基于Hi-C测序的拓扑关联结构域识别与可视化

Identification and Visualization of Topologically Associating Domains Based on High-throughput Chromosome Conformation Capture

报 告 人：李平静

导　　师：刘健 教授

南开大学 生物信息与智能医学中心

2024年05月

南開大學
Nankai University

提纲

一、研究背景及研究内容

二、基于深度学习的Hi-C数据增强算法

三、基于Hi-C接触矩阵的TAD识别方法

四、三维基因组可视化方法

五、总结与展望

一、研究背景：一维线性测序

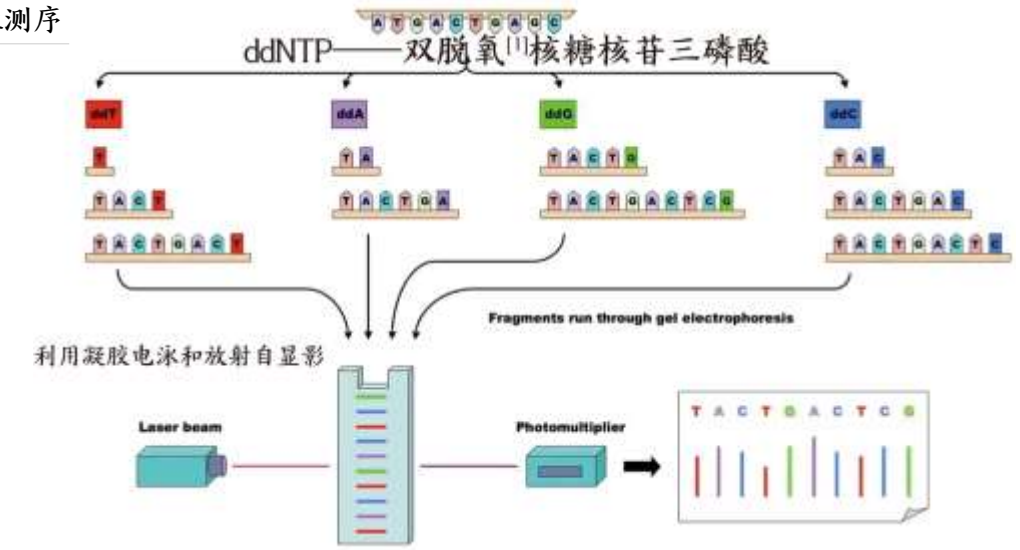
基因测序

基因测序是指对生物体的基因组进行分析和测定，以确定其中的碱基（ATCG）序列，从而了解其遗传信息和基因组结构。

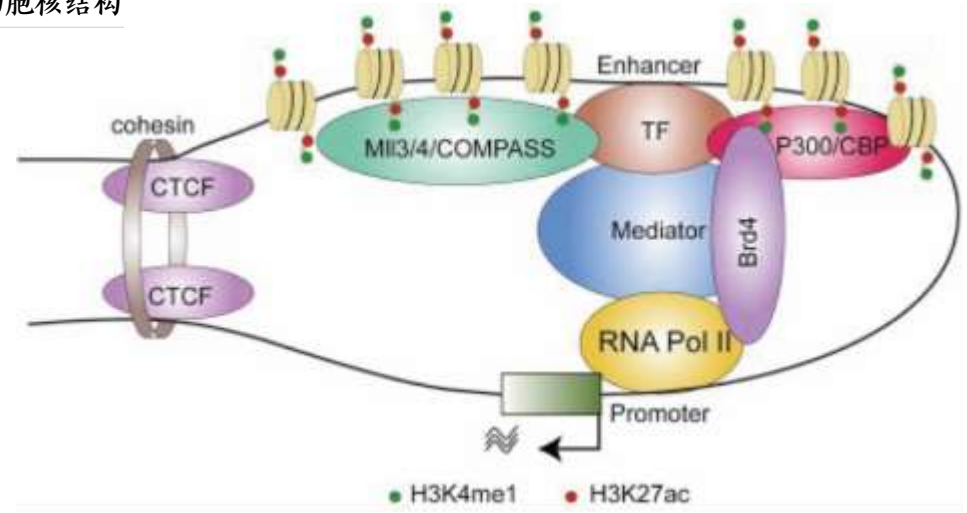
- 基因组研究
- 疾病诊断与预防
- 药物研发
- 个体化医学
- 进化研究

测序技术是
现代生物信息学的基础

线性测序



复杂的细胞核结构



以染色体坐标

为度量的

一维线性序列

以染色体位点接触

为度量的

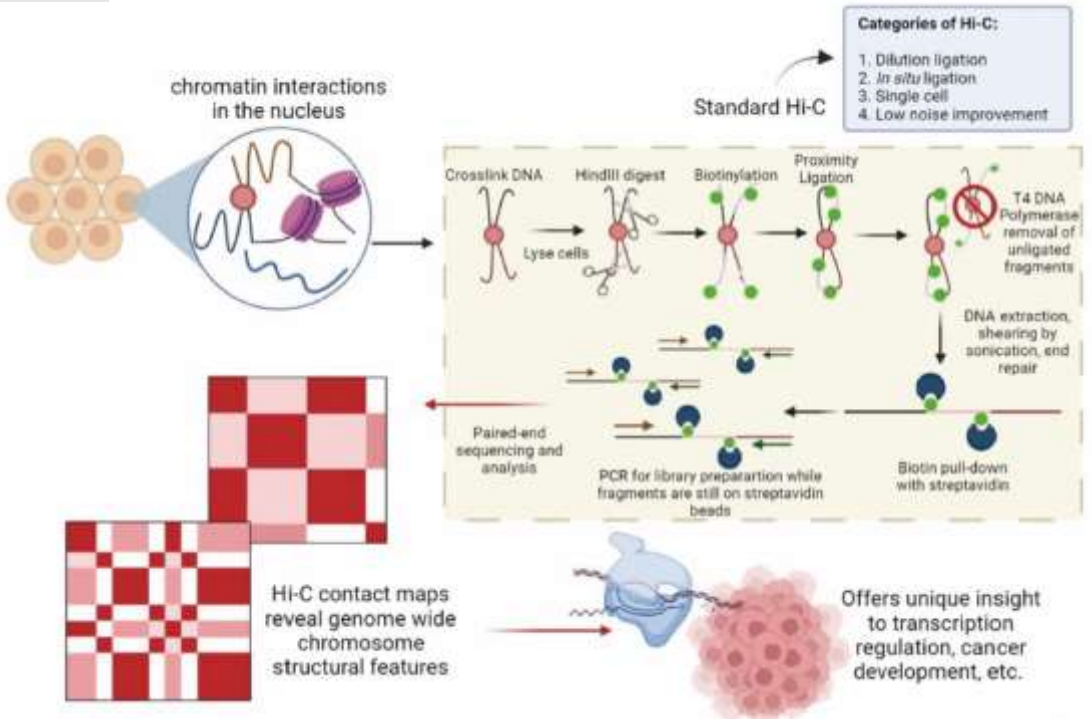
三维空间结构分解

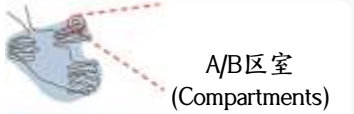
HiC测序技术

1、双脱氧：第3号碳原子上的羟基被氢原子取代，参与DNA合成后，无法提供3’-OH末端而使反应终止
2、4种ddNTP：为了测定哪些位置是A、T、C、G

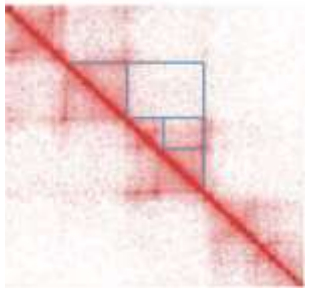
一、研究背景：Hi-C测序与三维空间结构

Hi-C生物实验

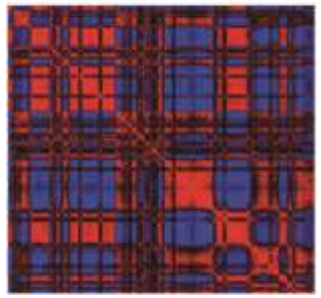




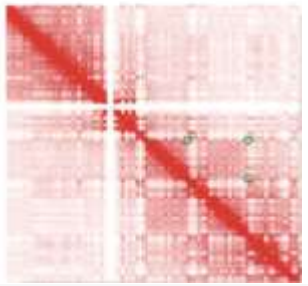
A/B区室
(Compartments)



拓扑关联结构域
(TAD)



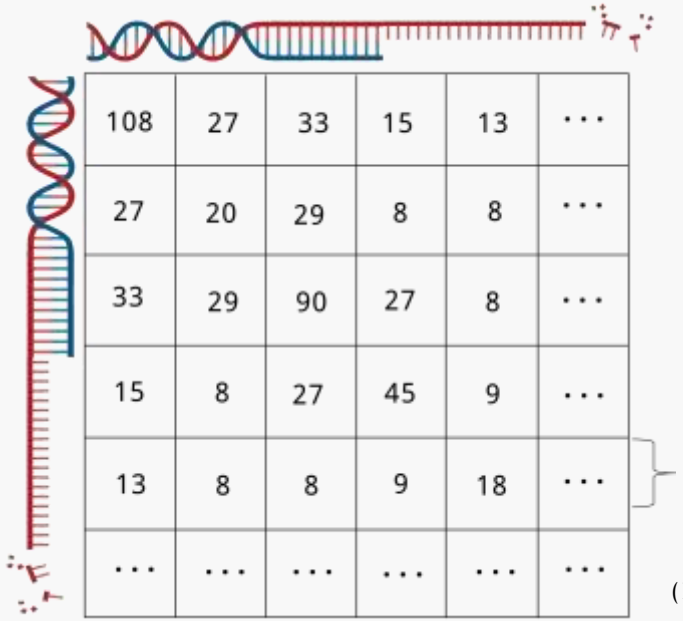
染色质环
(Chromatin Loops)



二维接触矩阵：

代表染色体复杂空间结构在基因片段水平上的分解

三维基因组的研究内容为三维，其数据结构为二维！

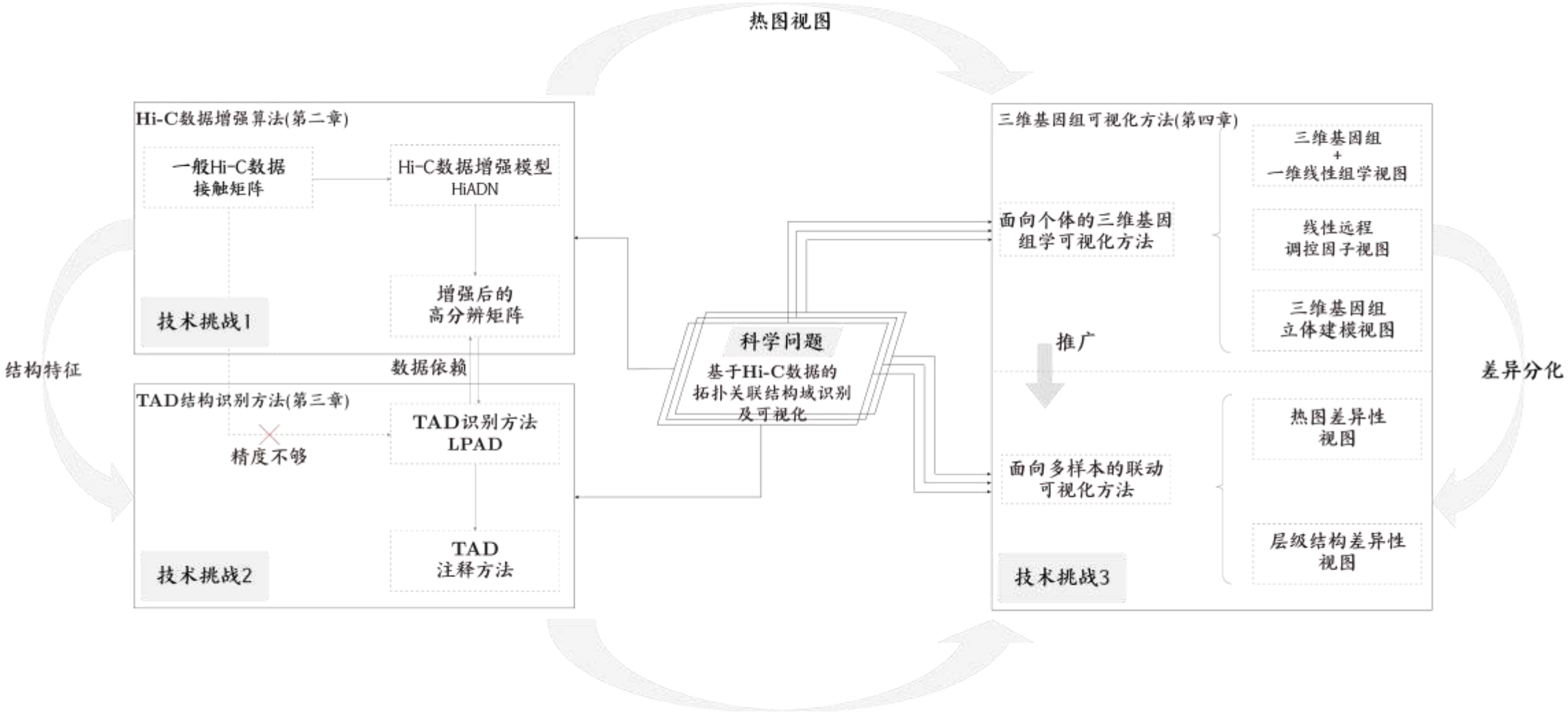


108	27	33	15	13	···
27	20	29	8	8	···
33	29	90	27	8	···
15	8	27	45	9	···
13	8	8	9	18	···
···	···	···	···	···	···

bin

分辨率
(Resolution)

