

DOI:10.12113/202210025

# 基于 NGS 的染色质测序在肿瘤研究中的应用

蒲鹏, 李国强, 刘法涛, 刘颖斌\*

(上海交通大学医学院附属仁济医院 胆胰外科, 上海市肿瘤研究所,  
癌基因与相关基因国家重点实验室, 上海 200127)

**摘要:**染色质是真核生物细胞核内由核酸和蛋白质组成的复合结构, 有着精密且复杂的三维结构。染色质除基本的 DNA 序列外, 内部还存在着不同化学修饰, DNA-蛋白质相互作用, DNA-DNA 相互作用和 DNA-RNA 相互作用, 以上这些若发生改变都可能在肿瘤发生发展过程中起到至关重要的作用。通过不同的染色质测序方法, 可以解析出这些改变, 并进一步加深研究者对肿瘤形成机制的理解, 最终应用于肿瘤的治疗。本文对常见的染色质测序技术部分原理和应用进行综述。

**关键词:** NGS; 染色质; 肿瘤

中图分类号: Q75 文献标志码: A 文章编号: 1672-5565(2024)02-079-14

## Application of NGS-based chromatin sequencing in cancer research

PU Peng, LI Guoqiang, LIU Fatao, LIU Yingbin\*

(Department of Biliary-Pancreatic Surgery, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai Cancer Institute, State Key Laboratory of Oncogenes and Related Genes, Shanghai Jiao Tong University, Shanghai 200127, China)

**Abstract:** All eukaryotic genomes are organized into a nucleoprotein complex called chromatin, which have a sophisticated and intricate three-dimensional structure. In addition to the basic DNA sequence, there also exist DNA modifications, DNA-protein interactions, DNA-DNA interactions and DNA-RNA interactions in chromatin, all of which, if altered, may play a vital role in cancer initiation and development. Through different chromatin sequencing techniques, researchers are able to parse these alterations and enhance our understanding of tumorigenesis, which could ultimately be applied to the cancer therapy. In this paper, we review some experimental principles and applications of common chromatin sequencing techniques.

**Keywords:** NGS; Chromatin; Cancer

染色质是真核生物细胞核内由核酸和蛋白质组成的复合结构<sup>[1]</sup>, 其中核酸包括 DNA 和少量 RNA, 蛋白质则包括组蛋白和非组蛋白<sup>[2]</sup>。

人类基因组序列若线性展开长约 2 m, 但细胞核只是微米级别, 这提示我们 DNA 势必经过精密且复杂的折叠才能组装进细胞核内。关于染色质的结构, 目前主流观点认为: 每约 147 bp 线性 DNA 首先缠绕八聚核小体 1.75 圈组成基本单元核小

体<sup>[3]</sup>, 核小体再经螺旋盘绕形成 30 nm 螺线管结构, 随后在 cohesin 和 CTCF (CCCTC-binding factor) 等蛋白作用下内聚成环 (Loops), 这样使得线性距离较远的基因座 (Chromatin loci) 在空间上得以接近, 接下来环继续折叠形成拓扑相关结构域 (Topologically associated domains, TADs), 最终被分配到不同的隔室 (Compartment A/B)<sup>[4]</sup>。

鉴于染色质结构和功能的复杂性, 染色质测序

收稿日期: 2022-10-31 修回日期: 2023-03-19; 网络首发日期: 2023-06-02.

网络首发地址: <https://kns.cnki.net/kcms2/detail/23.1513.Q.20230601.1404.006.html>

基金项目: 国家自然科学基金 (No.3213000192, 81874181).

\* 通信作者: 刘颖斌, 男, 主任医师, 博导. 研究方向: 胆道系统、胰腺恶性肿瘤临床与基础研究. E-mail: laoniulyb@shsmu.edu.cn.

引用格式: 蒲鹏, 李国强, 刘法涛, 等. 基于 NGS 的染色质测序在肿瘤研究中的应用 [J]. 生物信息学, 2024, 22(2): 79-92.

PU Peng, LI Guoqiang, LIU Fatao, et al. Application of NGS-based chromatin sequencing in cancer research [J]. Chinese Journal of Bioinformatics, 2024, 22(2): 79-92.

涉及多种 NGS 技术,主要包括以下方面(图 1):

1) 基因组 DNA 序列:全基因组测序<sup>[5]</sup> (Whole genome sequencing, WGS), 全外显子测序<sup>[6]</sup> (Whole exome sequencing, WES)。

2) DNA 甲基化修饰:全基因组重亚硫酸盐甲基化测序<sup>[7-9]</sup> (Whole genome bisulfite sequencing, WGBS), 氧化结合重亚硫酸盐测序<sup>[10]</sup> (Oxidative bisulfite sequencing, oxBS-Seq)。

3) DNA-蛋白质相互作用:染色质免疫共沉淀测序<sup>[11-13]</sup> (Chromatin immunoprecipitation sequencing, ChIP-Seq), 染色质内源性切割与高通量测序<sup>[14]</sup> (Chromatin endogenous cleavage with high-throughput sequencing, ChEC-Seq), CUT&Tag<sup>[15]</sup> (Cleavage under targets and tagmentation)。

4) 染色质可及性:转座酶染色质可及性高通量测序分析<sup>[16]</sup> (Assay for transposase-accessible

chromatin with high throughput sequencing, ATAC-seq), 微球菌核酸酶消化测序<sup>[17-18]</sup> (Micrococcal nuclease digestion with deep sequencing, MNase-seq), 脱氧核糖核酸酶 I 超敏位点测序<sup>[19]</sup> (DNase I hypersensitive sites sequencing, DNase-seq), 甲醛辅助分离调控元件<sup>[20]</sup> (Formaldehyde-assisted isolation of regulatory elements, FAIRE-seq)。

5) DNA-DNA 相互作用:基于染色体构象捕获 (Chromosome conformation capture, 3C) 的测序技术, 主要包括 3C<sup>[21]</sup>, 4C<sup>[22-23]</sup>, 5C<sup>[24]</sup>, HiC<sup>[25]</sup>, 与 ChIP-Seq 结合有 HiChIP<sup>[26]</sup>, ChIA-PET<sup>[27]</sup> 等。

6) DNA-RNA 相互作用:原位全 RNA-DNA 相互作用测序<sup>[28]</sup> (*in situ* global RNA interactions with DNA by deep sequencing, GRID-seq), RNA-基因组原位定位<sup>[29]</sup> (*in situ* mapping of RNA-genome interactome, iMARGI)。

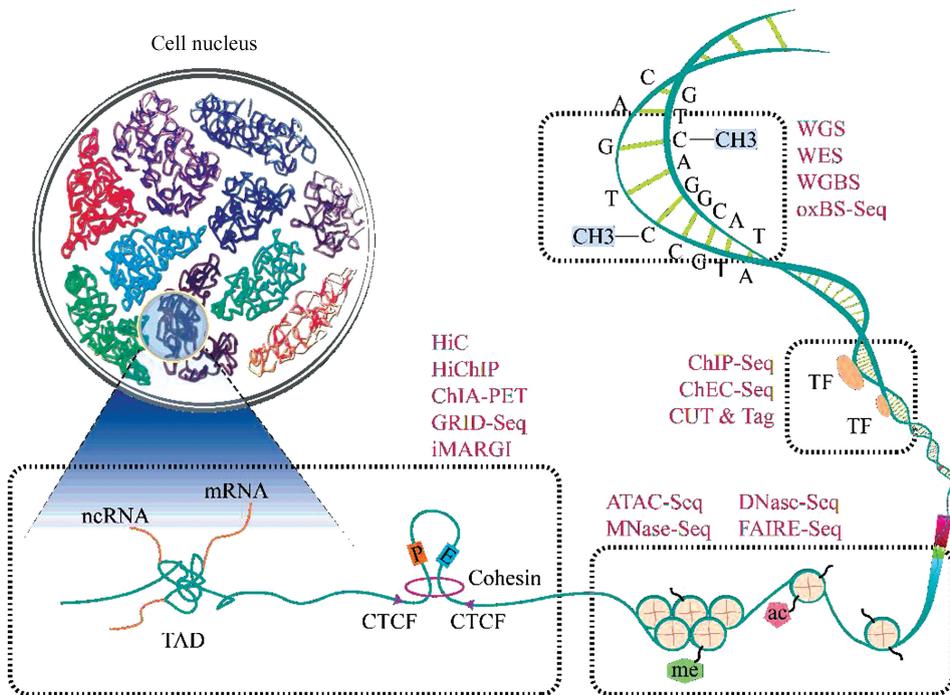


图 1 染色质不同层次结构及相应测序技术

Fig.1 Different hierarchical chromatin structures and corresponding sequencing techniques

## 1 染色质测序技术原理及应用

### 1.1 基因组 DNA 序列

#### 1.1.1 WGS

WGS 是利用 NGS 基本原理针对全基因组进行测序,不同的测序平台使用的方法可能略有区别,但主要流程都包括 DNA 纯化(DNA purification)、DNA 片段化(DNA shearing)、片段选择(Size selection)、

末端修复(End repair)、加接头(Adaptor ligation)、PCR 扩增(PCR amplification),文库质控(Quality assessment of libraries)和上机测序(Sequencing)<sup>[30]</sup>。最终研究人员再通过数据分析<sup>[31-32]</sup>可以获得样本 DNA 的变异信息,包括点突变(Single nucleotide polymorphism, SNP)、短片段插入缺失(Small deletion and insertion, indel)、拷贝数变异(Copy number variations, CNVs)、结构变异(Structural variations, SVs)等,而获得这些突变信息则有助于研究者更好

地理解与阐述肿瘤发生发展的基础机制。

### 1.1.2 WES

外显子组约占基因组的 1%,但 85% 疾病相关突变存在于外显子<sup>[33]</sup>。与 WGS 不同的是,WES 只需设计外显子探针即可捕获外显子序列并进行高通量测序。相较于 WGS,WES 更经济,数据量更小,应用更加广泛,但其包含的信息不如 WGS 全面,非编码区域和内含子相关致病突变无法检出<sup>[34]</sup>。

