877–885.
- [61] Kelso TWR, Porter DK, Amaral ML, et al. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *Elife*, 2017, 6: e30506.
- [62] 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.
- [63] Liang ZY, Li GP, Wang ZJ, et al. BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat Commun*, 2017, 8(1): 1622.
- [64] Burton A, Torres-Padilla ME. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol*, 2014, 15(11): 723–735.
- [65] Xue PC, Yu W, Ke L, et al. Key role for CTCF in establishing chromatin structure in human embryos. *Nature*, 2019, 576(7786): 306–310.
- [66] Ke YW, Xu YN, Chen XP, et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell*, 2017, 170(2): 367–381.e20.
- [67] Wang L, Zhang J, Duan JL, et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell*, 2014, 157(4): 979–991.
- [68] Lin AW, Barradas M, Stone JC, et al. Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. *Genes Dev*, 1998, 12(19): 3008–3019.
- [69] Narita M, Nuñez S, Heard E, et al. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell*, 2003, 113(6): 703–716.
- [70] Chandra T, Ewels PA, Schoenfelder S, et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep*, 2015, 10(4): 471–483.
- [71] Zirkel A, Nikolic M, Sofiadis K, et al. HMGB2 loss upon senescence entry disrupts genomic organization and induces CTCF clustering across cell types. *Mol Cell*, 2018, 70(4): 730–744.e6.

- [72] Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 2015, 518(7539): 331–336.
- [73] Zhang YX, Li T, Preissl S, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet*, 2019, 51(9): 1380–1388.
- [74] Kemeny S, Tatout C, Salaun G, et al. Spatial organization of chromosome territories in the interphase nucleus of trisomy 21 cells. *Chromosoma*, 2018, 127(2): 247–259.
- [75] Song M, Yang XY, Ren XJ, et al. Mapping *cis*-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet*, 2019, 51(8): 1252–1262.
- [76] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin*, 2018, 68(6): 394–424.
- [77] Chen WQ, Zheng RS, Baade PD, et al. Cancer statistics in China, 2015. *CA A Cancer J Clin*, 2016, 66(2): 115–132.
- [78] Chen H, Li CY, Peng XX, et al. A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell*, 2018, 173(2): 386–399.e12.
- [79] Lewis DR, Chen HS, Cockburn MG, et al. Early estimates of SEER cancer incidence, 2014. *Cancer*, 2017, 123(13): 2524–2534.
- [80] Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*, 2019, 69(5): 363–385.
- [81] Wang ZF, Tu KL, Xia L, et al. The open chromatin landscape of non-small cell lung carcinoma. *Cancer Res*, 2019, 79(19): 4840–4854.
- [82] Ahmad A. Breast cancer statistics: recent trends// Ahmad A, Ed. *Breast Cancer Metastasis and Drug Resistance*. Cham: Springer, 2019, 1152: 1–7.
- [83] Meaburn KJ, Misteli T. Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol*, 2008, 180(1): 39–50.
- [84] Meaburn KJ, Gudla PR, Khan S, et al. Disease-specific gene repositioning in breast cancer. *J Cell Biol*, 2009, 187(6): 801–812.
- [85] 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(1): 214.
- [86] Zhou Y, Gerrard DL, Wang JB, et al. Temporal dynamic reorganization of 3D chromatin architecture in hormone-induced breast cancer and endocrine resistance. *Nat Commun*, 2019, 10(1): 1522.
- [87] The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*, 2015, 163(4): 1011–1025.
- [88] Rhie SK, Perez AA, Lay FD, et al. A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nat Commun*, 2019, 10(1): 4154.
- [89] Guo Y, Perez AA, Hazelett DJ, et al. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. *Genome Biol*, 2018, 19(1): 160.
- [90] Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. *Lancet*, 2019, 394(10207): 1467–1480.
- [91] Boland CR, Shin SK, Goel A. Promoter methylation in the genesis of gastrointestinal cancer. *Yonsei Med J*, 2009, 50(3): 309–321.
- [92] 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.
- [93] Seaman L, Chen HM, Brown M, et al. Nucleome analysis reveals structure-function relationships for colon cancer. *Mol Cancer Res*, 2017, 15(7): 821–830.
- [94] Bian SH, Hou Y, Zhou X, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, 2018, 362(6418): 1060–1063.
- [95] Manna ED, Seixas AA, De Araújo RP, et al. Primary adenosquamous carcinoma of the stomach. *Rev Assoc Med Bras*, 1998, 44(2): 152–154.
- [96] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA A Cancer J Clin*, 2017, 67(1): 7–30.
- [97] Yao F, Kausalya JP, Sia YY, et al. Recurrent fusion genes in gastric cancer: *CLDN18-ARHGAP26* induces loss of epithelial integrity. *Cell Rep*, 2015, 12(2): 272–285.