一、研究内容：基于Hi-C数据的拓扑关联结构域识别及可视化



提纲

一、研究背景及研究内容

二、基于深度学习的Hi-C数据增强算法

三、基于Hi-C接触矩阵的TAD识别方法

四、三维基因组可视化方法

五、总结与展望

二、HiADN：基于深度学习的Hi-C数据增强算法

研究背景 → 方法设计 → 实验结果 → 本章小结

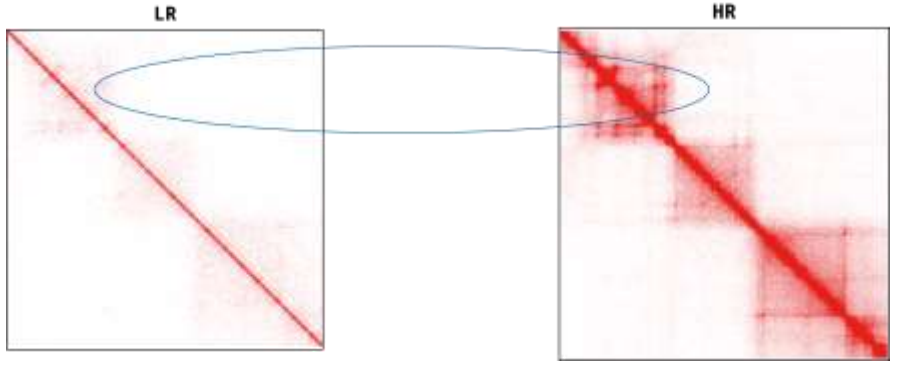
Problem

* 测序技术导致的bias（酶、GC含量等）
* 测序深度/获取到的DNA片段少
* 低分辨率下无法识别三维空间结构
* 低分辨率下遗漏关键的相互作用

Objective

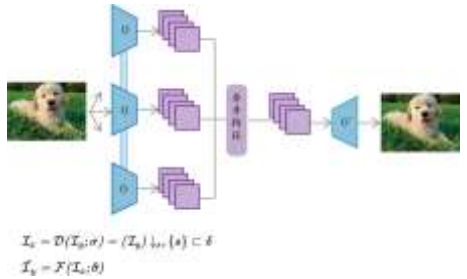
* 计算方法增强/预测高分辨的矩阵
* 利用现有的稀疏的矩阵

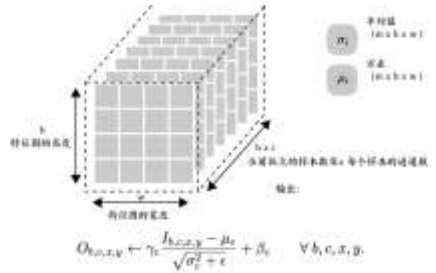
高低分辨率热图



方法	分类	方法	效果
HiCPlus	CNN[1]	3-layers	首次提出
HiCNN	CNN	54-layers	局部+全局连接
SRHiC	CNN	ResNet	ResNet，降低计算量
DFHiC	CNN	空洞卷积	全局捕获
HicGAN	GAN[2]	第一个GAN	第一个GAN
DeepHiC	GAN	感知损失	纹理平滑（TV_LOSS）
HiCSR	GAN	DAE损失	矩阵特征
VEHiCLE	GAN	TAD损失	下游分析结果

1、CNN（卷积神经网络）：这里特指超分辨率中的卷积网络
2、GAN（生成对抗模型）
3、HiCPlus：Hi-C数据增强开篇之作，(2018)Nature Communications

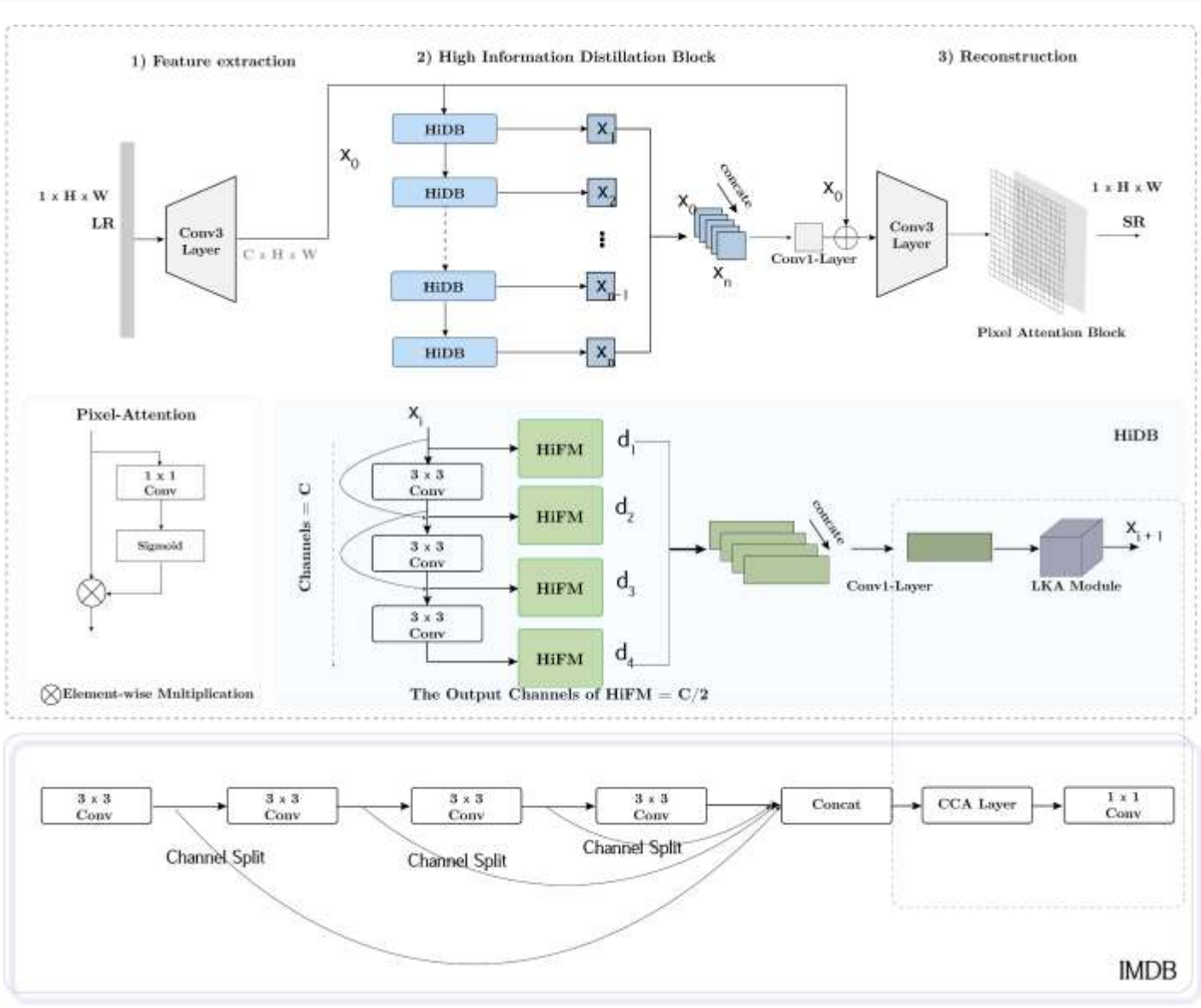




繁重的模型　残影　远程交互　纹理平滑

二、HiADN：基于深度学习的Hi-C数据增强算法

研究背景 → **方法设计** → 实验结果 → 本章小结



设计思路

1、轻量级网络IMDB

2、减少隐藏层通道数：Hi-C单通道

3、设计HiFM结构（Rethink IMDB）

：提取Hi-C矩阵局部细节

4、设计卷积分解操作（LKA）

：一种全新的注意力机制

5、移除所有的BN-layer

6、引入SRB（小残差模块）

：加速训练，提高泛化

7、使用MS-SSIM损失

：更加关注于纹理（层级结构）

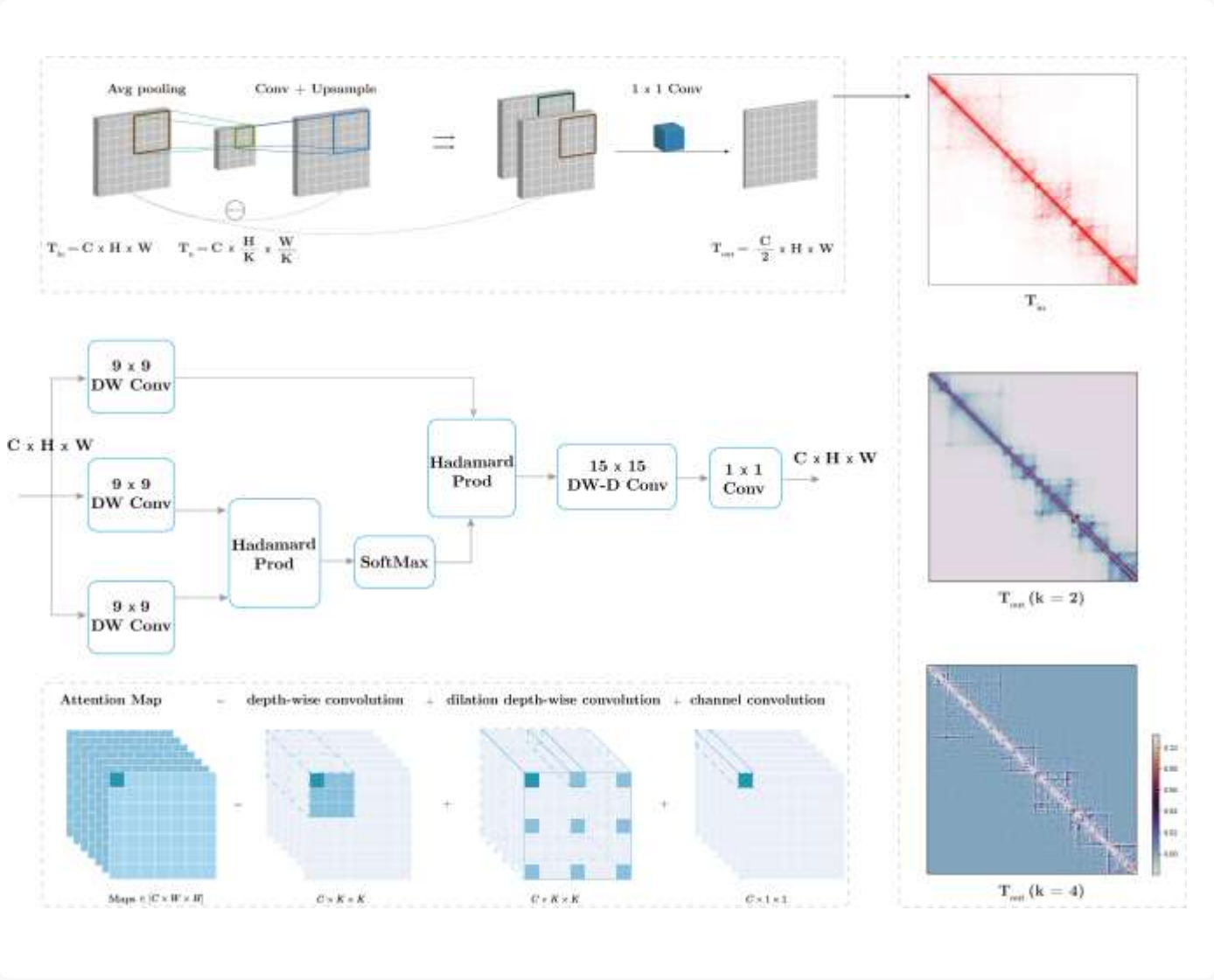
$$\mathcal{F}_0 = f_s(\mathcal{I}_{\mathcal{LR}})$$

$$\mathcal{F}_n = \Phi^n(\Phi^{n-1}(\ldots\Phi^1(\mathcal{F}_0)))$$

$$\mathcal{F}_d = f_3(f_1([\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n]) + \mathcal{F}_0)$$

$$\mathcal{I}_{\mathcal{SR}} = f_p(\mathcal{F}_d)$$

二、HiADN：基于深度学习的Hi-C数据增强算法



效果

HiFM

。Tout成功地保留了特征图的精细

。减少和平滑对角线的值

。从而减轻模型的学习负担

LKA

。局部特征

。全局、超远程感知

。通道注意力

。参数少

	Standard Conv.	*LKA*	*Reduction*	*Ratio(%)*
C=48, K=3	20,736	**3,168**	17,568	84.72
C=48, K=9	186,624	**10,080**	176,544	94.60
C=64, K=3	36,864	**5,248**	31,616	85.76
C=64, K=9	331,776	**14,464**	317,302	95.64