### 1.1.3 应用

目前,多种肿瘤类型和超过 5 万例的肿瘤基因组已完成了测序,研究人员并依此建立了多个肿瘤相关数据库并描绘出了各肿瘤的突变图谱,如 TCGA<sup>[35]</sup>,ICGC<sup>[36]</sup>和 COSMIC<sup>[37]</sup>等。TCGA 数据库覆盖超 30 个常见癌种,其中肺癌、乳腺癌、结直肠癌和血液病更是多达近万例,数据库提供包括基因组突变信息、拷贝数变化、表观遗传改变和基因表达谱等数据,在肿瘤分子分型、肿瘤特异标志物和药物靶向治疗等方面参考作用极大。ICGC 是国际癌症基因组联合体,其目标是在全球范围全面阐述全球人类多达 50 种不同癌症类型中存在的基因组变化。截至目前,ICGC (Release 28) 包含 24 298 个肿瘤基因组,检测到 81 782 588 个体细胞突变,涉及 57 905 个突变基因。COSMIC 是世界上最大最全的关于肿瘤体细胞突变及其作用的数据库。最新版本 (v97, 2022-11-29) 中涵盖超过 150 万个肿瘤样本的 23 443 841 个基因组变异,16 015 511 个非编码突变、19 428 个融合突变,321 804 个基因组重排,1 207 190 个拷贝数变异,9 215 470 个表达异常和 7 930 489 差异甲基化 CpGs 的详细信息。

研究表明,不同的肿瘤突变图谱表现不尽相同。例如,在乳腺癌中根据 ER+/-与否,突变谱略有不同,但高频突变均主要为 *TP53*, *PIK3CA*, *MYC*, *ERBB2*, *CCND1*, *ZNF703/FGFR1*, *GATA3* 和 *RB1* 等<sup>[38-40]</sup>。而在胰腺导管腺癌中,*KRAS* 突变比例则高达 90%,其余靠前的有 *TP53*, *CDKN2A*, *SMAD4* 和 *BRCA1/2*<sup>[41]</sup>等。笔者团队通过 WES 技术鉴定出 *ERBB2/3* 通路突变是胆囊癌的高频突变<sup>[42]</sup>,此后在印度的一项研究中这一结果也进一步得到了证实<sup>[43]</sup>,随后我们分别在细胞水平<sup>[44-45]</sup>和单细胞转录组水平<sup>[46]</sup>发现 *ERBB2/3* 突变与 *PD-L1*, *MDK* 表达相关,从而影响胆囊癌免疫微环境。进一步笔者团队还描绘了胆囊神经内分泌肿瘤的突变谱,其中高频突变为 *TP53*, *SNX27*, *PCDHB14*, *RB1*, *SRPK1* 和 *ZNF107* 等<sup>[47]</sup>。

通过对肿瘤驱动基因的突变情况进行描述,有助于肿瘤发生发展基础机制的研究和分子分型,最

终实现靶向治疗、精准治疗<sup>[48-50]</sup>。

## 1.2 DNA 甲基化

实际上,除了 DNA 序列本身发生突变外,DNA 还会发生许多化学修饰,目前已知的修饰至少有 17 种<sup>[51]</sup>,其中以胞嘧啶 (C) 的 5 号碳原子甲基化 (5-methylcytosine, 5mC) 修饰丰度最高,因此 5mC 也被称为第 5 种碱基。这些修饰能在不改变 DNA 序列的前提下引起 DNA 构象、DNA 稳定性和 DNA-蛋白质互作的改变,从而影响基因表达,在肿瘤进程中发挥重要作用。

### 1.2.1 WGBS

5mC 甲基化修饰大多发生在 CpG 双核苷酸,可在转录水平抑制基因表达。超过半数基因的启动子区域会有 CpG 的富集,这些区域称为 CpG 岛 (CpG island, CpGi)<sup>[52]</sup>。WGBS 的基本原理是利用重亚硫酸盐处理 DNA,其中未被甲基化修饰的胞嘧啶 (C) 会脱氨基转换为尿嘧啶 (U),再经后续的 PCR 扩增,U 会转换为 T,从而区分甲基化 C 和未甲基化 C,最后结合高通量测序技术,即可获得单碱基分辨率的全基因组 5mC 修饰图谱<sup>[7-9, 53]</sup>。

### 1.2.2 oxBS-Seq

5mC 在 TET 酶作用下产生 5-羟甲基胞嘧啶 (5-hydroxymethylcytosine, 5hmC), 5hmC 可能是活化 DNA 去甲基过程的中间产物,但同时也是一种稳定的 DNA 修饰,具体作用机制还不十分清楚,可能与 DNA 的主动/被动去甲基化有关<sup>[54]</sup>。WGBS 无法区分 5mC 和 5hmC,而 oxBS-Seq 首先将 5hmC 氧化为 5-甲酰胞嘧啶 (5-formylcytosine, 5fC),再经重亚硫酸盐处理则会转换为 U,通过与 BS-Seq 的结果取差集即可实现 5mC 的精准检测和 5hmC 的检出<sup>[10, 53, 55]</sup>,从而获得单碱基分辨率的 5hmC 修饰图谱。

### 1.2.3 应用

DNA 甲基化多起到抑制转录的作用,主要有影响转录因子与启动子结合、直接与转录抑制因子结合、影响组蛋白修饰和染色质结构等机制<sup>[56-57]</sup>,而 DNA 甲基表观修饰的异常和失衡在肿瘤的发生和发展中发挥重要作用。例如:研究人员利用 WGBS 描绘了前列腺癌甲基化图谱,发现 *RXRG*, *FH*, *CCDC6*, *RARA* 和 *GPX3* 等抑癌基因呈高甲基化并低表达,而 *NCOA4*, *BIRC2*, *FGF6* 和 *HGF* 等癌基因呈低甲基化并高表达<sup>[57-58]</sup>;苏建忠教授团队通过对 210 例早期乳腺癌患者的 cfDNA 甲基化情况进行分析,成功构建了一个新的早期乳腺癌预测模型,结合影像学可提高诊断的准确性<sup>[59]</sup>;Berman B P 等<sup>[60]</sup>研究人员发现与正常相比,结直肠癌包含更多低甲基化区域,而这些区域在局部表现为高甲基化但在

长距离(>100 kb)下表现为低甲基化,并且这些区域与染色质高级结构相关。

DNA 甲基化修饰癌症发生发展密切相关<sup>[61]</sup>,利用 WGBS 等技术,研究人员可对 DNA 表观修饰图谱进行描绘<sup>[62]</sup>,丰富肿瘤发生机制,为肿瘤早期诊断、治疗提供另一思路。

### 1.3 DNA-蛋白质相互作用

染色质内的蛋白质包括组蛋白和非组蛋白,非组蛋白主要有转录因子(Transcription factors, TFs)、各类酶以及维持染色质拓扑结构的蛋白等。其中,组蛋白还可以发生多种修饰,如乙酰化、磷酸化、甲基化、苏木化、泛素化<sup>[63]</sup>等,这些修饰多见于 H3 的赖氨酸(K)和精氨酸(R)位点。蛋白质-DNA 相互作用同样属于表观遗传学范畴,在转录调控、DNA 复制、DNA 损伤修复、DNA 包装和修饰等方面发挥重要作用<sup>[64]</sup>。

#### 1.3.1 ChIP-Seq

ChIP-Seq 是 ChIP 和高通量测序结合的技术,首先利用甲醛交联蛋白质-DNA,随后分离 DNA,并用超声波使其片段化,再通过特异性抗体与靶蛋白形成免疫复合物沉淀,最后解交联纯化得到目的 DNA 序列<sup>[11-13, 65]</sup>。再经测序和分析后,即可得到靶蛋白与目的 DNA 序列的互作图谱。

#### 1.3.2 ChEC-Seq

尽管 ChIP-Seq 已在研究蛋白质-DNA 互动中发挥重要作用,但该技术仍存在一定的缺陷,主要原因在于使用了甲醛交联,甲醛处理会增加目的蛋白与其他可溶性蛋白的结合,从而增加非特异性信号<sup>[66]</sup>。ChEC-Seq 是基于 MNase 的分析技术,首先需要构建靶蛋白-MNase 融合蛋白,低  $Ca^{2+}$  浓度下 MNase 不发挥作用,当添加合适的  $Ca^{2+}$  后, MNase 会切割靶蛋白两端的 DNA 序列,从而释放出目的序列<sup>[14, 67]</sup>。ChEC-Seq 避免了甲醛交联和超声打断 DNA 的步骤,有效地规避了 ChIP 假阳性的问题,但该技术针对不同蛋白均需构建蛋白-MNase 融合体,技术难度较高。

#### 1.3.3 CUT&Tag

CUT&Tag 是 Henikoff 博士研发的一种新的蛋白质-DNA 互作分析方法<sup>[15]</sup>,与 ChIP-Seq 相比,具有操作简便,细胞需求量低,背景噪音低等优点。其基本原理<sup>[15]</sup>是利用了预装 DNA 接头的 protein A/G-Tn5 转座酶融合蛋白,首先一抗与靶蛋白特异性结合,二抗放大信号,随后融合蛋白与二抗结合,通过加入  $Mg^{2+}$  即可激活 Tn5 转座酶的活性,切割与靶蛋白结合的 DNA 序列并插入接头序列,最后分离标签化的 DNA 片段,即可用于建库测序。

