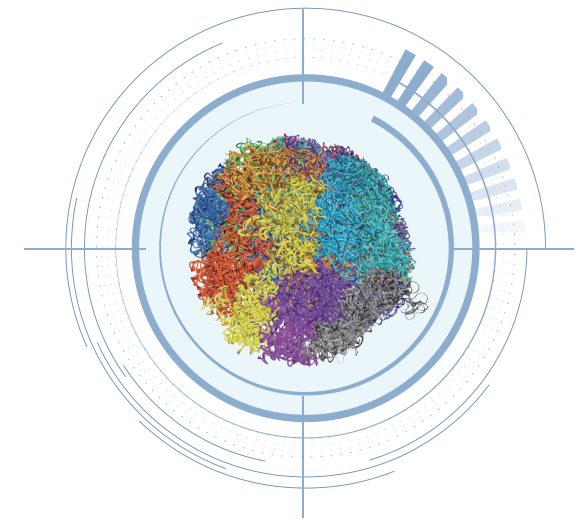


www.frasergen.com

三维基因组创新技术联盟

Hi-C



武汉菲沙基因信息有限公司

Wuhan Frasergen Bioinformatics Co., Ltd.

武汉市东湖高新技术开发区高新大道666号 邮编:430075

T:027-87224696 F:027-87224785 E:support@frasergen.com

武汉菲沙基因组医学有限公司

Wuhan Frasergen Genomic Medicine Co., Ltd.

武汉市东湖高新技术开发区高新大道666号B8栋 邮编:430075

Building B8, 666 Gaoxin Road, East Lake High-tech Zone, Wuhan, China 430075

北京研发中心

Beijing Research Center

北京市海淀区中关村南大街12号科技综合楼 邮编:100081

12 Zhongguancun South St., Haidian District, Beijing, China 100081



扫码关注更多信息

Frasergen 菲沙基因

Hi-C技术

高通量染色体构象捕获 (High-throughput chromosome conformation capture, Hi-C) 技术, 是指对捕获到的含有染色质构象信息的片段进行高通量测序, 获得全基因组范围的染色质空间互作信息, 从而研究全基因组的三维结构, 及其在基因转录、调控、复制和修复等生物过程中的功能。2009年, 美国 Job Dekker 研究团队首次开发了 Hi-C 技术, 并利用该技术绘制出了人类正常淋巴细胞基因组的三维结构。自此, 基因组学研究全面进入 3D 时代。

技术原理

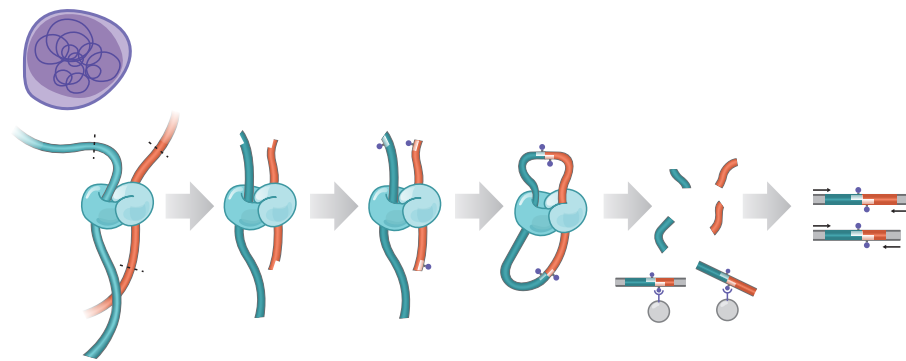
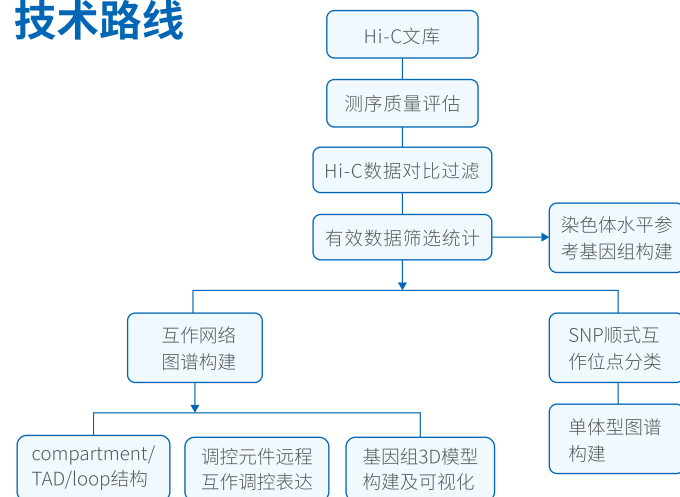