与当前先进模型的定量比较

参数数量：

526K参数，SOTA三分之一

训练时间：

轻量，50个周期达到收敛，

远快于GAN模型，pipeline只需要2小时

预测效果：

数据集：GM12878、K562和CH12-LX

采样比例：16、32和100

交叉实验：✓

指标：PSNR、SSIM、DISTS

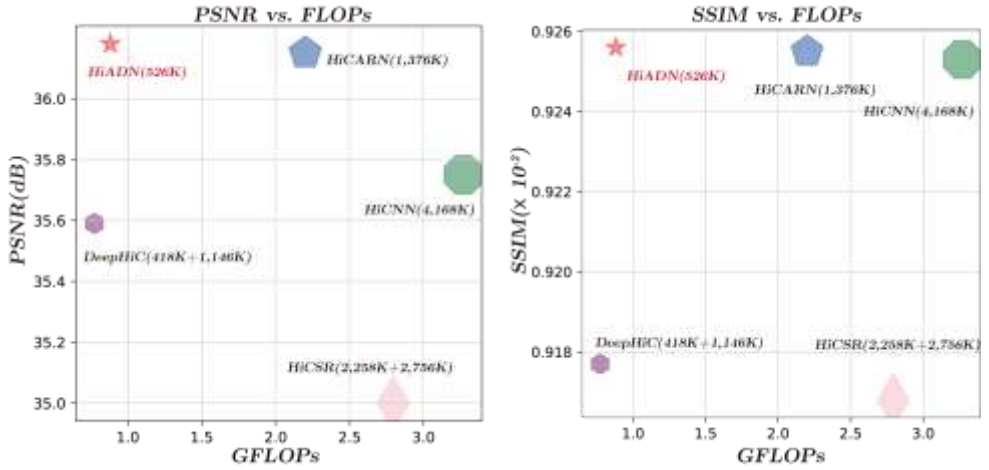
生物：Genome DISCO，HiC-Spector、QuASAR-Rep

均取得第一或第二

HiADN是一个轻量级网络

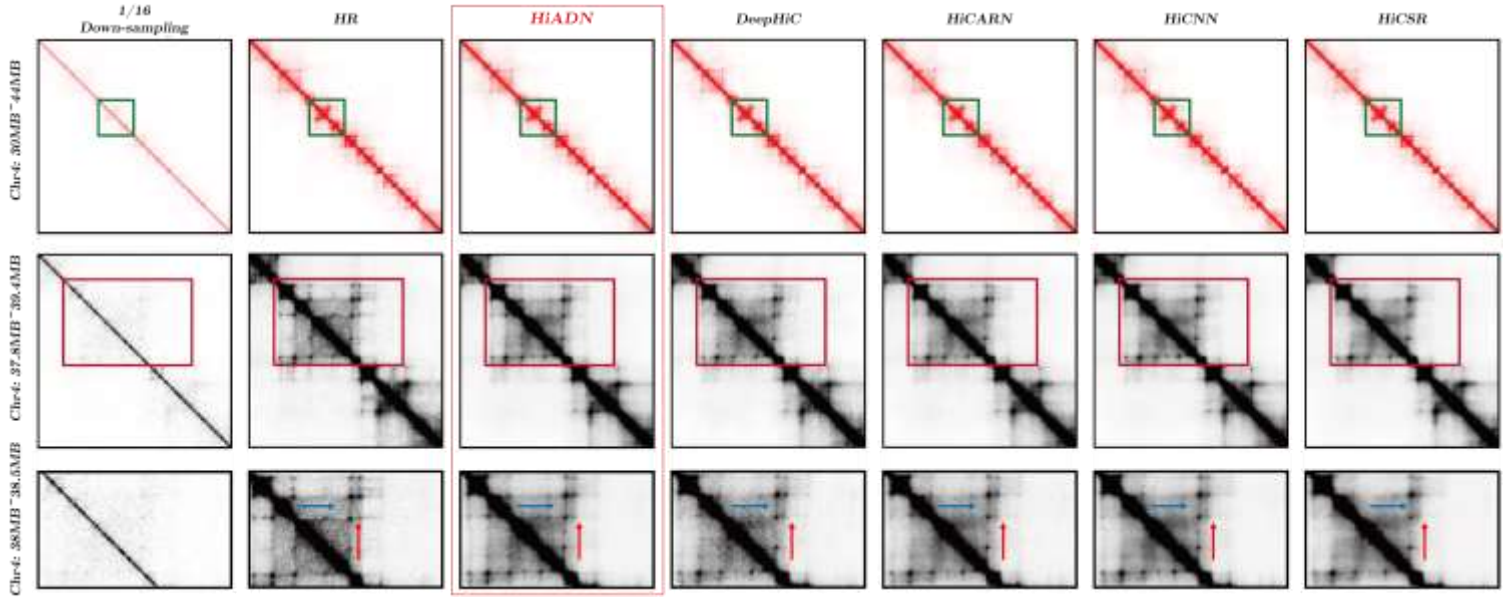
Method	*Layers*	*Params.*	*FLOPs*	*Training time*
HiCARN	52	1377K	2.2G	42 × 100
HiCNN	51	4169K	3.3G	156 × 200
DeepHiC	24+13	418K+115K	0.8G	33 × 200
HiCSR	84+21	2258K+2756K	3.1G	109 × 600
HiADN	**383**	**526K**	**0.8G**	**43 × 50**

HiADN在性能和资源取得平衡

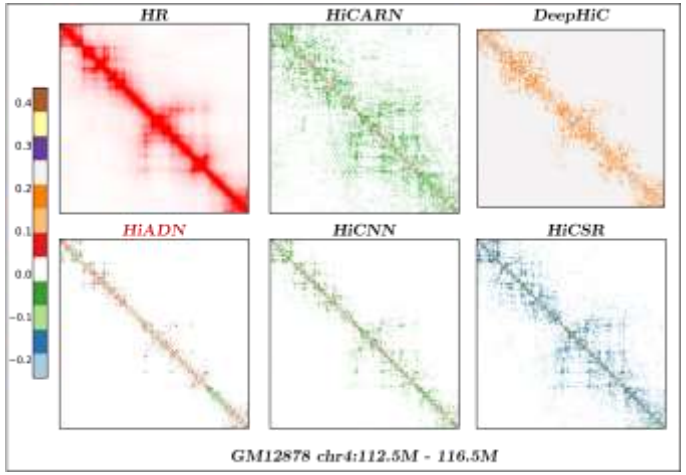


二、HiADN：基于深度学习的Hi-C数据增强算法

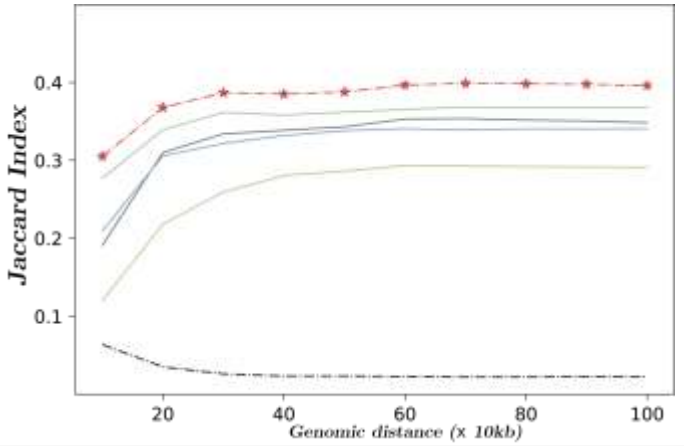
视觉效果比较



SR－HR 差值比较



不同基因距离下，显著交互位点再现



结果

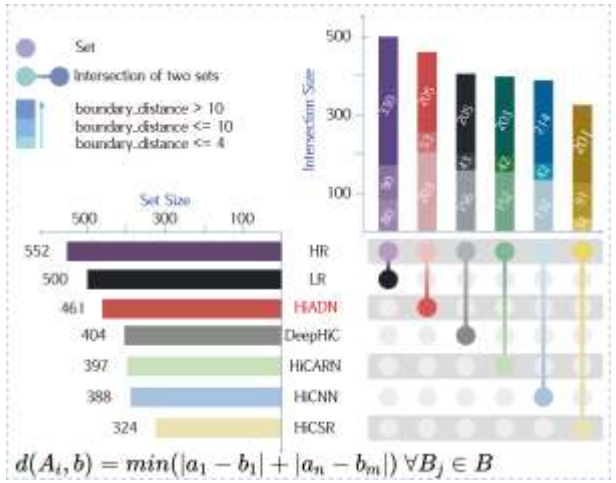
- ✓ HiADN能够更加精确的重建精细的三维结构（Sub-TAD、Stripes）
- ✓ 空间结构的重建得益于HiFM结构（消融实验）
- ✓ 在恢复显著性交互上，HiADN比其他方法领先约0.05~0.1左右