### 1.3.4 应用

基因组正常功能的发挥很大程度上依赖于动态的 DNA-蛋白互作的调节,一旦这些平衡被打破,细胞内环境稳态将遭到破坏,从而引起病变甚至癌变。例如, Koeffler 等<sup>[68]</sup>在食管癌细胞系中以 H3K4me1, H3K4me3, H3K27ac 以及转录因子等为靶标进行 ChIP-Seq 分析,鉴定了 *ALDH3A1* 超级增强子调控通路, *TP63*, *SOX2* 和 *KLF5* 等转录因子起到关键调控作用;武汉大学吴旻教授<sup>[69]</sup>团队以 H3K27ac, H3K4me3 为靶标对 73 对结直肠癌标本进行 ChIP-Seq 分析,同时整合 RNA-Seq 和 WGS 数据,描绘了结直肠癌中增强子和超级增强子的改变情况,并在细胞系中验证了相关癌基因 *KLF3*, *MAFK* 和 *RUNX1* 等的生物学功能; Janssens 等<sup>[70]</sup>利用 CUT&Tag 对急性髓系和淋巴系白血病的易位基因 *KMT2A* 进行了研究,发现 *KMT2A* 与 *HOX9*, *MEIS1* 和 *MEF2C* 等基因发生融合,融合蛋白可改变染色质组蛋白修饰模式,从而诱导致癌性染色质重排,促进白血病的发生。

DNA-蛋白质相互作用在染色质功能调节方面起到非常重要的作用,主要通过组蛋白修饰<sup>[71-72]</sup>和转录因子结合改变<sup>[73]</sup>等方式实现,描绘肿瘤的这些表观遗传变化将进一步阐明肿瘤发生发展机制,为肿瘤药物开发提供新的靶点。

### 1.4 染色质可及性

染色质可及性是 DNA 可被启动子、增强子、绝缘子等顺式调控元件和转录因子、结构蛋白等蛋白质物理接近的性质。尽管可接近 DNA 仅占总 DNA 序列的 2%~3%,但却包含 90% 以上的转录结合区域<sup>[74]</sup>。染色质的开放与关闭处于一个动态平衡的状态,其异常改变可能在肿瘤发生发展中发挥重要作用。

#### 1.4.1 ATAC-Seq

ATAC-Seq 主要原理<sup>[16, 74]</sup>是用预装接头的 Tn5 转座酶捕获开放染色质(Open chromatin)区域, Tn5 可实现对基因组 DNA 的切割并将标签插入两端,闭合染色质(Closed chromatin)由于其结构致密而无法被切割,以此完成了对可及染色质序列的富集,随后对这些 DNA 序列进行高通量测序即可描绘出全基因组的染色质可及性图谱。ATAC-Seq 操作简便,所需细胞量较少,但 Tn5 转座酶的价格较为昂贵,结果存在酶切割的序列偏好问题<sup>[75]</sup>。

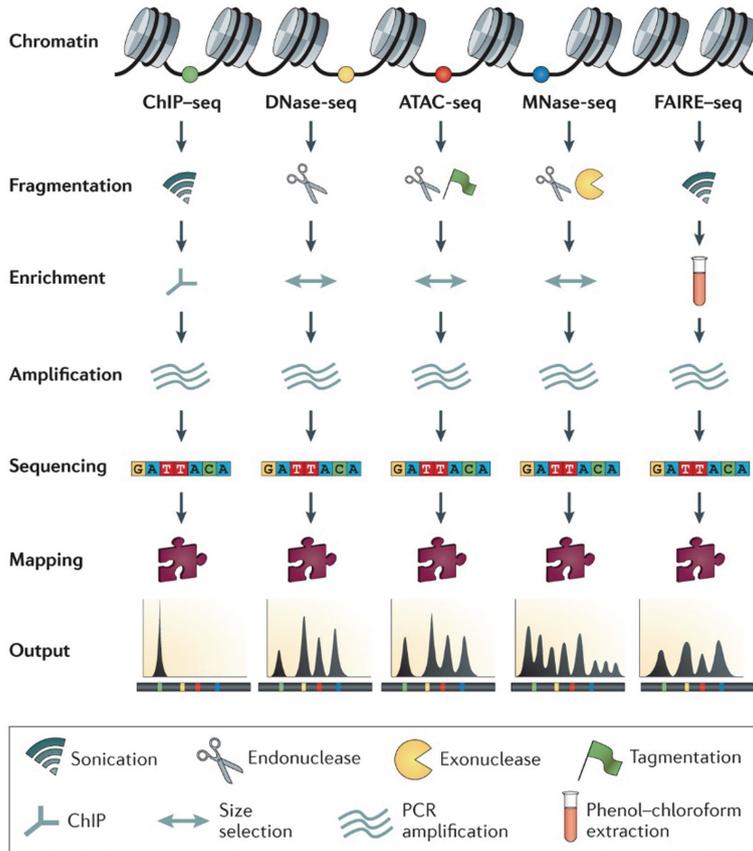
#### 1.4.2 MNase/DNase-Seq

MNase 具有核酸内切酶和外切酶的特性,优先靶向核小体之间的连接 DNA(Linker DNA),使用 MNase 消化基因组 DNA 可实现对核小体序列的富

集<sup>[17-18, 74]</sup>, MNase-Seq 主要用于核小体 DNA 的研究。DNase-Seq 是使用 DNase I 切割开放染色质区域超敏位点,结合组蛋白或转录因子区域受到蛋白质的保护作用不会被切割,纯化切割后的 DNA 加接头序列,可进行高通量测序<sup>[19, 74, 76]</sup>。MNase/DNase-Seq 原理简单,所需细胞量较多,且 DNase I 切割 DNA 时具有一定的偏好性,其分析的可靠性受到了一定的质疑,但 MNase/DNase-Seq 的结果,可间接推测出核小体以及转录因子的结合位点<sup>[75]</sup>。

### 1.4.3 FAIRE-Seq

FAIRE-Seq 的主要原理是缠绕组蛋白的核小体 DNA 与游离 DNA 在苯酚-氯仿中的溶解度不同。先用甲醛固定染色质后,再经超声打断,核小体 DNA 主要分布在有机相和水相之间,而游离 DNA 则分布于水相中,富集纯化游离 DNA 建库测序,以此完成对开放染色质的分析<sup>[20, 77]</sup>。FAIRE-Seq 克服了 MNase 和 DNase I 的序列偏好性<sup>[75]</sup>,但仍存在信噪比低的问题。各技术比较如图 2 所示。



Nature Reviews | Genetics

图 2 ChIP-Seq, DNase-Seq, ATAC-seq, MNase-Seq 和 FAIRE-Seq 技术比较<sup>[75]</sup>

Fig.2 An overview of ChIP-seq, DNase-seq, ATAC-seq, MNase-seq and FAIRE-seq experiments<sup>[75]</sup>

### 1.4.4 应用

目前关于肿瘤染色质可及性最大规模的研究来自于斯坦福大学 Howard Y Chang 教授和 William J. Greenleaf 教授研究团队<sup>[78]</sup>。作为 ATAC-Seq 的两位主要开发者,他们利用该技术绘制了来自 TCGA 样本库中的 410 例肿瘤样本涉及 23 种不同癌症类型的全基因组染色质可及性图谱,他们鉴定了 562 709 个可重复泛癌染色质可及性峰,其中约 65% 的泛癌峰与过往研究中发现的调控元件一致,表明 ATAC-Seq 不仅能很好地重复过去的研究,同时还能发现大量新的开放染色质位点。染色质开放/可及性是活性 DNA 的标志,通过结合基因组突变图谱,化学修饰图谱,转录谱等综合分析,可进一步描绘肿

瘤基因调控网络,推动肿瘤基础机制的研究和肿瘤精准治疗。

### 1.5 DNA-DNA 相互作用

Job Dekker 于 2002 年研发了染色体构象捕获 (3C) 技术<sup>[21]</sup>, 3D-基因组的研究自此拉开了序幕。随着后来基于 3C 的技术及测序方法不断开发与应用,染色质三维结构和染色质互作也逐渐被人们解析。基于 3C 的技术本质上检测的是 DNA-DNA 相互作用,如备受关注的增强子-启动子环 (Enhancer-Promotor loop, E-P loop)。

#### 1.5.1 HiC

HiC 不同于 3C, 4C 或 5C 的主要是因为 3C 是检测一对一 (One vs one) 的互作, 4C 是一对

多 (One vs all), 5C 是多对多 (Many vs many), 而 HiC 理论上可以检测所有 (All vs all) 的 DNA 互作<sup>[21-25, 79-81]</sup>。这些技术之所以被称为基于 3C 的技术, 是因为有着共同的处理步骤, 主要包括甲醛固定、内切酶消化、连接、DNA 纯化等步骤, 如图 3 所

示。HiC 不同的是首先用生物素标记的脱氧腺苷 (Biotin labeled dATP) 填补了酶切末端, 再连接成环, 解交联后用超声将 DNA 打断, 最后富集生物素标记的 DNA 片段建库测序, 以此实现全基因组 DNA-DNA 相互作用的检测。