图 1 Hi-C 实验原理图

- 1 使用多聚甲醛处理细胞, 固定 DNA 的构象;
- 2 裂解细胞后, 使用限制性内切酶处理交联的 DNA, 产生粘性末端;
- 3 DNA 末端补平修复, 并同时引入生物素, 标记寡核苷酸末端;
- 4 使用 DNA 连接酶连接 DNA 片段;
- 5 蛋白酶消化解除与 DNA 的交联状态, 纯化 DNA 并随机打断为 300~500bp 的片段;
- 6 使用亲和素磁珠捕获标记的 DNA, 进行二代建库测序。

技术路线



Hi-C 交互网络构建的应用

- 全基因组空间互作信息, 一网打尽
- 寻找靶基因的调控元件, 揭示转录调控机制
- 分析染色质高级结构的基础
- 构建基因组3D结构模型

Hi-C 染色体构建优势

- Hi-C 组装至 Chromosome 水平, 优于 Bionano 的 scaffold 水平
- 无需遗传图谱中的作图群体, 单个个体即可实现更高锚定率的染色体构建
- 可对已经组装的基因组进行纠错, 可构建单体型
- 周期短、成本低, 可同时进行交互网络分析, 性价比超高

应用方向

01 揭示基因组空间结构调控基因表达的机制

构建全基因组互作图谱, 寻找大量的线性距离遥远而空间邻近的调控元件, 研究基因组空间结构 (compartment、TAD、loop 结构) 对基因表达和调控的影响和机制。

02 重构细胞核内染色质3D结构图

利用一定的数学模型, 将二维的基因组互作信息转化成三维空间结构的物理坐标, 重构全基因组的三维折叠结构并绘图, 直观的观察染色质在细胞核内的空间构象。

03 染色体水平参考基因组的构建

基于染色质内互作强度远大于染色质间互作, 且染色质内互作强度随距离增加而减弱的规律, 将基因组 contigs 聚类、排序、定向, 从而 scaffolding 至染色体水平。

04 构建单体型图谱

通过筛选杂合的 SNP 位点, 利用 Hi-C 数据中的顺式双端序列信息, 将同源染色体之间远距离 SNP 位点 (基因型) 进行连锁分类, 分成单体型图谱。

05 疾病和癌症发生机理研究

研究病理细胞或组织的基因组三维结构, 并与 WGS、RNA-seq、ATAC-seq、ChIP-seq 等多组学联合分析, 深度解析疾病和癌症的发生机理。

06 细胞生长发育或分化的调控机制研究

比较动植物不同处理条件或不同发育阶段染色质空间结构变化, 揭示环境变化或生长发育过程中基因组三维结构的适应性变化。

07 植物抗逆、抗病、农艺性状形成过程中的调控机制

比较正常对照组与不同环境刺激下的处理组或具有性状差异的植物亚品种, 细胞内染色质空间结构差异, 并与转录组、组蛋白修饰数据联合分析, 深度解析植物对环境刺激的应答及重要农艺性状形成的基因表达调控机制。

08 细菌、真菌等微生物耐药、致病、代谢产物调控机制

比较不同培养条件下或药物处理前后, 微生物细胞内基因组三维结构的变化, 包括互作结构域 CID/TAD 大小和边界变化、关键基因之间的互作变化, 联合转录组数据揭示耐药、致病、代谢产物合成等过程的基因表达调控机制。