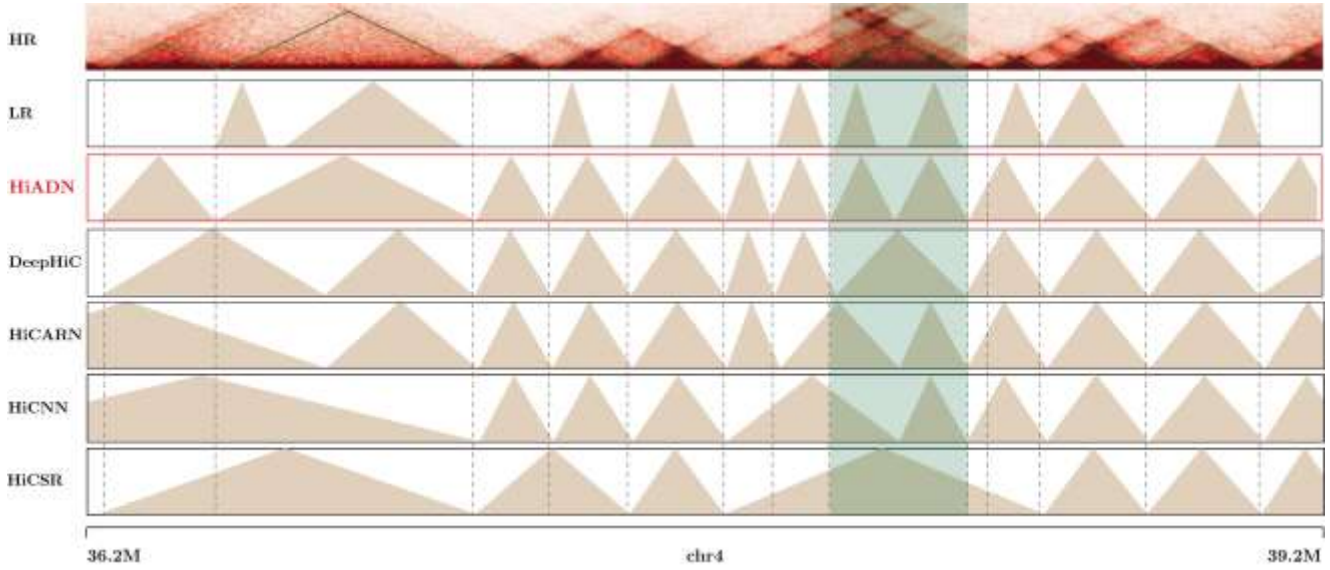
二、HiADN：基于深度学习的Hi-C数据增强算法

三维结构重建比较

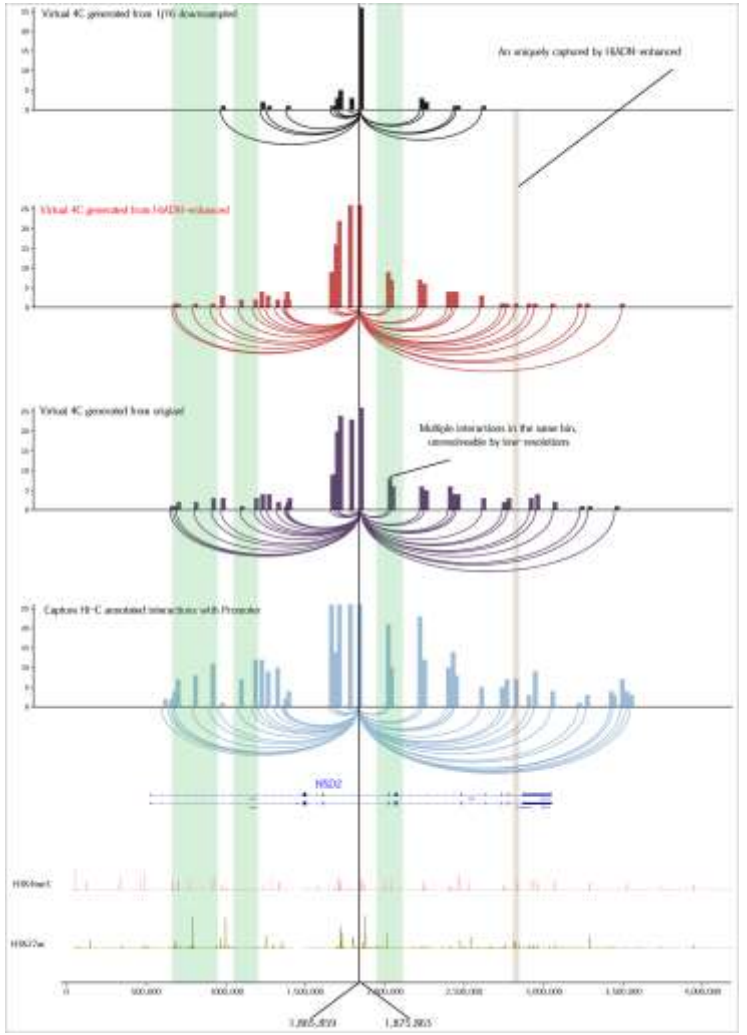
- HiADN在TAD重建上具有重大优势
- HiADN造成的假阳性结果最少

TAD: 46.37% vs. 36.05%
Loops: 31.78% vs. 29.96%
Stripes: 10(+/39)、12 (-)





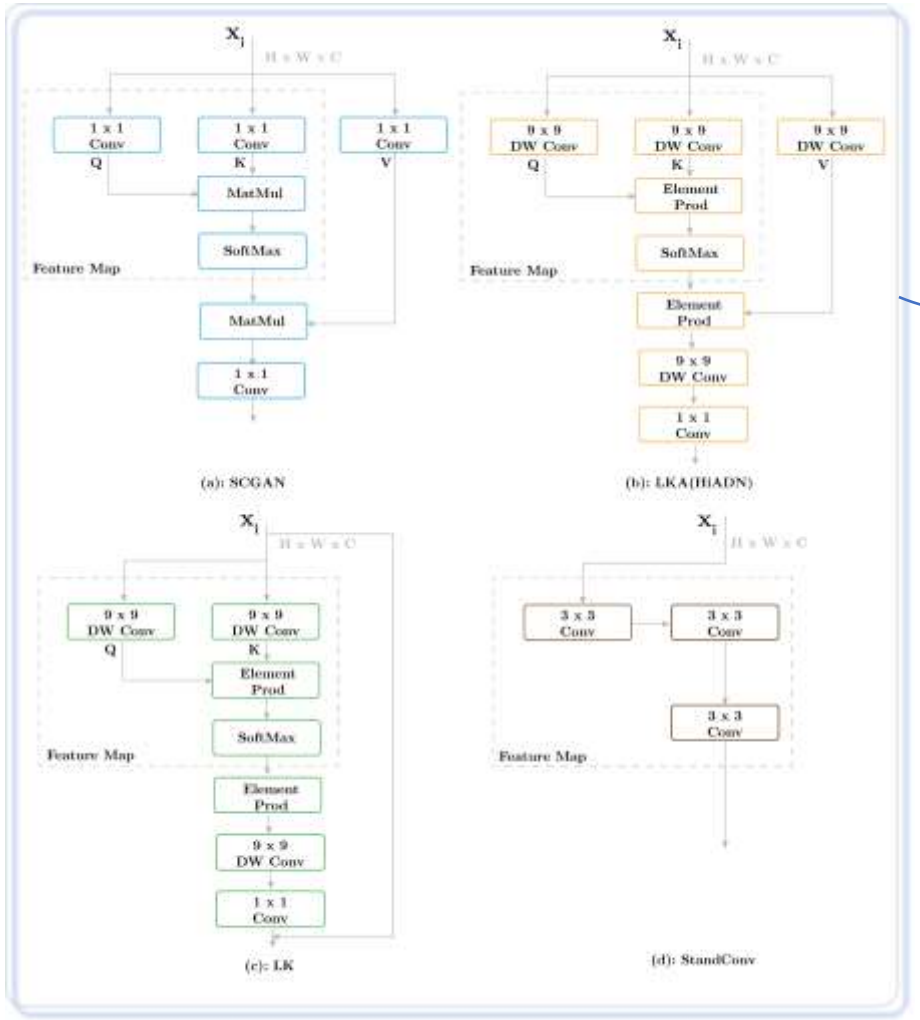
捕捉遗漏的增强子与启动子之间的重要交互



二、HiADN：基于深度学习的Hi-C数据增强算法

研究背景 → 方法设计 → **实验结果** → 本章小结

四种注意力机制



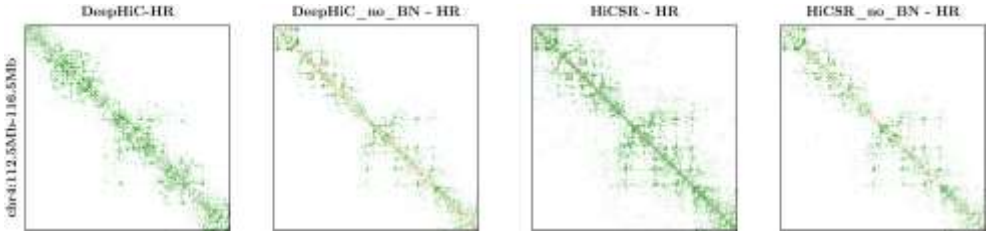
LKA（注意力）和HIFM的消融实验

方法	参数量	GM12878	K562	CH12-LX
HiFM	503K	36.16 0.9150 0.1561	34.07 0.9422 **0.1556**	34.81 0.9653 0.1310
SCGAN	426K	36.02 0.9231 0.1571	34.18 0.9438 0.1599	34.64 0.9649 0.1340
LK	526K	36.18 0.9254 0.1555	33.28 0.9432 0.1578	35.08 0.9662 0.1318
Conv	660K	36.04 0.9242 0.1564	33.31 0.9435 0.1595	34.79 0.9605 0.1308
HiADN*	526K	**36.18 0.9256 0.1555**	**34.18 0.9466** 0.1577	**35.25 0.9668 0.1298**

HiFM引入约**2.3万**的参数量，PSRN提升**0.44dB**，SSIM增加**0.02**，DISTS降低**0.0019**,TAD重建数量提高**60**个
标准卷积引入额外的**14万**参数（HiADN的**26.6%**）,LKA只引入其约**1/10**的参数

批标准化（BN）的消融实验

方法	参数量	GM12878	K562	CH12-LX
DeepHiC	417K	35.59 0.9177 0.1790	33.86 0.9400 0.1656	34.65 0.9652 0.1345
DeepHiC*	417K	35.70 0.9191 0.1678	35.06 0.9438 0.1603	35.24 0.9687 0.1310
HiCSR	2254K	31.65 0.8868 0.1977	33.24 0.9089 0.1937	32.06 0.8707 0.1853
HiCSR*	2254K	36.13 0.9253 0.1825	34.25 0.9467 0.1759	35.13 0.9726 0.1365



移除BN后，TAD和Loops的预测准确率显著提升**62%**

本章小结

为了克服生物实验影响，有效地利用现有数据，本文引入了一种基于卷机神经网络的框架 HiADN，用于提高Hi-C接触矩阵的分辨率。本章搭建了骨干网络，极大的降低了计算资源并提高了计算效率。本章所提出的卷积注意力机制模块，通过使用卷积分解和融合操作对传统的自注意力机制进行替换，能够有效地捕捉Hi-C数据中的全局模式。

由于缺少末端配序列导致接触矩阵稀疏无法将多数位点连接起来，这使得在鉴定染色体拓扑关联结构域时边界发生转移，与典型的大小不匹配。本文注意到以往的TAD识别算法往往关注于使用局部上下游区域（通常为 5~10 个 bin）的交互频率作为该bin的特征用于后续识别，在此背景下如果某两个位点的连接缺失对于该特征是有影响的。

定量结果

Method	scale	GM12878 PSNR	GM12878 SSIM	GM12878 DISTS↓	K562 PSNR	K562 SSIM	K562 DISTS↓	CH12-LX PSNR	CH12-LX SSIM	CH12-LX DISTS↓
LR	1/16	20.75	0.5469	0.2350	24.78	0.7158	0.2080	28.69	0.8810	0.1544
HiCSR	1/16	36.01	0.9255	0.1877	33.47	**0.9534**	0.1706	32.06	**0.9712**	0.1489
HiCARN	1/16	36.15	0.9158	0.1674	33.44	0.9529	0.1609	34.51	0.9644	0.1304
HiCNN	1/16	35.97	0.9159	0.1787	33.10	0.9507	0.1603	32.92	0.9662	0.1403
DFHiC	1/16	35.93	0.9127	0.1854	33.08	0.9486	0.1724	33.33	0.9644	0.1541
DeepHiC	1/16	35.59	0.9170	0.1555	33.86	0.9400	0.1656	34.65	0.9635	0.1298
HiADN*	1/16	**36.18**	**0.9256**	**0.1555**	**34.18**	0.9446	**0.1577**	**35.25**	0.9668	**0.1298**
LR	1/32	20.08	0.5029	0.2531	24.21	0.6883	0.2248	27.50	0.8659	0.1683
HiCSR	1/32	34.97	0.9066	0.2116	31.99	0.9294	0.1946	29.01	0.9392	0.1764
HiCARN	1/32	35.23	0.9143	0.1840	31.24	**0.9381**	0.1764	31.23	0.9561	0.1499
HiCNN	1/32	34.90	**0.9151**	0.1948	30.94	0.9324	0.1772	31.78	**0.9591**	0.1488
DFHiC	1/32	35.05	0.9126	0.2010	32.25	0.9369	0.1855	31.08	0.9580	0.1594
DeepHiC	1/32	34.80	0.8996	0.1984	33.13	0.9199	0.1794	31.18	0.9561	0.1435
HiADN*	1/32	**35.31**	0.9146	**0.1705**	**33.16**	0.9330	**0.1718**	**31.86**	0.9562	**0.1427**
LR	1/100	19.85	0.4800	0.2846	24.01	0.6722	0.2590	27.14	0.8569	0.1948
HiCSR	1/100	33.80	0.8914	0.2261	30.06	**0.9269**	0.2034	27.64	0.9450	0.1788
HiCARN	1/100	33.83	0.8980	0.2027	29.40	0.9188	**0.1854**	28.15	**0.9502**	**0.1575**
HiCNN	1/100	33.62	0.8968	0.2171	29.98	0.9213	0.1987	28.91	0.9460	0.1731
DFHiC	1/100	33.58	0.8964	0.2189	29.93	0.9201	0.1999	28.77	0.9488	0.1699
DeepHiC	1/100	33.49	0.8856	0.2253	**30.08**	0.9166	0.1962	28.76	0.9474	0.1652
HiADN*	1/100	**33.93**	**0.8983**	**0.1869**	29.95	0.9177	0.1922	**28.92**	0.9489	0.1580

提纲

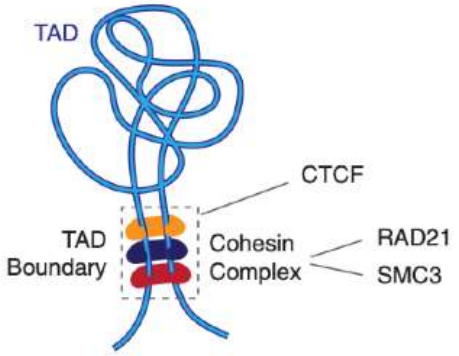
一、研究背景及研究内容
二、基于深度学习的Hi-C数据增强算法
三、基于Hi-C接触矩阵的TAD识别方法
四、三维基因组可视化方法
五、总结与展望