- [98] Shu Y, Zhang WH, Hou QQ, et al. Prognostic significance of frequent CLDN18-ARHGAP26/6 fusion in gastric signet-ring cell cancer. *Nat Commun*, 2018, 9(1): 2447.
- [99] Xue RD, Chen L, Zhang C, et al. Genomic and transcriptomic profiling of combined hepatocellular and intrahepatic cholangiocarcinoma reveals distinct molecular subtypes. *Cancer Cell*, 2019, 35(6): 932–947.e8.
- [100] Ružić M, Pellicano R, Fabri M, et al. Hepatitis C virus-induced hepatocellular carcinoma: a narrative review. *Panminerva Med*, 2018, 60(4): 185–191.
- [101] Medical Administration and Hospital Administration of The Health and Family Planning Commission, PRC. Guidelines for the Diagnosis and Treatment of primary Liver Cancer (2017 edition). *Chin J Dig Surg*, 2017, 16(7): 635–647 (in Chinese). 中华人民共和国卫生和计划生育委员会医政医管局. 原发性肝癌诊疗规范(2017年版). *中华消化外科杂志*, 2017, 16(7): 635–647.
- [102] Moreau P, Cournac A, Palumbo GA, et al. Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. *Nat Commun*, 2018, 9(1): 4268.
- [103] Zhou B, Ho SS, Greer SU, et al. Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. *Nucleic Acids Res*, 2019, 47(8): 3846–3861.
- [104] Hideshima T, Bergsagel PL, Kuehl WM, et al. Advances in biology of multiple myeloma: clinical applications. *Blood*, 2004, 104(3): 607–618.
- [105] Wu PZ, Li TT, Li RF, et al. 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat Commun*, 2017, 8(1): 1937.
- [106] Parsons DW, Jones S, Zhang XS, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 2008, 321(5897): 1807–1812.
- [107] The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med*, 2015, 372(26): 2481–2498.
- [108] Flavahan WA, Drier Y, Liao BB, et al. Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature*, 2016, 529(7584): 110–114.
- [109] Xiao DS, Huang J, Pan Y, et al. Chromatin remodeling factor LSH is upregulated by the LRP6-GSK3 $\beta$ -E2F1 axis linking reversely with survival in gliomas. *Theranostics*, 2017, 7(1): 132–143.
- [110] Yu L, Xu JL, Liu J, et al. The novel chromatin architectural regulator SND1 promotes glioma proliferation and invasion and predicts the prognosis of patients. *Neuro Oncol*, 2019, 21(6): 742–754.
- [111] Hnisz D, Weintraub AS, Day DS, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 2016, 351(6280): 1454–1458.
- [112] Kloetgen A, Thandapani P, Ntziachristos P, et al. Dynamic 3d chromosomal landscapes in acute leukemia. *Cancer Res*, 2018, 78(13): 2998.
- [113] Li YJ, He Y, Liang ZY, et al. Alterations of specific chromatin conformation affect ATRA-induced leukemia cell differentiation. *Cell Death Dis*, 2018, 9(2): 200.

(本文责编 陈宏宇)