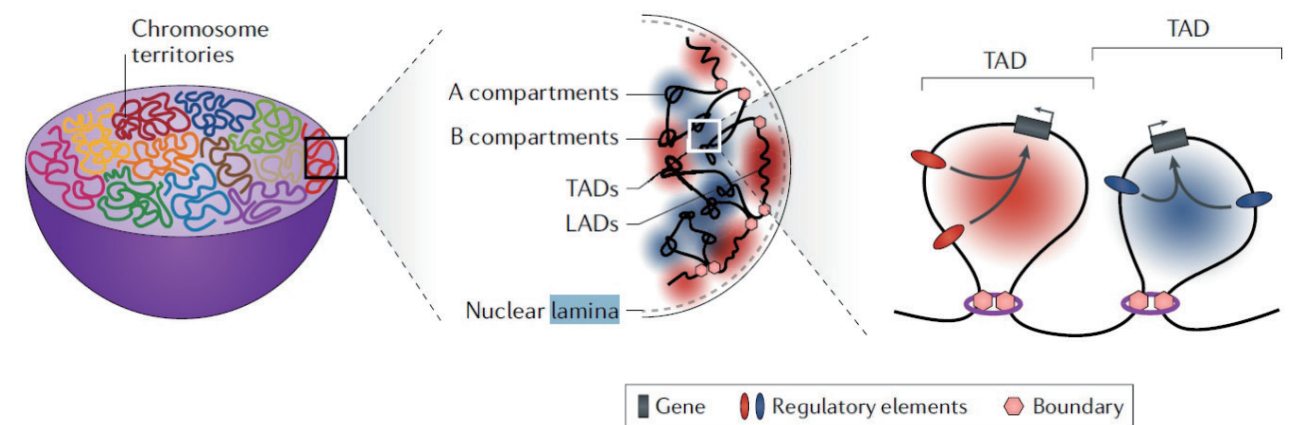


图 2 细胞核内基因组三维结构示意图

应用案例

案例一 基因组线性序列的结构变异导致三维结构改变产生手指畸形

Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, 2015.

采用 Hi-C 技术结合多组学数据已广泛应用于表型产生的机制研究中。如下图中 (Hi-C 联合 WGS 及 RNA-seq), 三个手指畸形病人 (短指、并指、多指), 基因组中 3 个 TAD 内涉及到的 4 个基因发生了缺失、重复、倒位, 这些染色体结构变异 (SV) 影响了基因组高级结构 TAD 的稳定性, 造成异常的基因互作, 改变了增强子的靶基因, 最终使不同基因的正常表达分别导致了不同疾病类型的产生。

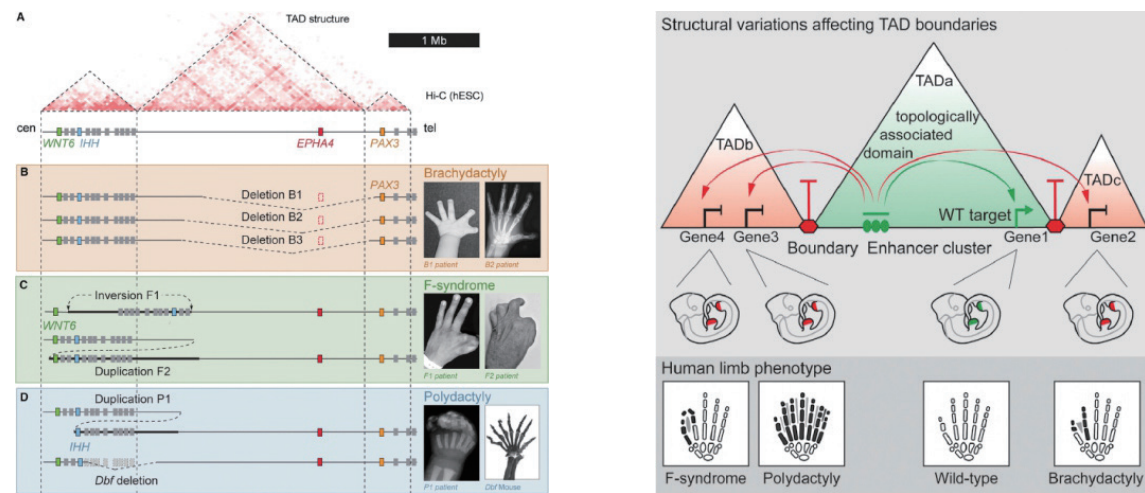


图3 DNA的序列变异 (SV) 与手指畸形

图4 增强子与靶基因的异常互作导致手指畸形

案例二 青鳉鱼胚胎发育过程中基因组三维结构的变化

CTCF looping is established during gastrulation in medaka embryos. *Genome Research*, 2021.