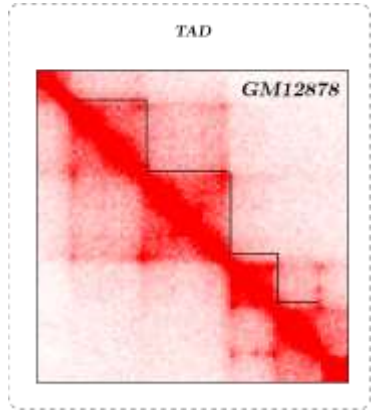
三、LPAD：基于Hi-C接触矩阵的TAD识别方法

染色体由许多不同长度的结构域组成，边界广泛存在着大量绝缘子结合蛋白CTCF、管家基因、SINE逆转座子等的调控因子

- 域内“自聚簇”（Self-association）：域内基因的“协同调控,结构内部的基因持有共同的调控元件
- 域间“绝缘性”（Insulation）：TAD 边界阻止着域间的基因调控作用

染色体三维结构中基本调控单元





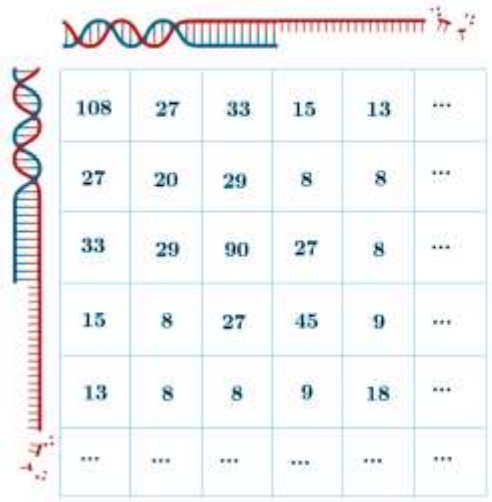
方法	分类	特点	缺陷
DI	线性得分	HMM	难用，参数
TopDom*	线性得分	滑窗	参数选择
MSTD	聚类	谱聚类	假阳高

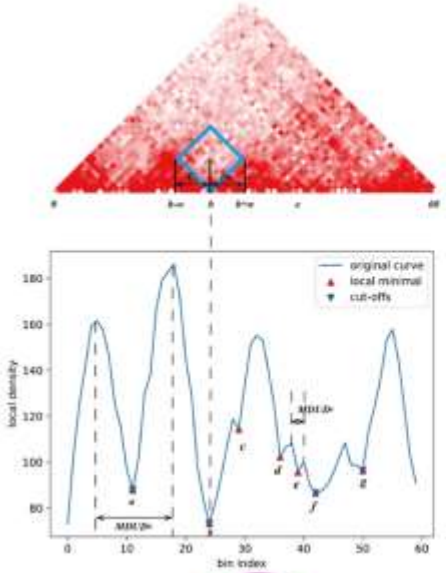
- 聚类模型，往往选择第i个bin的上下游一段组成一个向量，采取相似度计算指标和聚类算法进行计算
- 参数难选择：是指需要用户指定窗口大小，而未明确不同分辨率下，如何选择。并且选择不同的参数结果差异巨大

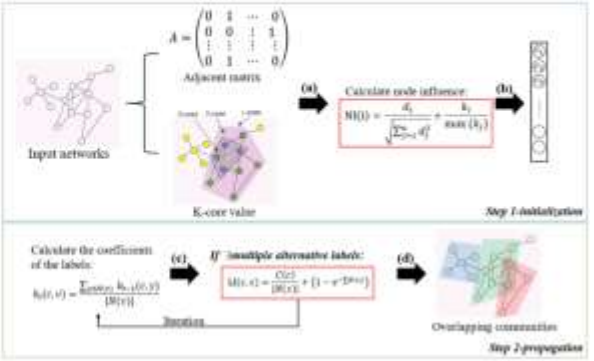
二、LPAD：基于Hi-C接触矩阵的TAD识别方法

研究背景 → **方法设计** → 实验结果 → 本章小结

LPAD算法流程







1 → 2 → 3 → 4

问题转化：

1、Hi-C接触矩阵看作临界矩阵
2、TAD看作图中社区

前提：

1、 Hi-C接触矩阵是对称矩阵
2、TAD内部高度交互

随机游走策略

1、TAD作为染色体折叠形成的结构，如何利用全局信息？
2、接触矩阵转化为邻接矩阵？
3、如何处理对角线处极大值？

近万个结点，如何加速？

$$\rho_k = \frac{1}{w^2}\sum_{l=1}^{w}\sum_{m=1}^{w}\mathcal{P}^t(\mathcal{U}_k(l),\mathcal{D}_k(m))$$

$$\epsilon_i = d[\mathcal{U}_{max}, i] + d[i, \mathcal{D}_{max}]$$

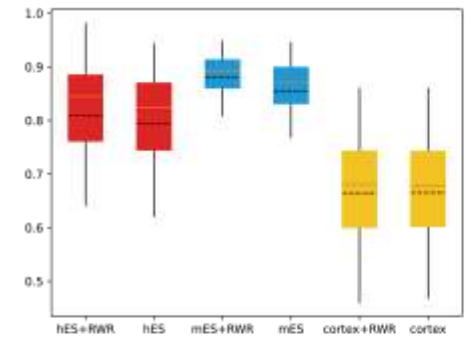
$$k = n - \sum \mathcal{L}_{0/1}(2w - \rho_k)$$

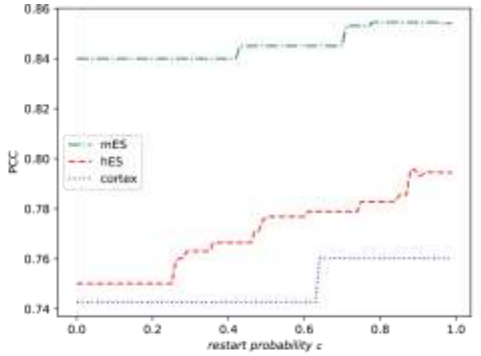
社区发现算法

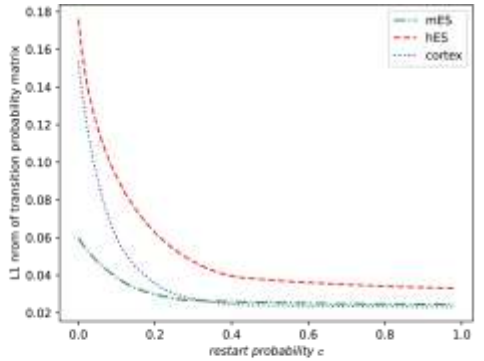
标签传播：

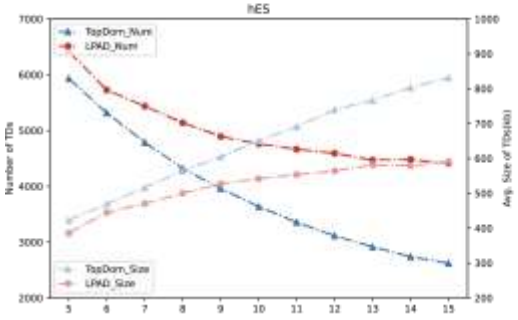
1、每个结点分配标签
2、计算直接相连的邻居+距离惩罚
3、得分排序，取最大，更新
4、迭代至收敛（或最大步骤）

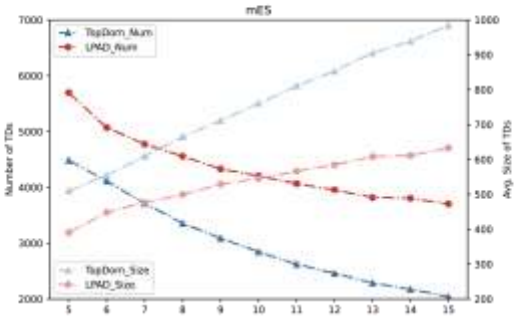
关于随机重启游走和窗口大小的消融实验

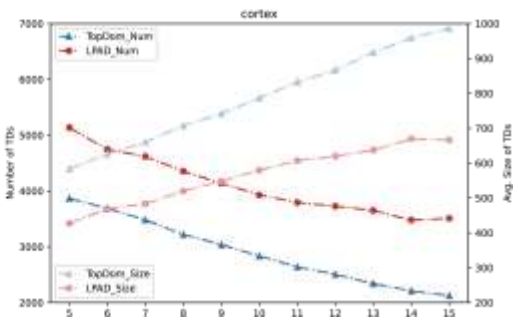












$$PCC(TAD_i) = \frac{2}{n(n-1)} \sum_{l=i}^{m} \sum_{k=i+1}^{m} corr(C(l), C(k))$$

实验结论

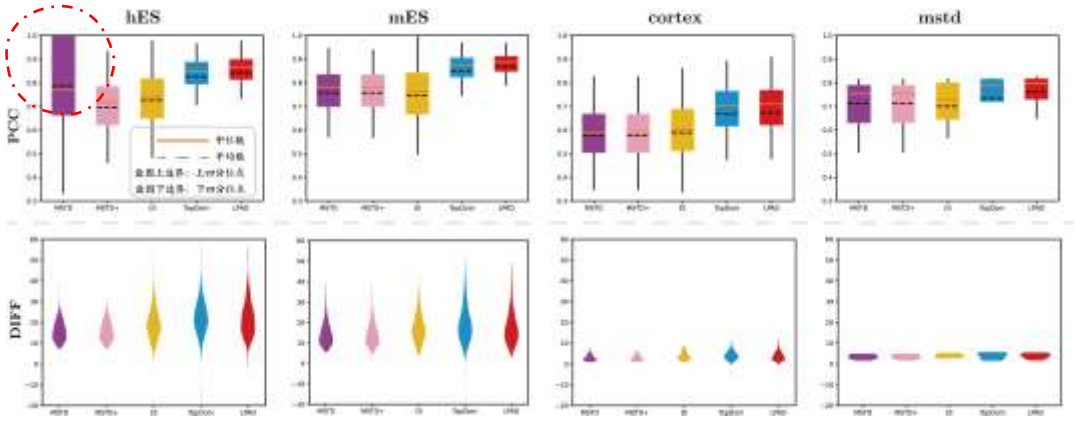
✓RWR能提升识别的TAD质量

✓RWR中重启概率取值对于结果影响不大（> 0.8）

✓与TopDom相比，参数的选择对于结果的影响更小，具有一致性的识别结果

三、LPAD：基于Hi-C接触矩阵的TAD识别方法

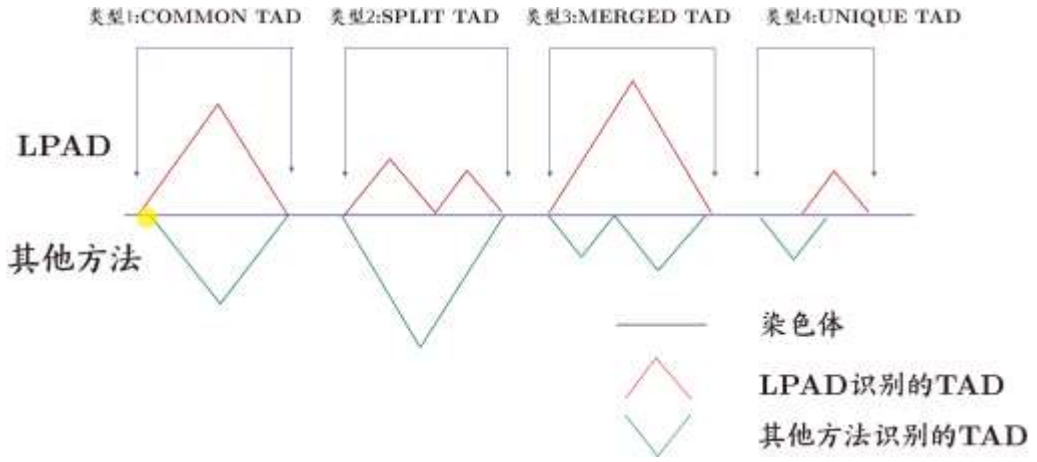
与当前先进的方法关于PCC和DIFF指标的比较



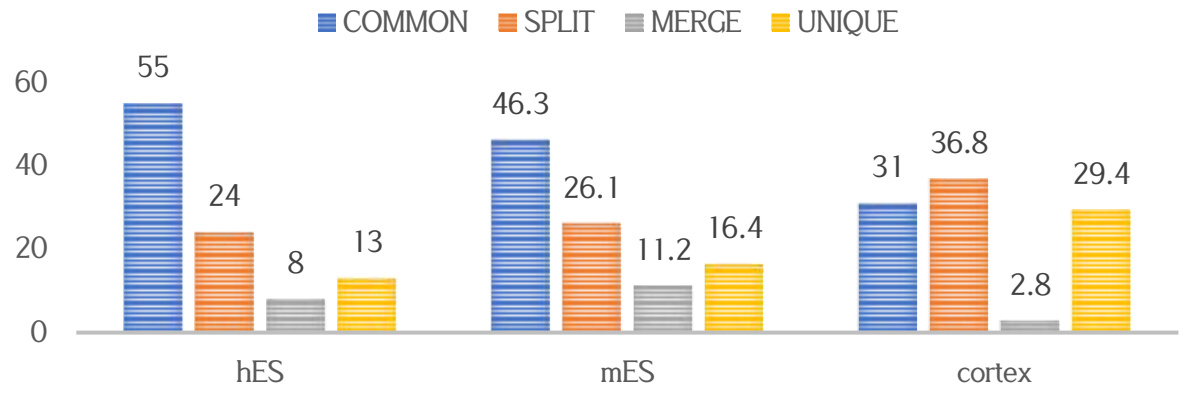
对于SPLIT类型的TAD，根据PCC/DIFF的计算公式可知，显然对一个TAD切分后，PCC/DIFF值越高。