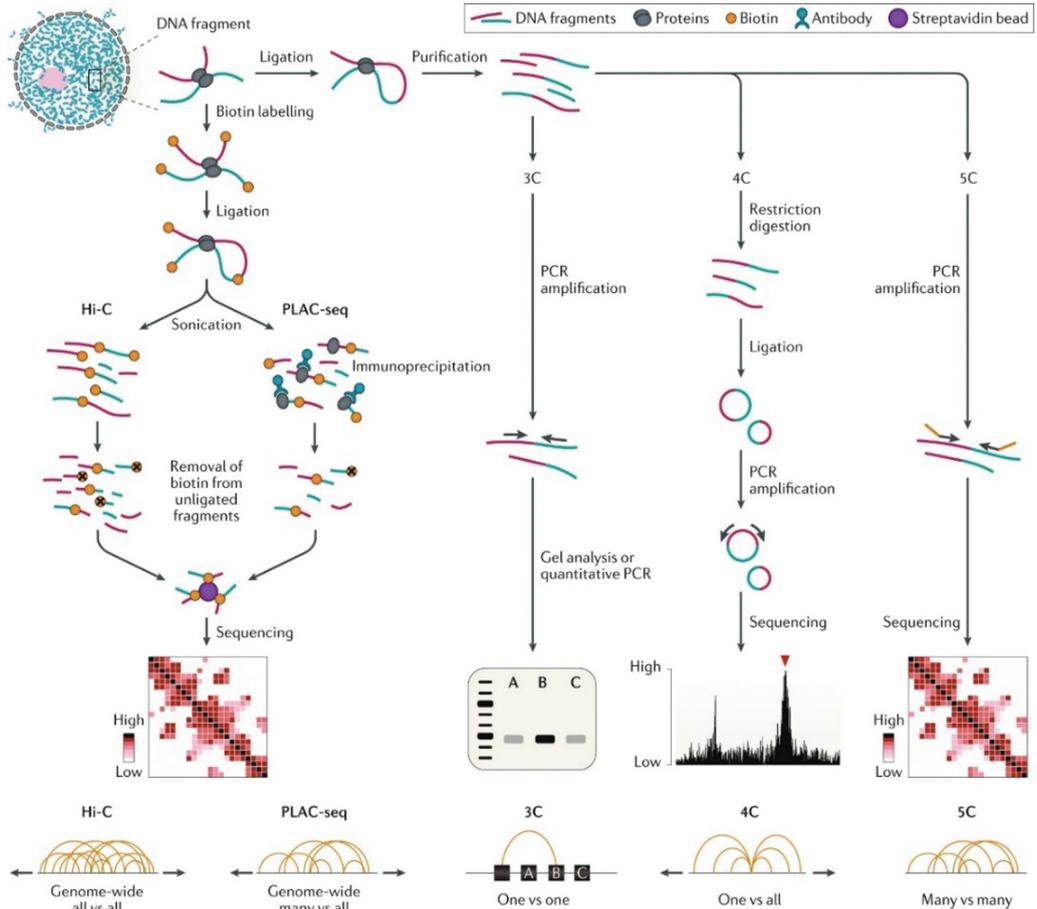


图 3 基于 3C 的染色质构象捕获技术<sup>[80]</sup>

Fig.3 3C-based chromosome conformation capture assays<sup>[80]</sup>

### 1.5.2 HiChIP/ChIA-PET

HiChIP 和 ChIA-PET 检出的 DNA 互作集合均是 HiC 的子集, 主要用于分析特定蛋白相关的 DNA 互作。HiChIP 是 HiC 和 ChIP 的结合, 固定、酶切、连接后先用抗体富集靶蛋白-DNA 复合物, 随后再针对这些 DNA 建库测序<sup>[26]</sup>。ChIA-PET 的原理略有不同, 是 ChIP 与 PET (Paired-end Tag sequencing) 的结合, 甲醛固定细胞后, 超声打断 DNA, 随后特异抗体富集靶蛋白-DNA, 接下来在 DNA 末端添加包含标签以及 Mme I 酶切位点生物素标记的 linker DNA, 连接后用 Mme I 消化并富集含生物素片段, 即可测序。传统 ChIA-PET 只能检测  $2 \times 20$  bp 的短序列, 获得的信息量较少, 近几年, 有研究人员通过引入 Tn5 转座酶将检出序列扩大到  $2 \times 250$  bp, 提高了 DNA 检出效率, 同时也提高了覆盖 SNP 的可能<sup>[82]</sup>。

### 1.5.3 应用

由于价格昂贵、技术复杂以及常常需要结合多种测序方法综合分析, 基于 HiC 等技术的肿瘤染色质三维结构研究数量并不太多。Aristotelis 等<sup>[83]</sup>通过对急性 T 淋巴细胞白血病原代细胞、细胞系以及正常 T 细胞进行比较, 发现 T-ALL 与 T 细胞相比, 其 TAD 内部的交互以及 TAD 边界的绝缘效应存在差异, 而 CTCF 介导的 TAD 融合可能促进 MYC 启动子-超级增强子环的形成, 从而促进肿瘤的形成。P. Andrew<sup>[84]</sup>通过对结肠癌、黑色素瘤、食管癌细胞系进行 HiC 建库, 结合已发表的肿瘤 TAD 边界的数据并注释, 再整合覆盖 42 种肿瘤类型超 3 000 例肿瘤-正常配对样本的 WGS 数据, 发现肿瘤体细胞突变的分布受染色质三维结构的影响, TAD 边界的突变分布率 (20.6%) 高于 TAD 内部 (6.9%)。Bradley 等<sup>[85]</sup>通过对结直肠癌临床

样本以及细胞系样本进行了 HiC 建库,鉴定出超 25 000 个 E-P 环与肿瘤转录相关,同时提出在结直肠癌中除了 A compartment 和 B compartment 外还存在一种 intermediate compartment 的中间状态,极大地丰富了染色质三维结构理论模型。

同时,近年来研究人员发现染色质三维结构与结构变异和非编码序列改变密切相关,染色质的重排影响顺式调控元件 (Cis-regulatory elements, CREs) 从而激活癌基因的表达<sup>[86-88]</sup>。通过整合分析 WGS、ChIP-seq 和 HiC 等数据<sup>[89]</sup>,可描绘出这种改变。笔者团队近期同样致力于胆道系统肿瘤和胰腺癌染色质三维结构的研究,有望描绘出上述肿瘤的染色质三维结构图谱。

### 1.6 DNA-RNA 相互作用

#### 1.6.1 GRID-Seq

自 3-D 基因组研究开展以来,DNA-DNA 互作和 RNA-RNA 互作检测技术<sup>[90-91]</sup>相继得到了突破,但 DNA-RNA 互作检测技术迟迟未能突破。2017 年,付向东教授团队研发出 GRID-Seq 可实现全基因组 DNA-RNA 互作的研究<sup>[28, 92]</sup>。其核心是设计了一段生物素标记的包含 Mme I 酶切位点的特殊的 DNA-RNA linker,基本流程如图 4 所示。首先用羟琥珀酰亚胺戊二酸 (DSG) 和甲醛双固定细胞内 DNA-RNA 互作,然后用 Alu I 消化基因组 DNA,随后 linker 与 RNA 结合,再经连接和复制延长即可捕获 DNA-RNA 相互作用,最后经 Mme I 消化后,对约 85 bp 长的目的序列进行建库测序。

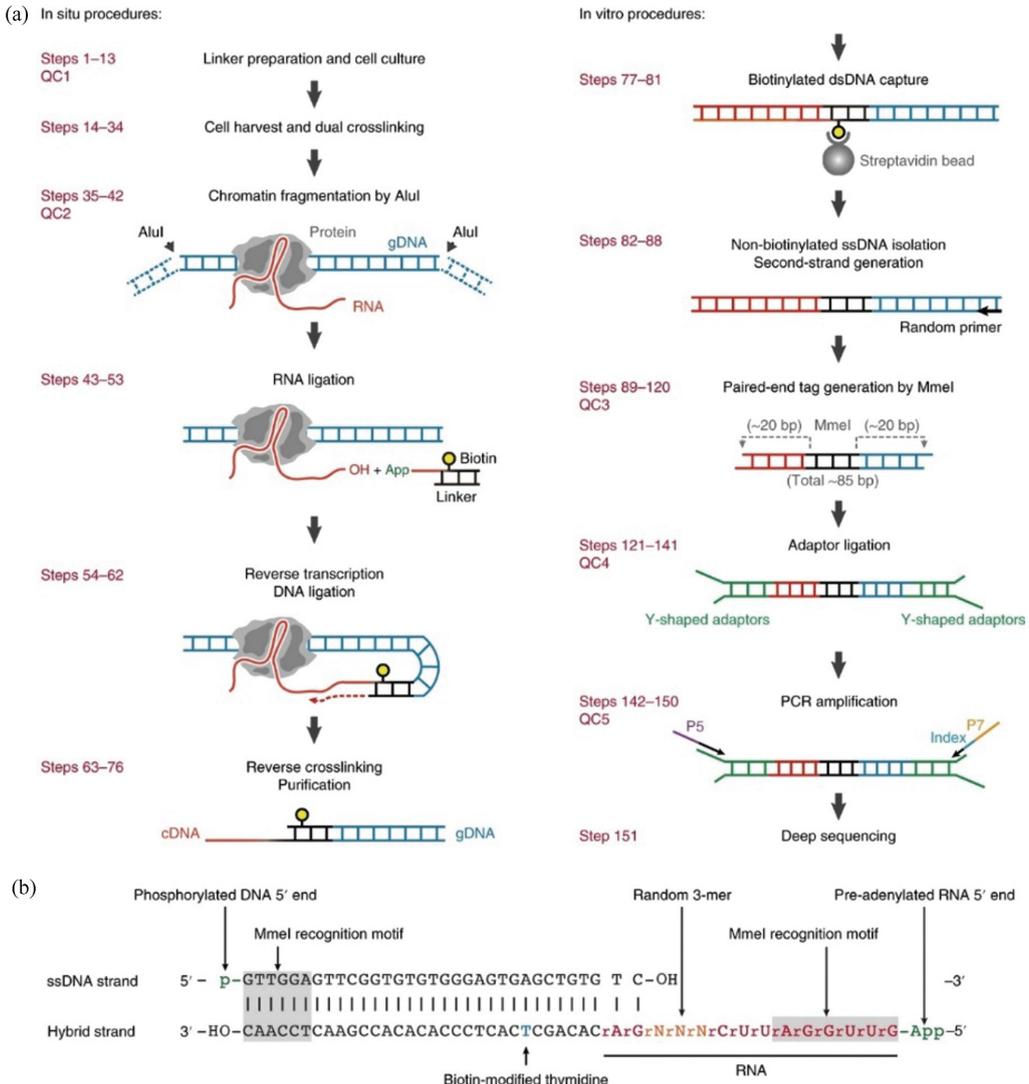


图 4 GRID-Seq 基本流程<sup>[92]</sup>

Fig.4 Schematics of the GRID-seq experiment<sup>[92]</sup>

### 1.6.2 iMARGI

iMARGI 基本流程与 GRID-Seq 类似,同样需要一段特殊的 linker 捕获染色质上的 RNA。与 GRID-Seq 不同的是,iMARGI 中 linker 的酶切位点为 BamH I,且需要环化捕获的 DNA-RNA 序列,经 BamH I 消化后,目的序列位于中间,两端为标签序列<sup>[29]</sup>。