染色质环在基因组调控中有着重要的作用, 然而哺乳动物早期胚胎的 ChIP-seq 和 loop 分辨率下的 Hi-C 分析极有难度, 因此发育期间由粘连蛋白 (cohesion) 介导的染色质环仍然未知。在该研究中, 研究者采用 Hi-C 技术提供了 12 个时间点 (原肠胚形成之前、形成期间以及形成之后) 的青鳉鱼早期胚胎发育过程的基因组三维结构特征及其动态变化, 发现原肠胚的形成 (Gastrulation) 引起胚胎基因组结构发生巨大的变化, 其中包括有绝缘蛋白 CTCF binding 附近形成第一个染色质环以及互作 domain 大小增加很多。相对而言, 胚胎形成期间的 CTCF binding 则是固定的。

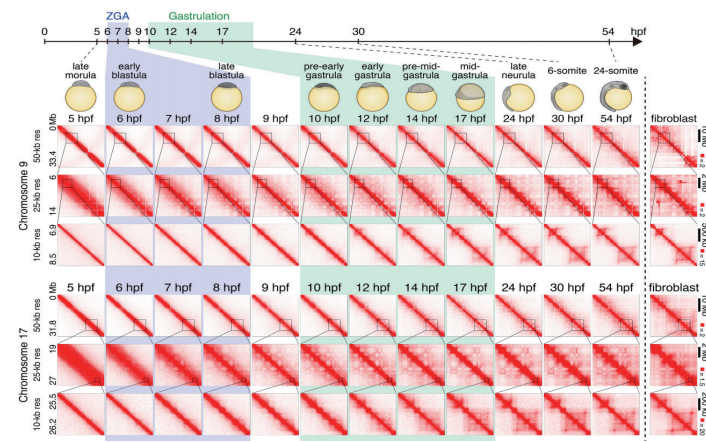


图5 青鳉鱼胚胎在发育成纤维细胞 (fibroblast) 前12个时间点的互作图谱 (ZGA, zygotic genome activation: 合子基因组激活; hpf, hr post fertilization: 受精x小时后)

案例三 Hi-C揭示不同条件处理儿童高级别胶质瘤的三维结构变化

Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Science Advances*, 2021.

儿童高级别胶质瘤 (pHGG) 包含有多型性胶质母细胞瘤 (GBM) 和突变型的弥漫性内在脑桥胶质瘤 (DIPG)。研究者采用 Hi-C 技术分析癌症亚型细胞之间, 以及患癌引起的特异性三维结构。研究发现, 相对于 DMSO 对照处理的细胞, BRD/BET 抑制剂 (BRD4i 和 dBET6) 处理的癌细胞中 DNA loop 结构大幅减少, 增强子 - 启动子互作受到破坏, A 和 B compartment 发生转换, 其中 dBET6 的影响更大。

研究者通过所有细胞系的 Hi-C 数据分析拷贝数变异 (CNV), 发现 DIPG 细胞系中有更多的 CNV, 随后根据 SV 数据进一步组装基因组 breaks, 识别到足以改变基因表达水平的增强子劫持事件, 并且抑制剂处理前后的 Hi-C 互作分析以及 FISH 实验证实该增强子劫持事件可作为药物研发的靶点。