对此部分本文将通过ChIP-seq蛋白质富集分析来探讨！

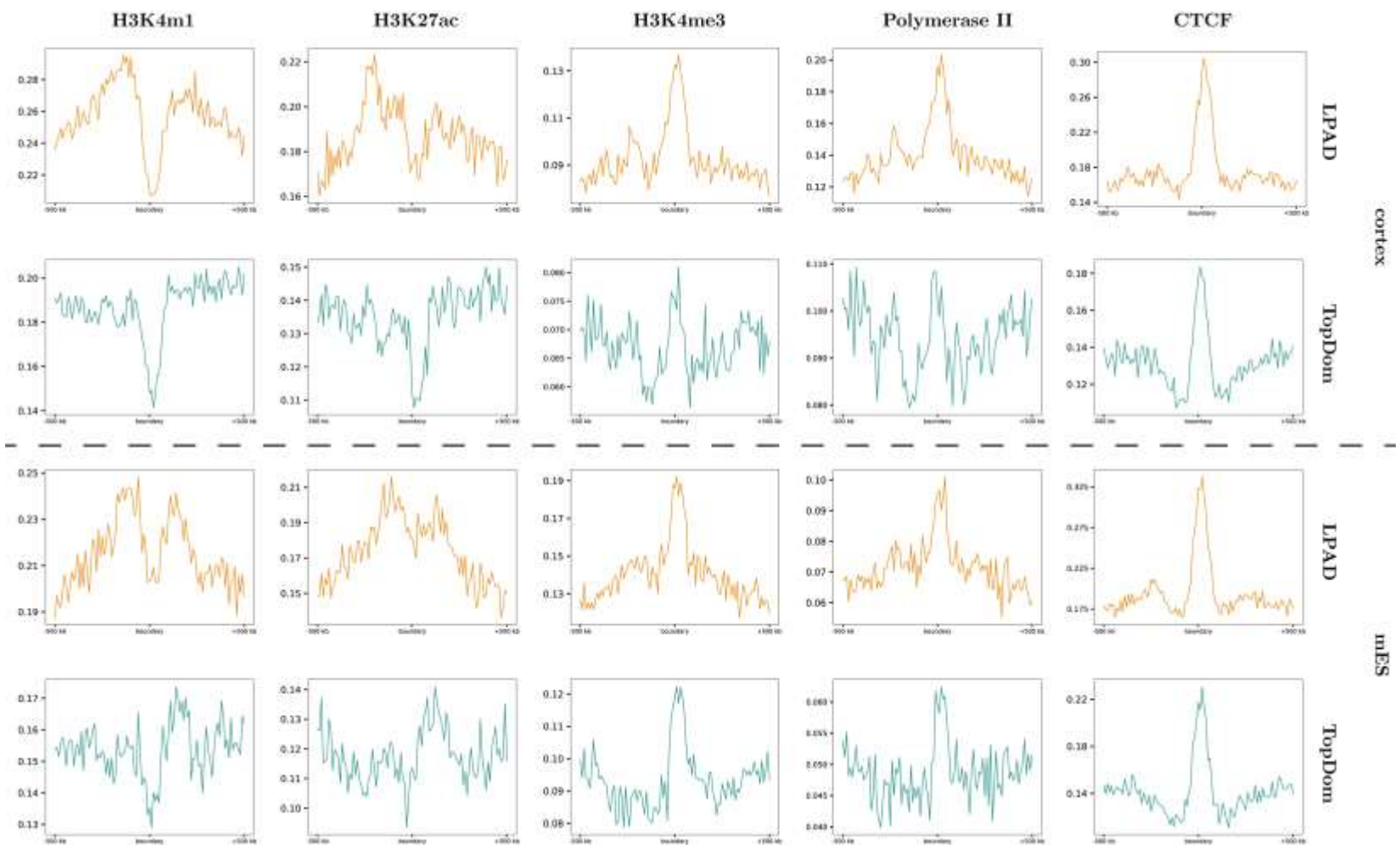
定义四种类型的TAD



与TOPDOM比较



SPLIT类型TAD边界组蛋白修饰富集分析



以10kb长度的片段作为一个分箱，计算了边界点上下游+/-500kb区域的ChIP-Seq峰值平均频率

结果

五种组蛋白修饰信号的富集/缺失分析

- H3k4me1
- H3k4me3
- H3k27ac：急剧消耗
- Polymerase II
- CTCT

TAD的形成与启动子具有密切关联
LPAD预测的TAD的边界具备相应的生物学功能

LPAD输出文本格式的TAD及其注释内容

```
# Annotation the TAD of 16:1170-1181, resolution:40000
[
    ...
    {
        "id": "NM_153029",
        "locus": "chr16:47,130,137-47,201,592",
        "strand": "-",
        "name": "N4BP1",
        "enhancer": "chr16:49094466-49095242,chr16:48753822-48755022,chr16:48939235-48941566",
        "promoter": "chr16:48610160-48610161",
        "super_enhancer": "",
        "diseases": "Rheumatoid Arthritis,Eosinophil count procedure,Blood basophil count (lab test)",
        "TF": "",
        "target": ""
    },
    {
        "id": "NM_001006610",
        "locus": "chr16:46,951,947-46,957,299",
        "strand": "-",
        "name": "SIAH1",
        "enhancer": "",
        "promoter": "chr16:48365886-48365887,chr16:48385409-48385410,chr16:48448397-48448398",
        "super_enhancer": "",
        "diseases": "Carcinogenesis,Liver carcinoma,Parkinson Disease,Breast Carcinoma",
        "TF": "EHMT2(-),TP53(?)",
        "target": ""
    }
    ...
]
```

本章小结

本章提出了TAD识别方法LPAD，基于图节点随机游走和社区发现的新模型。实验结果表明，LPAD能有效提高TAD的检测质量。在组蛋白修饰标记的富集分析实验中，LPAD显示了其准确性和优越性。

为了更好的展示结果，并将**注释内容**与 Hi-C 热图、一维线性注释、基因坐标等关联上，并提供多样本之间的差异化分析，在此基础上，下文将继续探索Hi-C热图与多组学整合可视化、基因注释结果可视化和差异分析可视化方案。

提纲

一、研究背景及研究内容

二、基于深度学习的Hi-C数据增强算法

三、基于Hi-C接触矩阵的TAD识别方法

四、三维基因组可视化方法

五、总结与展望

四、HiBrowser：三维基因组可视化方法

研究背景 → 方法设计 → 实验结果 → 本章小结

Hi-C测序的痛点与多组学优势

Hi-C测序为研究3D基因组组织和功能提供了前所未有的机会，当其数据结构形式潜在的忽略了序列（ATCG）的信息，因此将Hi-C和其他基因组学数据的综合分析能更好地理解染色质的结构和功能作用。

Hi-C发展方向：

1、多模式Hi-C、3D基因组注释数据可视化
2、多组学交互式
3、对照组比较分析（时间维度）
4、注释数据探索
5、私有数据、动态交互

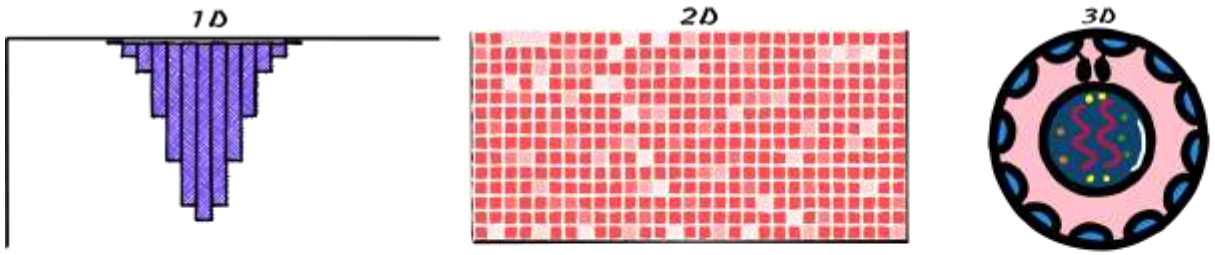
特征	*HiBrowser*	*GB*	*3DIV*	*NB*	*HiGlass*	*Juicebox*	*WashU*
动态交互	√			√	√	√	√
开放数据	√				√	√†	√†
多组学全景	√	√‡		√		√‡	√
多样本	√	√$			√$		
热图叠加	√					√∓	
立体建模	√			√			√
cREs	√		√±				
快速导航	√						
本地部署	√				√	√	√

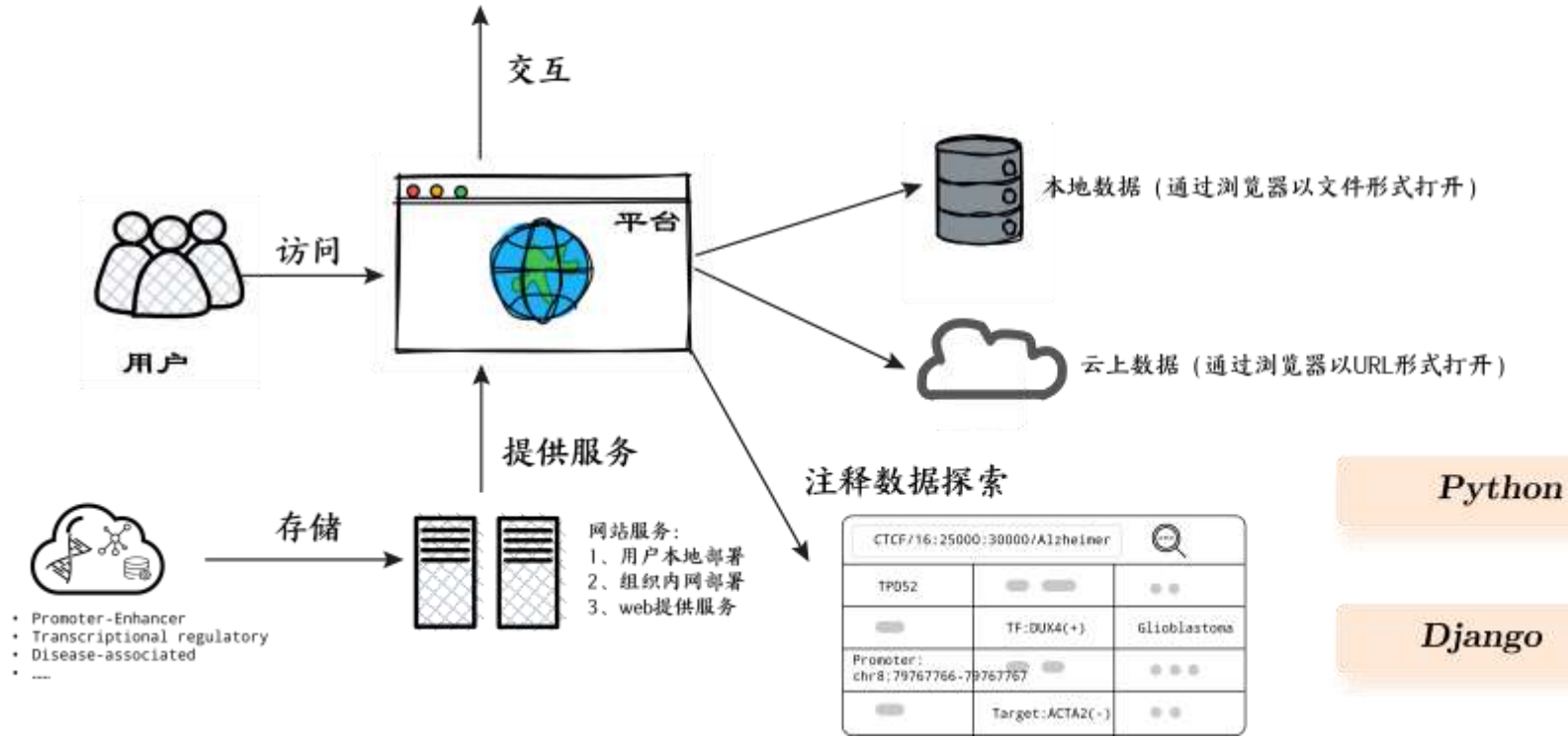
?