### 1.6.3 应用

目前,iMARGI 还未在肿瘤研究中得到应用,而 GRID-Seq 除了付向东教授以三阴性乳腺癌细胞系 MDA-MB-231 为模型建立 DNA-RNA 互作图谱<sup>[28]</sup>外,其他类型肿瘤研究中也还未得到应用。他们在该细胞系中共检出 868 条高度富集的 mRNA 和 72 条 ncRNA,可能在转录调控、稳定染色质构象等方面起到重要作用。DNA-RNA 互作在肿瘤发生发展中的机制还需要以上技术在多种肿瘤类型中得到更

多应用才能逐渐清晰。

## 2 国内研究成果

得益于染色质测序技术的飞速发展以及成本的降低,国内众多肿瘤研究专家和团队对各类肿瘤展开了系统研究,由于成果众多,不胜枚举,笔者在此仅简单总结近十年国内团队利用临床样本研究的部分代表性成果(表 1)。由于肿瘤组织的异质性和复杂性,ChIP-Seq、和 HiC 等技术还未在临床样本中得到广泛应用,多数研究目前仍使用细胞系作为研究对象。若能够结合稳定的细胞分选(流式分选或磁珠分选)或类器官培养技术,使得样本的细胞类型尽可能均一,上述技术还有很大的应用空间。

表 1 国内利用染色质测序技术在肿瘤研究中的成果

Table 1 Achievements in cancer research using NGS-based chromatin sequencing in China

肿瘤类型	样本量		染色质测序方法	发表年份	团队
	肿瘤	对照			
非小细胞肺癌	69	69	WES	2021	张力 <sup>[93]</sup>
非小细胞肺癌	2 984	3 020	WGS	2022	沈洪兵 <sup>[94]</sup>
肝细胞肝癌	10	10	WES/WGS	2016	张宁 <sup>[95]</sup>
肝细胞肝癌	34	6	WES	2019	匡铭 <sup>[96]</sup>
肝细胞肝癌	56	56	WGS	2022	郭林泉 <sup>[97]</sup>
肝细胞肝癌	40	40	WGS	2022	周俭 <sup>[98]</sup>
肝内胆管癌	6	6	WES	2018	高强 <sup>[99]</sup>
肝内胆管癌	262	262	WES	2022	樊嘉 <sup>[100]</sup>
食管鳞状细胞癌	158	158(blood)	WES/WGS	2014	詹启敏 <sup>[101]</sup>
食管鳞状细胞癌	113	113	Exome capture	2014	赫捷 <sup>[102]</sup>
食管鳞状细胞癌	104	104	WGS/WES	2015	崔永萍 <sup>[103]</sup>
食管鳞状细胞癌	508	508	WGS	2020	刘芝华 <sup>[104]</sup>
胆囊癌	57	57	WES	2014	刘颖斌 <sup>[42]</sup>
胆囊癌	157	157	WES	2019	刘颖斌 <sup>[45]</sup>
胆囊神经内分泌癌	15	15	WES	2021	刘颖斌 <sup>[47]</sup>
胰腺神经内分泌肿瘤	211	211	WGS/WES	2020	吴文铭 <sup>[105]</sup>
NK/T 细胞淋巴瘤	25	25(blood)	WES	2015	陈赛娟 <sup>[106]</sup>
NK 细胞白血病	39	4	WGS	2018	王前飞 <sup>[107]</sup>
NK/T 细胞淋巴瘤	128	-	WGS/WES	2020	赵维莅 <sup>[108]</sup>
弥漫大 B 细胞淋巴瘤	619	-	WES/WGS	2021	赵维莅 <sup>[109]</sup>
膀胱癌	99	99(blood)	WGS/WES	2013	蔡志明 <sup>[110]</sup>
前列腺癌	210	210	WGS	2018	孙颖浩 <sup>[111]</sup>
宫颈癌	104	-	WGS/HIVID	2015	马丁 <sup>[112]</sup>
颅咽管瘤	26	26(blood)	WGS	2021	彭勇 <sup>[113]</sup>
乳腺癌	210	210(blood)	WGBS	2021	苏建忠 <sup>[59]</sup>
胃癌	12	12	WGBS	2017	储海燕 <sup>[114]</sup>
肺癌,肺转移性肠癌	62	62	Targeted BS	2022	王洁 <sup>[115]</sup>
尿路上皮癌,膀胱癌	45	4	WGBS	2022	李学松 <sup>[116]</sup>
结直肠癌	73	73	ChIP-Seq	2021	吴旻 <sup>[69]</sup>
三阴性乳腺癌	22	22	scATAC-Seq	2021	刘芝华 <sup>[117]</sup>
复发性膀胱癌	13	13	scATAC-Seq	2021	贾广帅 <sup>[118]</sup>
T 淋巴细胞白血病	18	4	BL-Hi-C <sup>[119]</sup>	2021	吴虹 <sup>[120]</sup>

### 3 其他染色质测序技术

值得注意的是,上述的大多测序技术在肿瘤单细胞研究<sup>[121-124]</sup>中已经得到了应用,推动了肿瘤异质性研究的快速发展。笔者也认为未来测序技术的发展趋势是精细化,单细胞化。随着单细胞分离技术和基因编辑技术的日渐成熟,CRISPR screening 和单细胞测序的结合也孕育而生,目前已在转录组<sup>[125-127]</sup>和染色质可及性<sup>[128]</sup>方面得到了部分应用,CRISPR/Cas9 基因编辑系统功能强大,与高通量测序有机结合潜力巨大。同时也应注意,尽管二代测序技术极大程度地推动了肿瘤研究的进展,但由于其读长较短(几百 bp),在寻找串联重复(Tandem duplication)和大范围结构变异劣势较为明显<sup>[129]</sup>。近年来,基于 Oxford nanopore 和 PacBio SMRT 的单分子三代测序(Third generation sequencing, TGS)技术迅速发展,克服了二代测序读长短且需 PCR 扩增的缺点,可直接通过核酸分子读出长达数千 bp 的序列以及 DNA/RNA 的修饰信息,在寻找串联重复、结构变异和遗传信息从头组装等方面具有强大优势,但三代测序目前仍存在错误率较高、成本高昂和分析方法有限等问题<sup>[130-132]</sup>。由于本文主要探讨基于 NGS 的染色质测序技术,同时受于篇幅所限,单细胞染色质测序和三代测序技术的原理及应用在此不详细展开。

### 4 总结与展望

染色质不仅储存着遗传信息,同时还有着精密的高级结构。针对肿瘤染色质分子生物学的不同方面,研究人员利用不同的基于 NGS 的染色质测序技术可对其进行研究,进而发现肿瘤染色质结构和功能的变化,并据此阐述肿瘤发生发展的基础机制,推动肿瘤的分型、靶向治疗和精准治疗。

### 参考文献(References)

[1] FIERZ B, POIRIER M G. Biophysics of chromatin dynamics [J]. *Annual Review of Biophysics*, 2019, 48: 321-45. DOI:10.1146/annurev-biophys-070317-032847.

[2] 李国强,刘法涛,刘颖斌.染色质三维结构变化在肿瘤进展中的作用[J].*中华外科杂志*, 2020, 58(12):973-976. DOI:10.3760/cma.j.cn112139-20200319-00240.

LI Guoqiang, LIU Fatao, LIU Yingbin. Role of three-dimensional chromatin structure changes in tumor progression [J]. *Chinese Journal of Surgery*, 2020, 58(12): 973-976. DOI: 10.3760/cma.j.cn112139-20200319-00240.

[3] TESSARZ P, KOUZARIDES T. Histone core modifications regulating nucleosome structure and dynamics [J]. *Nature Reviews Molecular Cell Biology*, 2014, 15(11): 703-708. DOI:10.1038/nrm3890.

[4] MISTELI T. The self-organizing genome: Principles of genome architecture and function [J]. *Cell*, 2020, 183(1): 28-45. DOI:10.1016/j.cell.2020.09.014.

[5] BENTLEY D R, BALASUBRAMANIAN S, SWERDLOW H P, et al. Accurate whole human genome sequencing using reversible terminator chemistry [J]. *Nature*, 2008, 456(7218): 53-59. DOI:10.1038/nature07517.

[6] NG S B, TURNER E H, ROBERTSON P D, et al. Targeted capture and massively parallel sequencing of 12 human exomes [J]. *Nature*, 2009, 461(7261): 272-276. DOI:10.1038/nature08250.

[7] MEISSNER A, MIKKELSEN T S, GU H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells [J]. *Nature*, 2008, 454(7205): 766-770. DOI:10.1038/nature07107.

[8] LISTER R, O'MALLEY R C, TONTI-FILIPPINI J, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis [J]. *Cell*, 2008, 133(3): 523-536. DOI:10.1016/j.cell.2008.03.029.

[9] LISTER R, PELIZZOLA M, DOWEN R H, et al. Human DNA methylomes at base resolution show widespread epigenomic differences [J]. *Nature*, 2009, 462(7271): 315-322. DOI:10.1038/nature08514.

[10] BOOTH M J, BRANCO M R, FICZ G, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution [J]. *Science (New York, NY)*, 2012, 336(6083): 934-937. DOI: 10.1126/science.1220671.

[11] BARSKI A, CUDDAPAH S, CUI K, et al. High-resolution profiling of histone methylations in the human genome [J]. *Cell*, 2007, 129(4): 823-837. DOI: 10.1016/j.cell.2007.05.009.