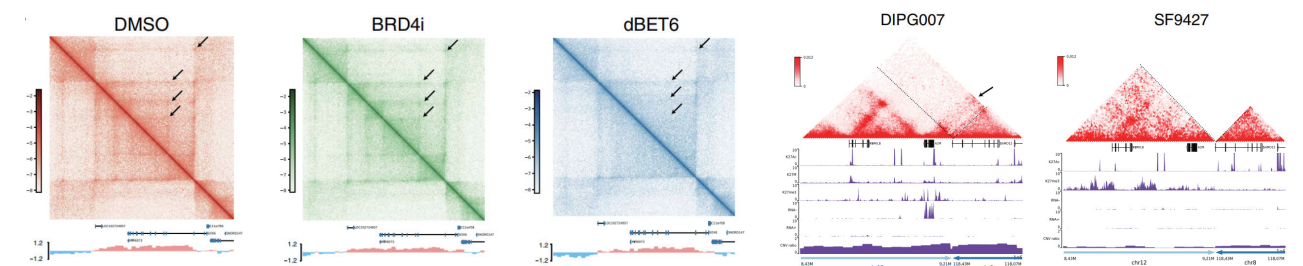


图6 肿瘤特异性互作 (箭头) 在BRD4i和dBET6处理24小时后被破坏, dBET6效果更强

图7 基于SV识别得到DIPG007中的增强子劫持事件

案例四 拟南芥基因组加倍对三维结构和基因表达的影响

The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Research*, 2019.

研究对象为野生型二倍体拟南芥 (Col-0) 和同源四倍体拟南芥 (4×Col-0) (秋水仙素处理获得), 对叶片组织分别进行 Hi-C、RNA-seq、ChIP-seq (H3K4me3、H3K27me3)。4×Col-0 的表型表现为叶片数量增多, 开花时期延长。三维结构特征为 4×Col-0 的基因组染色体体内互作减少, 染色体间互作增多。染色质的松紧程度发生转换, 松散区域 (LSD) 富集 H3K4me3 信号, 紧密区域 (CSD) 富集 H3K27me3 信号。差异互作和差异表达基因重叠区富集在饥饿、营养、细胞稳态等应答信号通路中。

发现开花抑制基因 FLC 的表达水平在 4×Col-0 中增加, 并使用 3C 实验验证了 FLC gene loop 结构的强度增加。

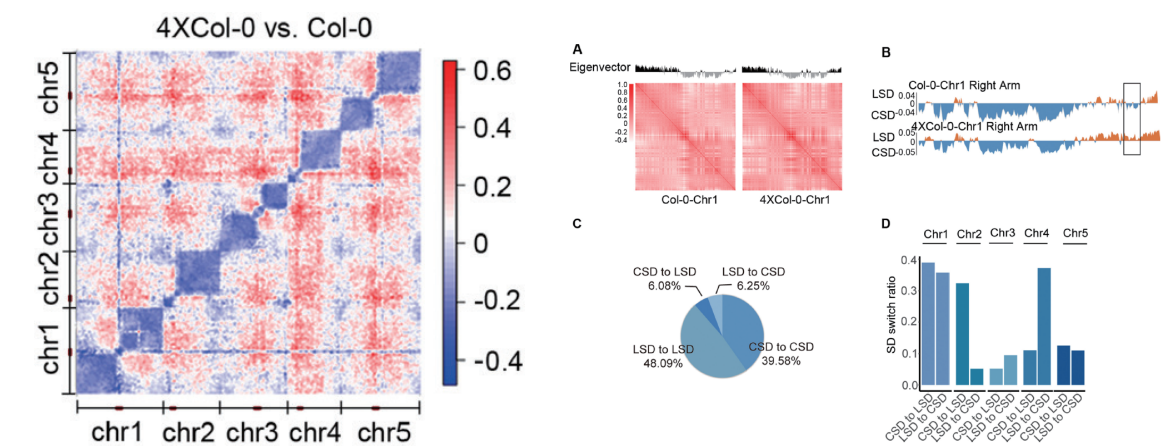


图8 4×Col-0基因组染色体体内互作减少, 染色体间互作增多

图9 染色质松散和紧密区域的转换

案例五 Hi-C深入解析胰腺癌的遗传基础

A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. *Genome Medicine*, 2021.