⋆ Juicebox 只可以与 Hi-C 热图交互。
† HiGlass 必须本地部署并配置数据目录，WashU 无法选择本地参考基因组。
‡ GB 嵌入 WashU 页面，无法移动，无法使用本地文件；Juicebox 支持少数的定量注释数据。
$ GB 为静态系统，多样本即多图片，HiGlass 无法取消同步状态、无法比对到不同参考基因组。
∓ Juicebox 不支持 AMB 热图叠加，不支持分别绘制多样本 2D 注释。
± 3DIV 仅含收集的 80 个样本对应 cREs 数据。

* 表中列出的浏览器基本包含主流浏览器，尽管它们支持某些特征，但更多专注于一维（1D）基因组学
* 未列出的浏览器，如TADkit、HiC3D已经停止更新，且现存最后版本已经无法使用
* WashU、Nucleome Browser、3D Genome Browser等大型组织提供的浏览器为本身组织的数据集服务

四、HiBrowser：三维基因组可视化方法





HTML　JS+CSS

Jquery.js　Tree.js

Cache

Straw.js

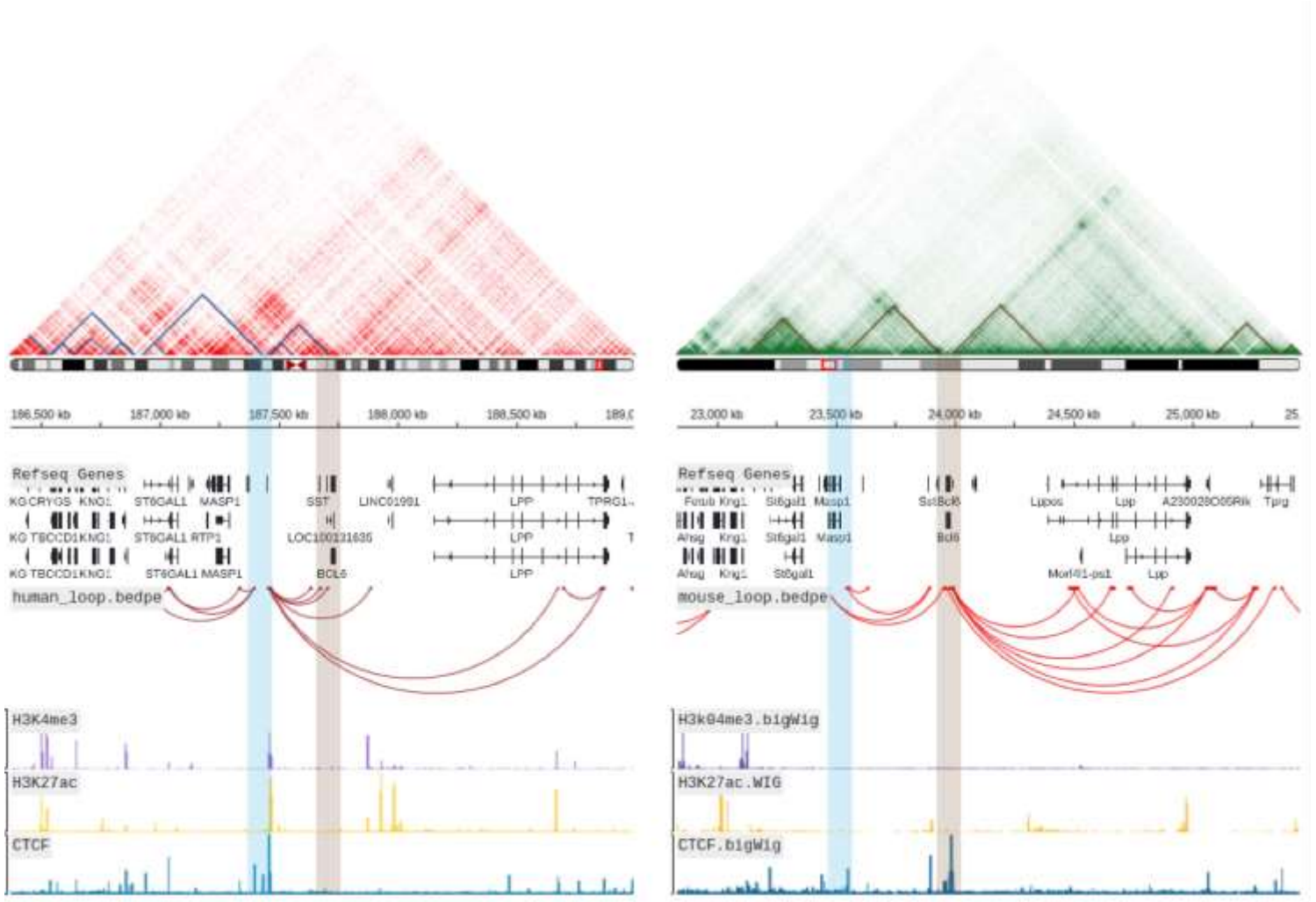
EventBus

Python　SqlLite

Django　Cloud

四、HiBrowser：三维基因组可视化方法

面向个体的一维和三维联合视图



特征

多组学联合分析：
1、上下结构（Hi-C和一维注释）
2、Hi-C热图加载三维空间结构注释
3、同步导航（动态交互、非图片）
4、双向导航（拖动、放大、缩小）

多样本联动分析
1、同时加载多个样本（不限于2个）
2、点击随时选择样本同步
3、比对到不同的参考基因组（坐标对齐）

Drag & Roll交互：
1、所有文件随时加载（本地、网络）
2、本地数据，无延迟（网络数据低延迟）
3、可视化自定义
不限于：
（1）：颜色
（2）：高度
（3）：展示形式
（4）：顺序
4、可交互：（不限于）放大、拖动、点击

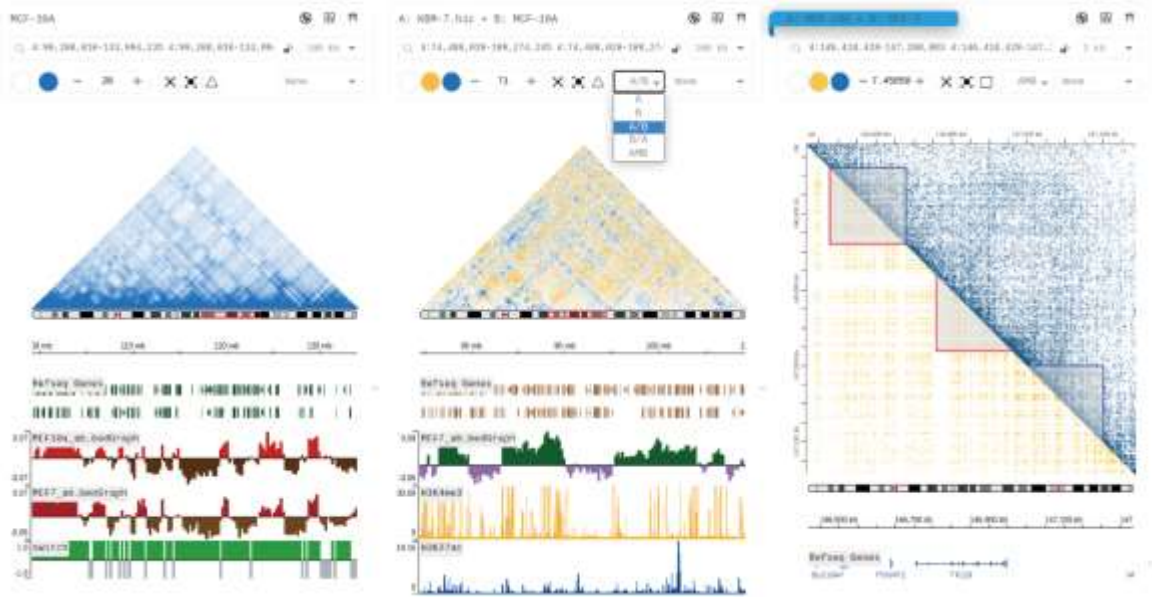
四、HiBrowser：三维基因组可视化方法

多维度差异表示视图

场景一
只加载一个Hi-C数据

场景二
加载两个Hi-C数据并叠加

场景三
加载两个Hi-C数据并对比



快速定位两个样本
交互显著上升或下降位点

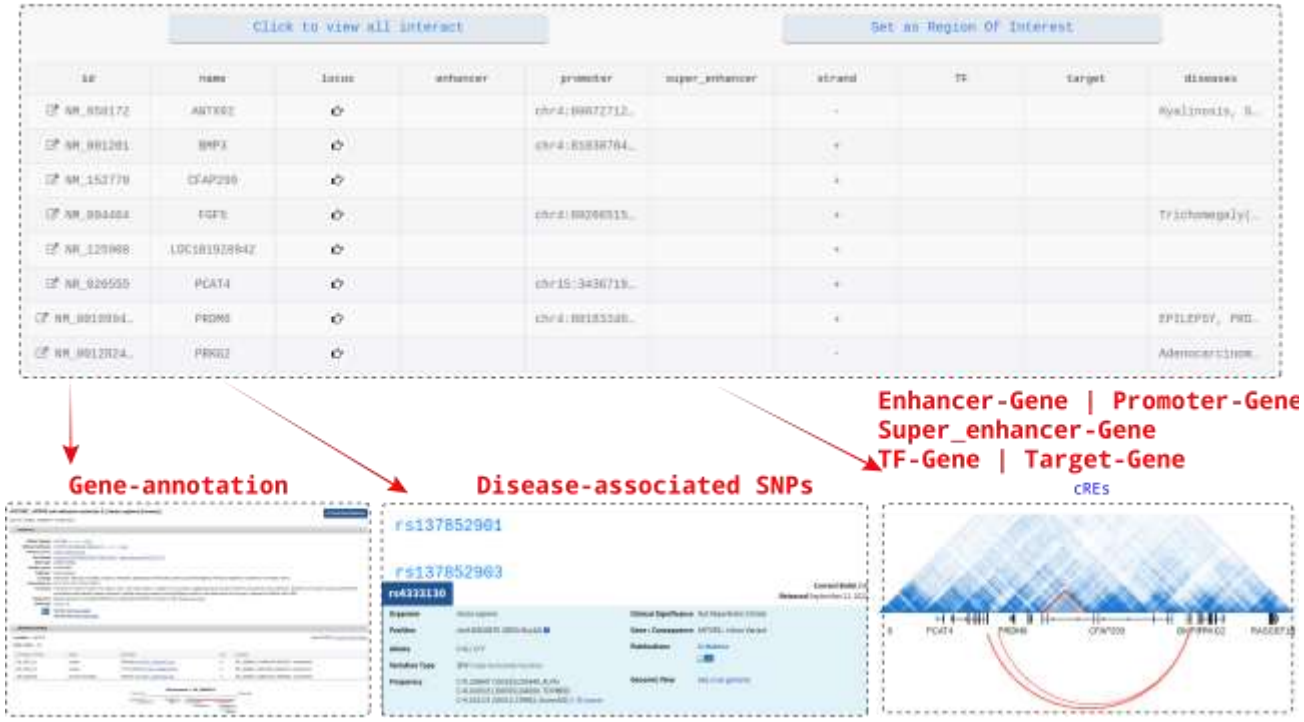
三维结果差异+注释表示视图

1、一键比较样本三维注释结果
2、快速定位A、B区室转变的位点
3、快速比较TAD边界转移位点
4、快速定位Loop锚点转移位点
5、快速定位差异性空间结构内的基因