[12] JOHNSON D S, MORTAZAVI A, MYERS R M, et al. Genome-wide mapping of in vivo protein-DNA interactions [J]. *Science (New York, NY)*, 2007, 316(5830): 1497-1502. DOI: 10.1126/science.1141319.

[13] MIKKELSEN T S, KU M, JAFFE D B, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells [J]. *Nature*, 2007, 448(7153): 553-560. DOI: 10.1038/nature06008.

[14] ZENTNER G E, KASINATHAN S, XIN B, et al. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo [J]. *Nature Communications*, 2015, 6: 8733. DOI:10.1038/ncomms9733.

[15] KAYA-OKUR H S, WU S J, CODOMO C A, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells [J]. *Nature Communications*, 2019, 10(1):1930. DOI:10.1038/s41467-019-09982-5.

[16] BUENROSTRO J D, GIRESI P G, ZABA L C, et al.

- Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position [J]. *Nature Methods*, 2013, 10(12): 1213–1218. DOI:10.1038/nmeth.2688.
- [17] JOHNSON S M, TAN F J, MCCULLOUGH H L, et al. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin[J]. *Genome Research*, 2006, 16(12): 1505–1516.
- [18] SCHONES D E, CUI K, CUDDAPAH S, et al. Dynamic regulation of nucleosome positioning in the human genome [J]. *Cell*, 2008, 132(5): 887–898. DOI:10.1016/j.cell.2008.02.022.
- [19] BOYLE A P, DAVIS S, SHULHA H P, et al. High-resolution mapping and characterization of open chromatin across the genome[J]. *Cell*, 2008, 132(2): 311–322. DOI:10.1016/j.cell.2007.12.014.
- [20] GIRESI P G, KIM J, MCDANIELL R M, et al. FAIRE (Formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin [J]. *Genome Research*, 2007, 17(6): 877–885. DOI: 10.1101/gr.5533506.
- [21] DEKKER J, RIPPE K, DEKKER M, et al. Capturing chromosome conformation[J]. *Science (New York, NY)*, 2002, 295(5558): 1306–1311. DOI: 10.1126/science.1067799.
- [22] ZHAO Z, TAVOOSIDANA G, SJÖLINDER M, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions [J]. *Nature Genetics*, 2006, 38(11): 1341–1347. DOI: 10.1038/ng1891.
- [23] SIMONIS M, KLOUS P, SPLINTER E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C) [J]. *Nature Genetics*, 2006, 38(11): 1348–1354. DOI: 10.1038/ng1896.
- [24] DOSTIE J, RICHMOND T A, ARNAOUT R A, et al. Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements[J]. *Genome Research*, 2006, 16(10): 1299–1309. DOI: 10.1101/gr.5571506.
- [25] LIEBERMAN-AIDEN E, VAN BERKUM N L, WILLIAMS L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome [J]. *Science (New York, NY)*, 2009, 326(5950): 289–293. DOI:10.1126/science.1181369.
- [26] MUMBACH M R, RUBIN A J, FLYNN R A, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture[J]. *Nature Methods*, 2016, 13(11): 919–922. DOI:10.1038/nmeth.3999.
- [27] LI G, FULLWOOD M J, XU H, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing[J]. *Genome Biology*, 2010, 11(2): R22. DOI:10.1186/gb-2010-11-2-r22.
- [28] LI Xiao, ZHOU Bing, CHEN Liang, et al. GRID-seq reveals the global RNA-chromatin interactome [J]. *Nature Biotechnology*, 2017, 35(10): 940–950. DOI:10.1038/nbt.3968.
- [29] WU Weixin, YAN Zhangming, NGUYEN T C, et al. Mapping RNA-chromatin interactions by sequencing with iMARGI[J]. *Nature Protocols*, 2019, 14(11): 3243–3272. DOI:10.1038/s41596-019-0229-4.
- [30] YOSHINAGA Y, DAUM C, HE G, et al. Genome sequencing[J]. *Methods In Molecular Biology (Clifton, NJ)*, 2018, 1775: 37–52. DOI:10.1007/978-1-4939-7804-5\_4.
- [31] ULINTZ P J, WU W, GATES C M. Bioinformatics analysis of whole exome sequencing data[J]. *Methods In Molecular Biology (Clifton, NJ)*, 2019, 1881: 277–318. DOI:10.1007/978-1-4939-8876-1\_21.
- [32] KOSUGI S, MOMOZAWA Y, LIU X, et al. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing [J]. *Genome Biology*, 2019, 20(1): 117. DOI:10.1186/s13059-019-1720-5.
- [33] CHOI M, SCHOLL U I, JI W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(45): 19096–19101. DOI:10.1073/pnas.0910672106.
- [34] NAKAGAWA H, FUJITA M. Whole genome sequencing analysis for cancer genomics and precision medicine [J]. *Cancer Science*, 2018, 109(3): 513–522. DOI:10.1111/cas.13505.
- [35] Cancer Genome Atlas Research Network. The cancer genome atlas [DB/OL]. (2008-09-04) [2020-12-01]. <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>.
- [36] HUDSON T J, ANDERSON W, ARTEZ A, et al. International network of cancer genome projects [J]. *Nature*, 2010, 464(7291): 993–998. DOI:10.1038/nature08987.
- [37] TATE J G, BAMFORD S, JUBB H C, et al. COSMIC: the catalogue of somatic mutations in cancer [J]. *Nucleic Acids Research*, 2019, 47(D1): D941–D947. DOI: 10.1093/nar/gky1015.
- [38] STEPHENS P J, TARPEY P S, DAVIES H, et al. The landscape of cancer genes and mutational processes in breast cancer [J]. *Nature*, 2012, 486(7403): 400–404. DOI:10.1038/nature11017.
- [39] PEREIRA B, CHIN S-F, RUEDA O M, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes [J]. *Nature Communications*, 2016, 7: 11479. DOI:10.1038/ncomms11479.
- [40] NIK-ZAINAL S, DAVIES H, STAAF J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences [J]. *Nature*, 2016, 534(7605): 47–54. DOI:10.1038/nature17676.
- [41] WADDELL N, PAJIC M, PATCH A M, et al. Whole ge-

- nomes redefine the mutational landscape of pancreatic cancer [J]. *Nature*, 2015, 518(7540): 495–501. DOI: 10.1038/nature14169.
- [42] LI Maolan, ZHANG Zhou, LI Xiaoguang, et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway [J]. *Nature Genetics*, 2014, 46(8): 872–876. DOI:10.1038/ng.3030.
- [43] PANDEY A, STAWISKI E W, DURINCK S, et al. Integrated genomic analysis reveals mutated ELF3 as a potential gallbladder cancer vaccine candidate [J]. *Nature Communications*, 2020, 11(1): 4225. DOI:10.1038/s41467-020-17880-4.
- [44] GENG Yajun, CHEN Shili, YANG Yang, et al. Long-term exposure to genistein inhibits the proliferation of gallbladder cancer by downregulating the MCM complex [J]. *Science Bulletin*, 2022, 67(8): 813–824. DOI: 10.1016/j.scib.2022.01.011.
- [45] LI Maolan, LIU Fatao, ZHANG Fei, et al. Genomic ERBB2/ERBB3 mutations promote PD-L1-mediated immune escape in gallbladder cancer; A whole-exome sequencing analysis [J]. *Gut*, 2019, 68(6): 1024–1033. DOI: 10.1136/gutjnl-2018-316039.
- [46] ZHANG Yijian, ZUO Chunman, LIU Liguang, et al. Single-cell RNA-sequencing atlas reveals an MDK-dependent immunosuppressive environment in ErbB pathway-mutated gallbladder cancer [J]. *Journal of Hepatology*, 2021, 75(5): 1128–1141. DOI:10.1016/j.jhep.2021.06.023.
- [47] LIU Fatao, LI Yongsheng, YING Dongjian, et al. Whole-exome mutational landscape of neuroendocrine carcinomas of the gallbladder [J]. *Signal Transduction and Targeted Therapy*, 2021, 6(1): 55. DOI: 10.1038/s41392-020-00412-3.
- [48] NEGRINI S, GORGOULIS V G, HALAZONETIS T D. Genomic instability—an evolving hallmark of cancer [J]. *Nature Reviews Molecular Cell Biology*, 2010, 11(3): 220–228. DOI:10.1038/nrm2858.
- [49] HUANG A, GARRAWAY L A, ASHWORTH A, et al. Synthetic lethality as an engine for cancer drug target discovery [J]. *Nature Reviews Drug Discovery*, 2020, 19(1): 23–38. DOI:10.1038/s41573-019-0046-z.
- [50] ZHURAVLEVA E, O'Rourke C J, ANDERSEN J B. Mutational signatures and processes in hepatobiliary cancers [J]. *Nature Reviews Gastroenterology & Hepatology*, 2022, 19(6): 367–382. DOI: 10.1038/s41575-022-00587-w.
- [51] RAIBER E A, HARDISTY R, VAN DELFT P, et al. Mapping and elucidating the function of modified bases in DNA [J]. *Nature Reviews Chemistry*, 2017, 1(9): 0069. DOI: 10.1038/s41570-017-0069.
- [52] JONES P A. Functions of DNA methylation: islands, start sites, gene bodies and beyond [J]. *Nature Reviews Genetics*, 2012, 13(7): 484–492. DOI:10.1038/nrg3230.
- [53] ZHAO Linyong, SONG Jinghui, LIU Yibin, et al. Mapping the epigenetic modifications of DNA and RNA [J]. *Protein & Cell*, 2020, 11(11): 792–808. DOI:10.1007/s13238-020-00733-7.
- [54] WU Xiaoji, ZHANG Yi. TET-mediated active DNA demethylation: Mechanism, function and beyond [J]. *Nature Reviews Genetics*, 2017, 18(9): 517–534. DOI: 10.1038/nrg.2017.33.
- [55] BOOTH M J, OST T W B, BERARDI D, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine [J]. *Nature Protocols*, 2013, 8(10): 1841–1851. DOI:10.1038/nprot.2013.115.
- [56] DAS P M, SINGAL R. DNA methylation and cancer [J]. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 2004, 22(22): 4632–4642. DOI: 10.1200/JCO.2004.07.151.
- [57] YU Y P, DING Ying, CHEN Rui, et al. Whole-genome methylation sequencing reveals distinct impact of differential methylations on gene transcription in prostate cancer [J]. *The American Journal of Pathology*, 2013, 183(6): 1960–1970. DOI:10.1016/j.ajpath.2013.08.018.
- [58] ZHAO S G, CHEN W S, LI Haolong, et al. The DNA methylation landscape of advanced prostate cancer [J]. *Nature Genetics*, 2020, 52(8): 778–789. DOI: 10.1038/s41588-020-0648-8.
- [59] LIU Jiaqi, ZHAO Hengqiang, HUANG Yukuan, et al. Genome-wide cell-free DNA methylation analyses improve accuracy of non-invasive diagnostic imaging for early-stage breast cancer [J]. *Molecular Cancer*, 2021, 20(1): 36. DOI:10.1186/s12943-021-01330-w.
- [60] BERMAN B P, WEISENBERGER D J, AMAN J F, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains [J]. *Nature Genetics*, 2011, 44(1): 40–46. DOI:10.1038/ng.969.
- [61] KOCH A, JOOSTEN S C, FENG Z, et al. Analysis of DNA methylation in cancer; Location revisited [J]. *Nature Reviews Clinical Oncology*, 2018, 15(7): 459–466. DOI: 10.1038/s41571-018-0004-4.
- [62] PAPANICOLAOU-SENGOS A, ALDAPE K. DNA methylation profiling: an emerging paradigm for cancer diagnosis [J]. *Annual Review of Pathology*, 2022, 17: 295–321. DOI:10.1146/annurev-pathol-042220-022304.
- [63] ZHANG Yanjun, SUN Zhongxing, JIA Junqi, et al. Overview of Histone Modification [J]. *Advances In Experimental Medicine and Biology*, 2021, 1283: 1–16. DOI:10.1007/978-981-15-8104-5\_1.
- [64] FERRAZ R A C, LOPES A L G, DA SILVA J A F, et al. DNA-protein interaction studies: a historical and comparative analysis [J]. *Plant Methods*, 2021, 17(1): 82. DOI: 10.1186/s13007-021-00780-z.
- [65] YAMAKAWA A, HOJO H, OHBA S. ChIP-Seq assays from mammalian cartilage and chondrocytes [J]. *Methods In*