研究者通过GWAS分析,得到了624个候选SNP,为进一步了解它们的生物学功能,以及识别其他的胰腺癌遗传易感位点,研究者基于胰腺癌组织的Hi-C三维数据,探究其6761个显著性互作,并将XBP1基因为目标区域,该区域与四个局部莫兰系数(LMI)选择的SNP相互作用。

研究发现,与正常胰腺相比,XBP1启动子上的H3K27Ac修饰在胰腺癌细胞系(PANC-1)中完全丢失并在一种胰腺癌前病变(PanIN-1B)中减少。进一步表征XBP1上游启动子区域,研究发现,在癌前病变和PANC-1细胞中相应目标区域中增强子/弱启动子明显丢失。这种活性的丧失使得XBP1表达在癌细胞中降低。

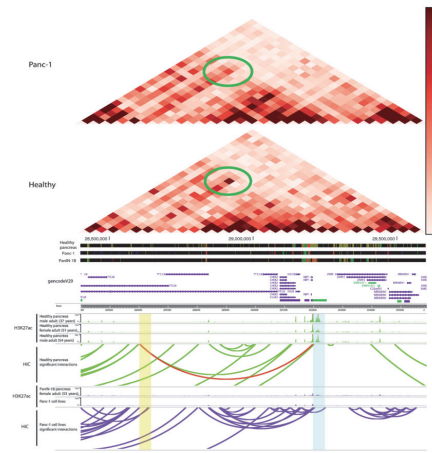


图8 PANC-1细胞和正常细胞的三维基因组结构,以及XBP1周围基因组区域对应的多组学关联

案例六 Hi-C解析苹果中花青素在光诱导下的生物合成

The Long Noncoding RNA MdLNC499 Bridges MdWRKY1 and MdERF109 Function to Regulate Early-Stage Light-Induced Anthocyanin Accumulation in Apple Fruit. *Plant Cell*, 2021.

花青素有助于植物着色,是日常饮食中来自于水果和蔬菜的重要抗氧化剂。在苹果(Malus domestica)中,花青素经由光诱导后进行生物合成,苹果果皮着色。

研究者鉴定了一种ERF(乙烯反应因子)蛋白-ERF109,它参与光诱导的花青素生物合成,并通过直接结合花青素相关基因启动子来促进着色。启动子:GUS(β -葡萄糖醛酸酶)报告基因分析和Hi-C测序表明,MdERF109和在其上游的长非编码RNA(lncRNA) MdLNC499位于一个交联信号很强的区域,并且在光诱导花青素形成的过程中,MdLNC499调控着MdERF109的表达。MdERF109蛋白也在苹果着色的早期诱导花青素相关基因的表达和花青素的积累。

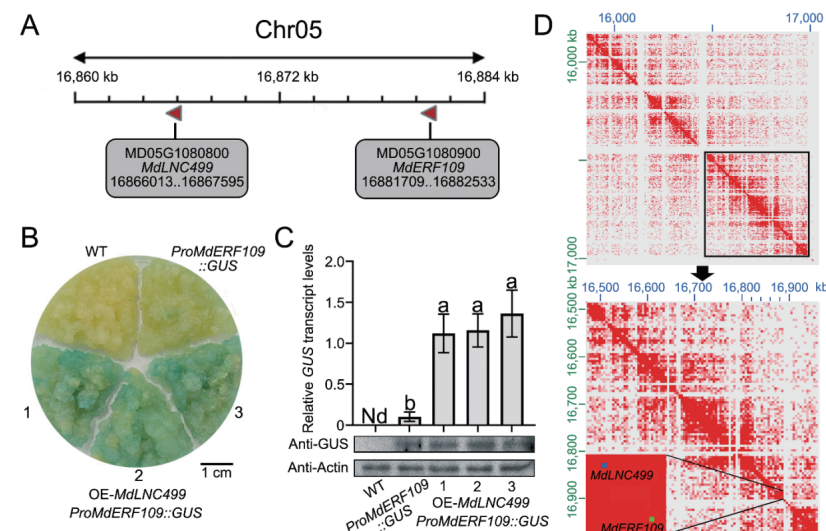


图9 MdERF109和MdLNC499的位置(9A)以及在染色体水平的互作(红色区域,9D);苹果中的瞬时表达(9B)和苹果愈伤组织中GUS蛋白的表达(9C)