四、HiBrowser：三维基因组可视化方法

线性远程调控因子的可视化表示

[Locus] 4:80,824,298-82,324,298的搜索结果

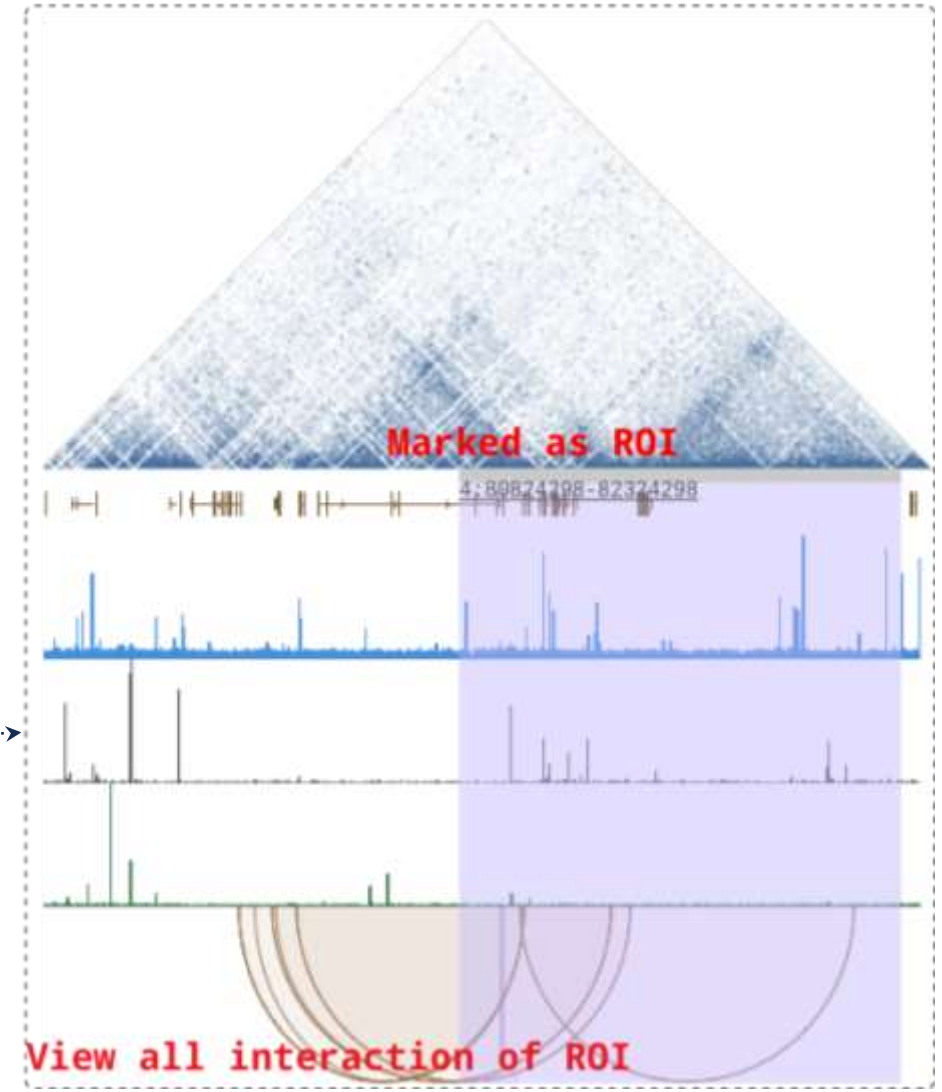


30,170种疾病、
1,167,518个调控因子

三种检索模式：

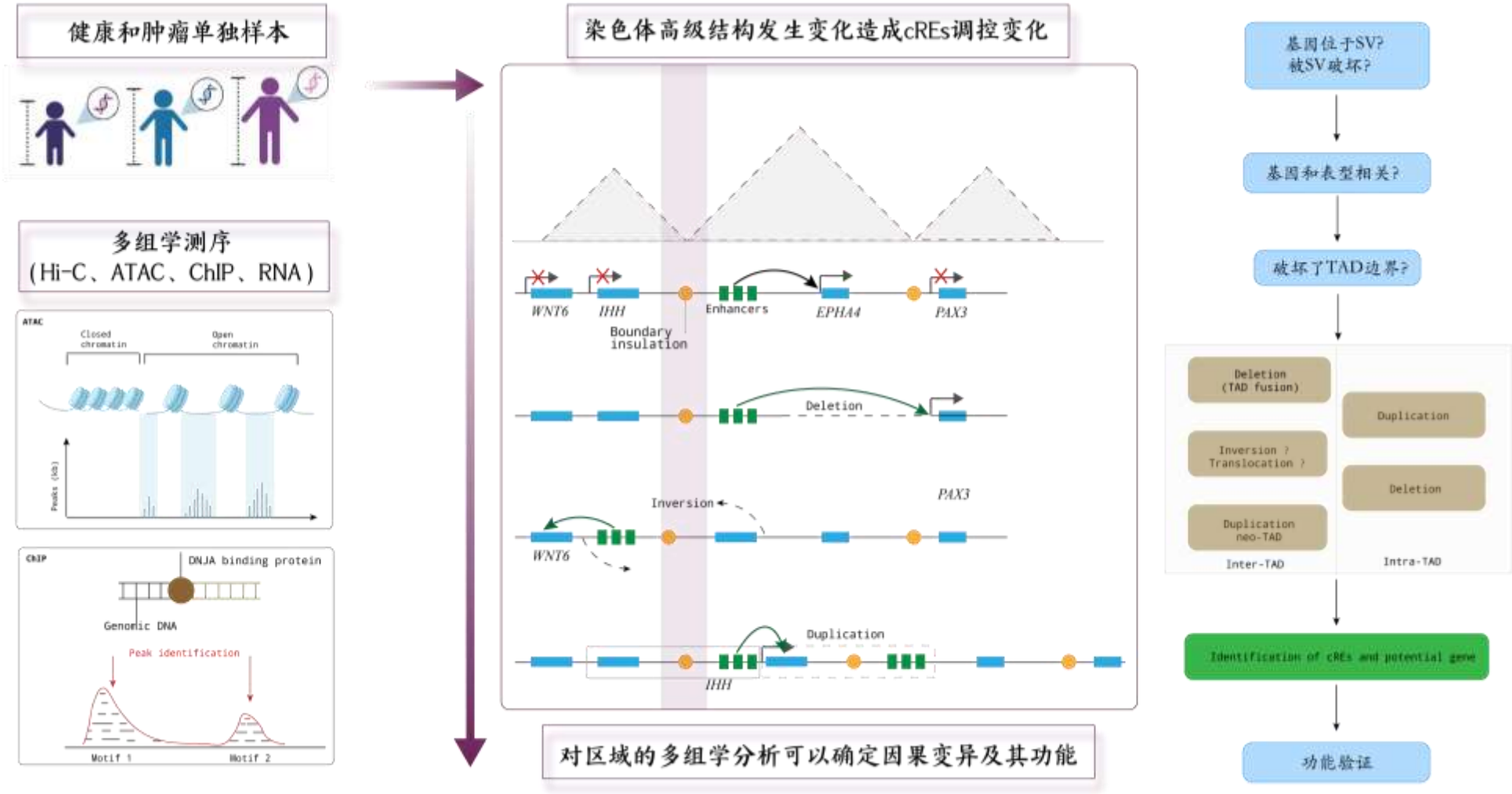
基因Symbol
疾病
染色体坐标

可靠的调控网络

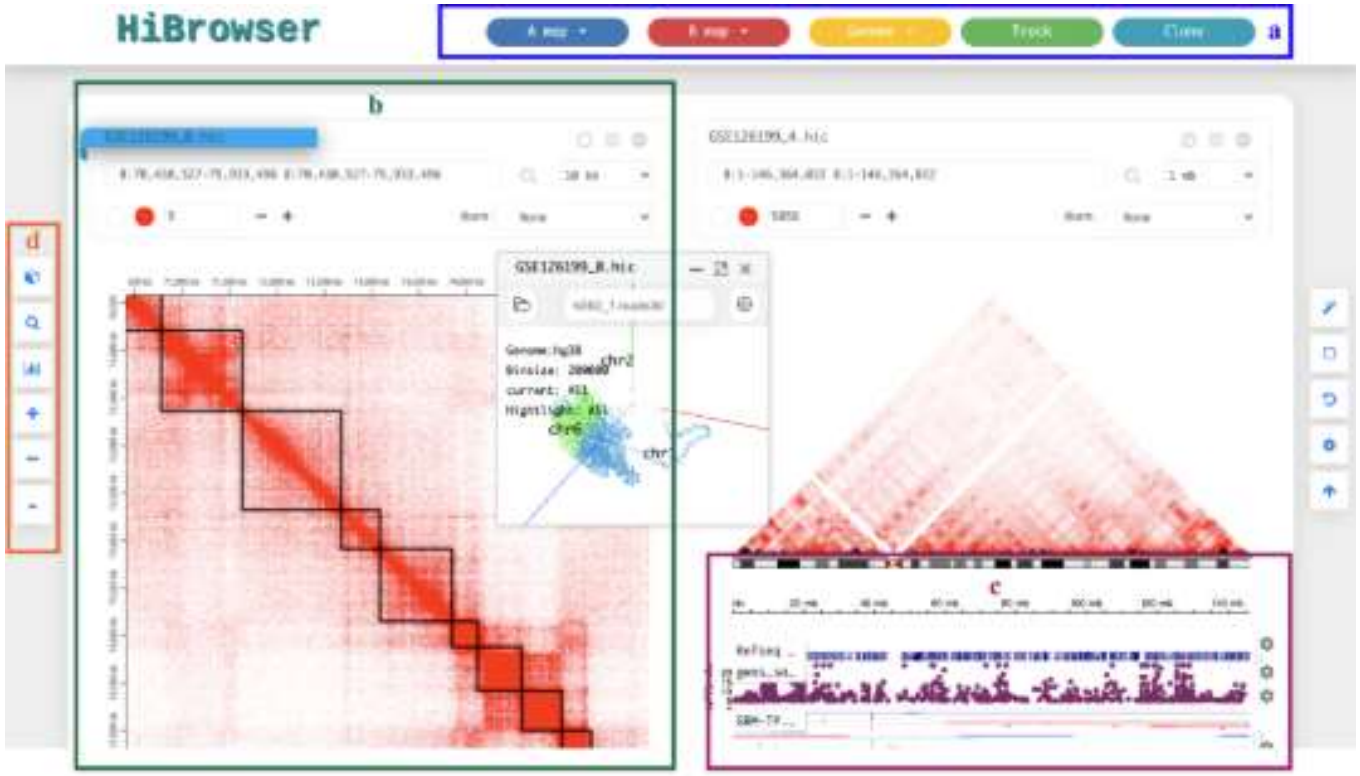


四、HiBrowser：三维基因组可视化方法

进行对照组（如癌症）的研究



HiBrowser页面



HiBrowser支持的注释轨迹类型

轨迹类型	描述
Annotation	非定量基因组注释如基因。这是最通用的轨迹类型
wig	定量基因组数据，如 ChIP 峰和比对覆盖率
alignment	排序、对齐的读对
variant	基因组变异
seg	分段拷贝数数据
mut	突变数据，主要来自癌症研究
interact	代表 2 个基因组基因座之间的关联或相互作用的弧
gwas	全基因组关联数据
arc	RNA 二级结构
junction	RNA 片段连接

本章小结

HiBrowser预设了Hi-C研究中常见的可视化视图和分析流程，使用户能够对多个相关样本进行微观层面的基因调控有意义的可视化分析。该系统有效满足了当前分子生物学研究中的可视化需求，有望加速三维基因组与基因调控领域的研究进程，对表观遗传功能的注释具有重要的影响和意义。

提纲

一、研究背景及研究内容

二、基于深度学习的Hi-C数据增强算法

三、基于Hi-C接触矩阵的TAD识别方法

四、三维基因组可视化方法

五、总结与展望

五、总结与展望

本文主要工作和创新点

- 基于深度学习的Hi-C数据增强算法
  - 重新构建了骨干网络模型
  - 提出了 HiDB 和 HiFM 结构
  - 设计卷积注意力代替自注意力
- 基于Hi-C接触矩阵的TAD识别方法
  - 构建带权无向图
  - 基于标签传播的社区发现
  - 人类组织细胞注释方法
- 三维基因组可视化方法
  - 一维和三维联合可视化表示
  - 多样本联动可视化表示
  - 差异可视化表示

本文不足及研究展望

1. 以HiFM/LKA模块为基础，嵌入先前方法中，探索混合模型的性能
2. 讨论[1]至少需要多少接触片段数开始才可以获得高质量的 Hi-C 数据
3. 对图社区发现算法的加速同样是值得探索的方向之一
4. 在不同的互作层级上构建三维结构图谱[2]，绘制完整的基因进化树

[1]: 本文对特定末端配对序列进行下采样预训练了9个模型
[2]: HiBrowser提出的层级结构差异性分析算法目前只预设了统计分析策略

THANK YOU FOR ATTENTION

Q&A

1、Liu, J.*, **Li, P.**, Sun, J., & Guo, J. (2023). LPAD: using network construction and label propagation to detect topologically associating domains from Hi-C data. *Briefings in Bioinformatics*, **24**(3), bbad165.

2、**Li, P.**, Liu, H., Sun, J., Lu, J., & Liu, J*. (2023). HiBrowser: an interactive and dynamic browser for synchronous Hi-C data visualization. *Briefings in Bioinformatics*, **24**(5), bbad283.

3、**Li, P.**, Guo, J., Feng, J., & Liu, J*. (2024). HiADN: Lightweight resolution enhancement of Hi-C data using high information attention distillation network. Submit to *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

报 告 人：李平静
导　　师：刘健 教授

2024年05月