- Molecular Biology (Clifton, NJ), 2021, 2245: 167–178. DOI:10.1007/978-1-0716-1119-7\_12.
- [66] BARANELLO L, KOUZINE F, SANFORD S, et al. ChIP bias as a function of cross-linking time [J]. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 2016, 24(2): 175–181. DOI:10.1007/s10577-015-9509-1.
- [67] SCHMID M, DURUSSEL T, LAEMMLI U K. ChIC and ChEC; genomic mapping of chromatin proteins [J]. *Molecular Cell*, 2004, 16(1): 147–157. DOI: 10.1016/j.molcel.2004.09.007.
- [68] JIANG Yanyi, JIANG Yuan, LI Chunquan, et al. TP63, SOX2, and KLF5 establish a core regulatory circuitry that controls epigenetic and transcription patterns in esophageal squamous cell carcinoma cell lines [J]. *Gastroenterology*, 2020, 159(4): 1311–1327. e19. DOI: 10.1053/j.gastro.2020.06.050.
- [69] LI Qinglan, LIN Xiang, YU Yali, et al. Genome-wide profiling in colorectal cancer identifies PHF19 and TBC1D16 as oncogenic super enhancers [J]. *Nature Communications*, 2021, 12(1): 6407. DOI:10.1038/s41467-021-26600-5.
- [70] JANSSENS D H, MEERS M P, WU S J, et al. Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia [J]. *Nature Genetics*, 2021, 53(11): 1586–1596. DOI:10.1038/s41588-021-00941-9.
- [71] MILLÁN-ZAMBRANO G, BURTON A, BANNISTER A J, et al. Histone post-translational modifications – cause and consequence of genome function [J]. *Nature Reviews Genetics*, 2022, 23(9): 563–580. DOI: 10.1038/s41576-022-00468-7.
- [72] ZHU Dongwei, ZHANG Yue, WANG Shengjun. Histone citrullination: A new target for tumors [J]. *Molecular Cancer*, 2021, 20(1): 90. DOI:10.1186/s12943-021-01373-z.
- [73] BUSHWELLER J H. Targeting transcription factors in cancer—from undruggable to reality [J]. *Nature Reviews Cancer*, 2019, 19(11): 611–624. DOI: 10.1038/s41568-019-0196-7.
- [74] KLEMM S L, SHIPONY Z, GREENLEAF W J. Chromatin accessibility and the regulatory epigenome [J]. *Nature Reviews Genetics*, 2019, 20(4): 207–220. DOI:10.1038/s41576-018-0089-8.
- [75] MEYER C A, LIU X S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology [J]. *Nature Reviews Genetics*, 2014, 15(11): 709–721. DOI:10.1038/nrg3788.
- [76] SONG Lingyun, CRAWFORD G E. DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells [J]. *Cold Spring Harbor Protocols*, 2010, 2010(2): 5384. DOI:10.1101/pdb.prot5384.
- [77] GIRESI P G, LIEB J D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements) [J]. *Methods (San Diego, Calif)*, 2009, 48(3): 233–239. DOI: 10.1016/j.ymeth.2009.03.003.
- [78] CORCES M R, GRANJA J M, SHAMS S, et al. The chromatin accessibility landscape of primary human cancers [J]. *Science (New York, NY)*, 2018, 362: 6413. DOI: 10.1126/science.aav1898.
- [79] DE WIT E, DE LAAT W. A decade of 3C technologies: Insights into nuclear organization [J]. *Genes & Development*, 2012, 26(1): 11–24. DOI:10.1101/gad.179804.111.
- [80] KEMPFFER R, POMBO A. Methods for mapping 3D chromosome architecture [J]. *Nature Reviews Genetics*, 2020, 21(4): 207–226. DOI:10.1038/s41576-019-0195-2.
- [81] NOORDERMEER D, DUBOULE D. Chromatin looping and organization at developmentally regulated gene loci [J]. *Wiley Interdisciplinary Reviews Developmental Biology*, 2013, 2(5): 615–630. DOI:10.1002/wdev.103.
- [82] LI Xingwang, LUO O J, WANG Ping, et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions [J]. *Nature Protocols*, 2017, 12(5): 899–915. DOI:10.1038/nprot.2017.012.
- [83] KLOETGEN A, THANDAPANI P, NTZIACHRISTOS P, et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia [J]. *Nature Genetics*, 2020, 52(4): 388–400. DOI:10.1038/s41588-020-0602-9.
- [84] AKDEMIR K C, LE V T, KIM J M, et al. Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure [J]. *Nature Genetics*, 2020, 52(11): 1178–1188. DOI:10.1038/s41588-020-0708-0.
- [85] JOHNSTONE S E, REYES A, QI Y, et al. Large-scale topological changes restrain malignant progression in colorectal cancer [J]. *Cell*, 2020, 182(6): 1474–1489. e23. DOI: 10.1016/j.cell.2020.07.030.
- [86] SPIELMANN M, LUPIÁN EZ D G, MUNDLOS S. Structural variation in the 3D genome [J]. *Nature Reviews Genetics*, 2018, 19(7): 453–467. DOI: 10.1038/s41576-018-0007-0.
- [87] DUBOIS F, SIDIROPOULOS N, WEISCHENFELDT J, et al. Structural variations in cancer and the 3D genome [J]. *Nature Reviews Genetics*, 2022, 22(9): 533–546. DOI: 10.1038/s41568-022-00488-9.
- [88] KHURANA E, FU Y, CHAKRAVARTY D, et al. Role of non-coding sequence variants in cancer [J]. *Nature Reviews Genetics*, 2016, 17(2): 93–108. DOI:10.1038/nrg.2015.17.
- [89] WANG Xiaotao, XU Jie, ZHANG Baozhen, et al. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes [J]. *Nature Methods*, 2021, 18(6): 661–668. DOI:10.1038/s41592-021-01164-w.
- [90] KUDLA G, GRANNEMAN S, HAHN D, et al. Cross-link-