近期已发表的Hi-C文章列表(节选)

物种	年份	期刊	科学问题
蝾螈	2021	Proceedings of the National Academy of Sciences	巨型蝾螈基因组中远距离转录调控元件和染色体的拓扑结构
玉米	2021	Journal of Experiment Botany (非沙项目文章)	经Hi-C技术比较两种玉米自交系的三维空间结构
人	2021	Nature	相分离转录因子(NUP98-HOX9)诱导原癌基因处富集依赖CTCF的染色质环
人	2021	Molecular Systems Biology	细胞周期关键因子HMGB1协调衰老进程的空间结构
鼠	2021	Nucleic Acids Research	保守的DMRT1转录因子在性细胞命运重编程中招募SOX9转录因子
人	2021	Science Advances	儿童高级别胶质瘤中两种亚型的三维结构以及药物处理后所引起的变化
酵母	2021	Genome Research	应用Hi-C得到表达不同SIR3等位基因的出芽酵母的全基因组互作图,确认Sir3结合处直接促进远距区域的长距离互作
人	2020	Nature Communications	辐射诱导的DNA损伤和修复对三维基因组结构的影响
六倍体小麦	2020	Genome Biology	小麦染色质三维结构
拟南芥	2020	Nature Communications	热应激诱导的转座子激活与拟南芥中的三维染色质结构重排有关
鼠	2020	Nature Communications	小鼠体细胞核移植胚胎中的染色质结构重组
谷氨酸棒状杆菌	2020	Nature Communications	Hi-C结合高分辨率显微镜得到二倍体谷氨酸棒状杆菌的基因组结构,Hi-C阐述了oriC复制起始位点附近SMC介导的长距离互作
需钠弧菌(非沙项目文章)	2020	Frontiers in Microbiology	Hi-C解析微重力处理对需钠弧菌的空间结构,基因表达的影响,发现前噬菌体区域有明显的基因表达及结构的变化影响

菲沙基因Hi-C组装项目(节选)

物种	应用技术	Contig/Scaffold N50
某食用菌	PacBio+Hi-C	Contigs的N50=2,580,783 bp, Scaffolds的N50=4,102,088 bp;Contig总长度染色体锚定率为95.55%;BUSCO完整性评估为97.6%。
油料作物	PacBio+Hi-C	Contig N50=6.2Mb, Scaffold N50=50Mb Contig总长度染色体锚定率为99.32% BUSCO完整性评估92%
中草药	PacBio+Hi-C	Contig N50=14.2Mb, Scaffold N50=88Mb Contig总长度染色体锚定率为98% BUSCO完整性评估90%
水生动物	PacBio+Hi-C	Contig N50=2.1Mb, Scaffold N50=66.2Mb Contig总长度染色体锚定率为96.7% BUSCO完整性评估93%
木本中药材	PacBio+Hi-C	Contig N50=1.7Mb, Scaffold N50=36.3Mb Contig总长度染色体锚定率为98% BUSCO完整性评估95%
高寄生虫	PacBio+Hi-C	Contig N50=3.0Mb, Scaffold N50=51.4Mb Contig总长度染色体锚定率为92% BUSCO完整性评估98%
某鱼	PacBio+Hi-C	Contig N50=2.2Mb, Scaffold N50=32.4Mb Contig总长染色体锚定率为95.7% BUSCO完整性评估99%

菲沙基因在国内率先开展三维基因组学研究,开发了完善的Hi-C实验及数据分析流程,参与了众多的国际前沿科学研究,举办了首届三维基因组学国际研讨会、三维基因组信息分析培训班。公司参与的全球首例人造单染色体真核生物酿酒酵母研究成果于2018年8月和12月相继在国际顶级期刊《Nature》、《Cell Research》发表。菲沙基因具有丰富的项目经验,可为广大科研学者提供高水平高质量的三维基因组研究服务。