- ing, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(24): 10010–10015. DOI:10.1073/pnas.1017386108.
- [91] CAI Zhaokui, CAO Changchang, JI Lei, et al. RIC-seq for global in situ profiling of RNA-RNA spatial interactions [J]. *Nature*, 2020, 582(7812): 432–437. DOI:10.1038/s41586-020-2249-1.
- [92] ZHOU Bing, LI Xiao, LUO Daji, et al. GRID-seq for comprehensive analysis of global RNA-chromatin interactions [J]. *Nature Protocols*, 2019, 14(7): 2036–2068. DOI: 10.1038/s41596-019-0172-4.
- [93] FANG Wenfeng, JIN Haoxuan, ZHOU Haoqiang, et al. Intratumoral heterogeneity as a predictive biomarker in anti-PD-(L)1 therapies for non-small cell lung cancer[J]. *Molecular Cancer*, 2021, 20: 37. DOI:10.1186/s12943-021-01331-9.
- [94] WANG Cheng, DAI Juncheng, QIN Na, et al. Analyses of rare predisposing variants of lung cancer in 6,004 whole genomes in Chinese[J]. *Cancer Cell*, 2022, 40(10): 1223–1239.e6. DOI:10.1016/j.ccell.2022.08.013.
- [95] XUE Ruidong, LI Ruoyan, GUO Hua, et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma[J]. *Gastroenterology*, 2016, 150(4): 998–1008. DOI:10.1053/j.gastro.2015.12.033.
- [96] XU L X, HE M H, DAI Z H, et al. Genomic and transcriptional heterogeneity of multifocal hepatocellular carcinoma [J]. *Annals of Oncology: Official Journal of the European Society For Medical Oncology*, 2019, 30(6): 990–997. DOI:10.1093/annonc/mdz103.
- [97] LIU Weiwei, ZHENG Lu, ZHANG Rongguiyi, et al. Circ-ZEB1 promotes PIK3CA expression by silencing miR-199a-3p and affects the proliferation and apoptosis of hepatocellular carcinoma[J]. *Molecular Cancer*, 2022, 21: 72. DOI:10.1186/s12943-022-01529-5.
- [98] ZHOU Shaolai, ZHOU Zhengjun, SONG Chengli, et al. Whole-genome sequencing reveals the evolutionary trajectory of HBV-related hepatocellular carcinoma early recurrence [J]. *Signal Transduction and Targeted Therapy*, 2022, 7: 24. DOI:10.1038/s41392-021-00838-3.
- [99] DONG Liangqing, SHI Yang, MA Lijie, et al. Spatial and temporal clonal evolution of intrahepatic cholangiocarcinoma [J]. *Journal of Hepatology*, 2018, 69(1): 89–98. DOI: 10.1016/j.jhep.2018.02.029.
- [100] DONG Liangqing, LU Dayun, CHEN Ran, et al. Proteogenomic characterization identifies clinically relevant subgroups of intrahepatic cholangiocarcinoma [J]. *Cancer Cell*, 2022, 40(1): 70–87.e15. DOI: 10.1016/j.ccell.2021.12.006.
- [101] SONG Yongmei, LI Lin, OU Yunwei, et al. Identification of genomic alterations in oesophageal squamous cell cancer [J]. *Nature*, 2014, 509(7498): 91–95. DOI:10.1038/nature13176.
- [102] GAO Yibo, CHEN Zhaoli, LI Jiagen, et al. Genetic landscape of esophageal squamous cell carcinoma [J]. *Nature Genetics*, 2014, 46(10): 1097–1092. DOI:10.1038/ng.3076.
- [103] ZHANG Ling, ZHOU Yong, CHENG Caixia, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma [J]. *American Journal of Human Genetics*, 2015, 96(4): 597–611. DOI:10.1016/j.ajhg.2015.02.017.
- [104] CUI Yongping, CHEN Hongyan, XI Ruibin, et al. Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma [J]. *Cell Research*, 2020, 30(10): 902–913. DOI:10.1038/s41422-020-0333-6.
- [105] HONG Xiafei, QIAO Sitan, LI Fuqiang, et al. Whole-genome sequencing reveals distinct genetic bases for insulinomas and non-functional pancreatic neuroendocrine tumours: Leading to a new classification system [J]. *Gut*, 2020, 69(5): 877–887. DOI: 10.1136/gutjnl-2018-317233.
- [106] JIANG Lu, GU Zhaohui, YAN Zixun, et al. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma [J]. *Nature Genetics*, 2015, 47(9): 1061–1066. DOI:10.1038/ng.3358.
- [107] HUANG Liang, LIU Dan, WANG Na, et al. Integrated genomic analysis identifies deregulated JAK/STAT-MYC-biosynthesis axis in aggressive NK-cell leukemia [J]. *Cell Research*, 2018, 28(2): 172–186. DOI: 10.1038/cr.2017.146.
- [108] XIONG Jie, CUI Bowen, WANG Nan, et al. Genomic and transcriptomic characterization of natural killer T cell lymphoma [J]. *Cancer Cell*, 2020, 37(3): 403–419.e6. DOI:10.1016/j.ccell.2020.02.005.
- [109] HUANG Yaohui, CAI Kun, XU Pengpeng, et al. CREBBP/EP300 mutations promoted tumor progression in diffuse large B-cell lymphoma through altering tumor-associated macrophage polarization via FBXW7–NOTCH–CCL2/CSF1 axis [J]. *Signal Transduction and Targeted Therapy*, 2021, 6: 10. DOI: 10.1038/s41392-020-00437-8.
- [110] GUO Guangwu, SUN Xiaojuan, CHEN Chao, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation [J]. *Nature Genetics*, 2013, 45(12): 1459–1463. DOI: 10.1038/ng.2798.
- [111] REN Shancheng, WEI Gonghong, LIU Dongbing, et al. Whole-genome and transcriptome sequencing of prostate cancer identify new genetic alterations driving disease progression [J]. *European Urology*, 2018, 73(3): 322–339. DOI:10.1016/j.euro.2017.08.027.
- [112] HU Zheng, ZHU Da, WANG Wei, et al. Genome-wide

- profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism [J]. *Nature Genetics*, 2015, 47(2): 158–163. DOI:10.1038/ng.3178.
- [113] HE Juan, ZENG Zhen, WANG Yuelong, et al. Characterization of novel CTNNB1 mutation in Craniopharyngioma by whole-genome sequencing [J]. *Molecular Cancer*, 2021, 20(1): 168. DOI:10.1186/s12943-021-01468-7.
- [114] MA Gaoxiang, LIU Hanting, HUA Qiuhan, et al. KC-NMA1 cooperating with PTK2 is a novel tumor suppressor in gastric cancer and is associated with disease outcome [J]. *Molecular Cancer*, 2017, 16(1): 46. DOI:10.1186/s12943-017-0613-z.
- [115] ZUO Ying, ZHONG Jia, BAI Hua, et al. Genomic and epigenomic profiles distinguish pulmonary enteric adenocarcinoma from lung metastatic colorectal cancer [J]. *EBioMedicine*, 2022, 82: 104165. DOI:10.1016/j.ebiom.2022.104165.
- [116] LI Juan, LIANG Yuan, FAN Jian, et al. DNA methylation subtypes guiding prognostic assessment and linking to responses the DNA methyltransferase inhibitor SGI-110 in urothelial carcinoma [J]. *BMC Medicine*, 2022, 20(1): 222. DOI:10.1186/s12916-022-02426-w.
- [117] ZHANG Yuanyuan, CHEN Hongyan, MO Hongnan, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer [J]. *Cancer Cell*, 2021, 39(12): 1578–1593.e8. DOI:10.1016/j.ccell.2021.09.010.
- [118] WANG Huanjun, MEI Yan, LUO Cheng, et al. Single-cell analyses reveal mechanisms of cancer stem cell maintenance and epithelial-mesenchymal transition in recurrent bladder cancer [J]. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 2021, 27(22): 6265–6278. DOI:10.1158/1078-0432.CCR-20-4796.
- [119] LIANG Zhengyu, LI Guipeng, WANG Zejun, et al. BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions [J]. *Nature Communications*, 2017, 8(1): 1622. DOI:10.1038/s41467-017-01754-3.
- [120] YANG Lu, CHEN Fengling, ZHU Haichuan, et al. 3D genome alterations associated with dysregulated HOXA13 expression in high-risk T-lineage acute lymphoblastic leukemia [J]. *Nature Communications*, 2021, 12(1): 3708. DOI:10.1038/s41467-021-24044-5.
- [121] WU S J, FURLAN S N, MIHALAS A B, et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression [J]. *Nature Biotechnology*, 2021, 39(7): 819–824. DOI:10.1038/s41587-021-00865-z.
- [122] TAAVITSAINEN S, ENGEDAL N, CAO S, et al. Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse [J]. *Nature Communications*, 2021, 12(1): 5307. DOI:10.1038/s41467-021-25624-1.
- [123] HOU Yu, GUO Huahu, CAO Chen, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas [J]. *Cell Research*, 2016, 26(3): 304–319. DOI:10.1038/cr.2016.23.
- [124] GROSSELIN K, DURAND A, MARSOLIER J, et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer [J]. *Nature Genetics*, 2019, 51(6): 1060–1066. DOI:10.1038/s41588-019-0424-9.
- [125] DATLINGER P, RENDEIRO A F, SCHMIDL C, et al. Pooled CRISPR screening with single-cell transcriptome readout [J]. *Nature Methods*, 2017, 14(3): 297–301. DOI:10.1038/nmeth.4177.
- [126] ADAMSON B, NORMAN T M, JOST M, et al. A multiplexed single-cell crispr screening platform enables systematic dissection of the unfolded protein response [J]. *Cell*, 2016, 167(7): 1867–1882.e21. DOI:10.1016/j.cell.2016.11.048.
- [127] REPLOGLE J M, NORMAN T M, XU A, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing [J]. *Nature Biotechnology*, 2020, 38(8): 954–961. DOI:10.1038/s41587-020-0470-y.
- [128] RUBIN A J, PARKER K R, SATPATHY A T, et al. Coupled single-cell crispr screening and epigenomic profiling reveals causal gene regulatory networks [J]. *Cell*, 2019, 176(1–2): 361–376.e17. DOI:10.1016/j.cell.2018.11.022.
- [129] KUMAR K R, COWLEY M J, DAVIS R L. Next-generation sequencing and emerging technologies [J]. *Seminars In Thrombosis and Hemostasis*, 2019, 45(7): 661–673. DOI:10.1055/s-0039-1688446.
- [130] SEARLE B, MÜLLER M, CARELL T, et al. Third-generation sequencing of epigenetic DNA [J]. *Angewandte Chemie (International Ed In English)*, 2022, 62(14): e202215704. DOI:10.1002/anie.202215704.
- [131] JUNG H, WINEFIELD C, BOMBARELY A, et al. Tools and strategies for long-read sequencing and de novo assembly of plant genomes [J]. *Trends In Plant Science*, 2019, 24(8): 700–724. DOI:10.1016/j.tplants.2019.05.003.
- [132] PETERSEN L M, MARTIN I W, MOSCHETTI W E, et al. Third-generation sequencing in the clinical laboratory: Exploring the advantages and challenges of nanopore sequencing [J]. *Journal of Clinical Microbiology*, 2019, 58(1): e01315–19. DOI:10.1128/JCM.01315–19.