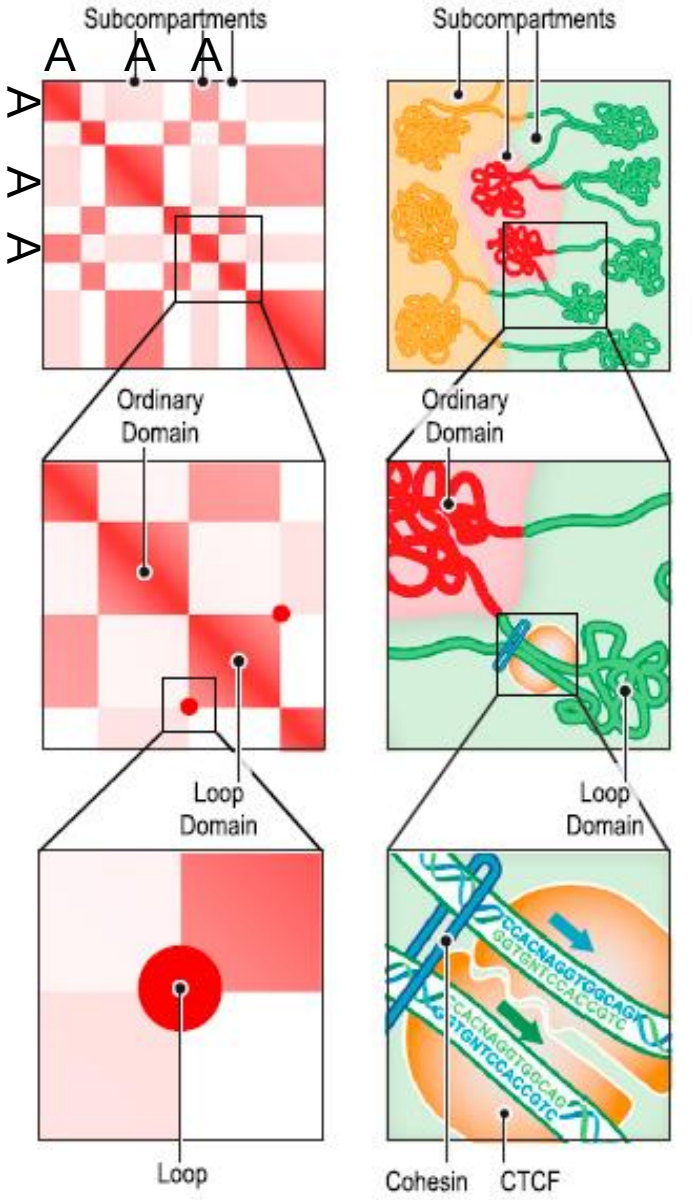
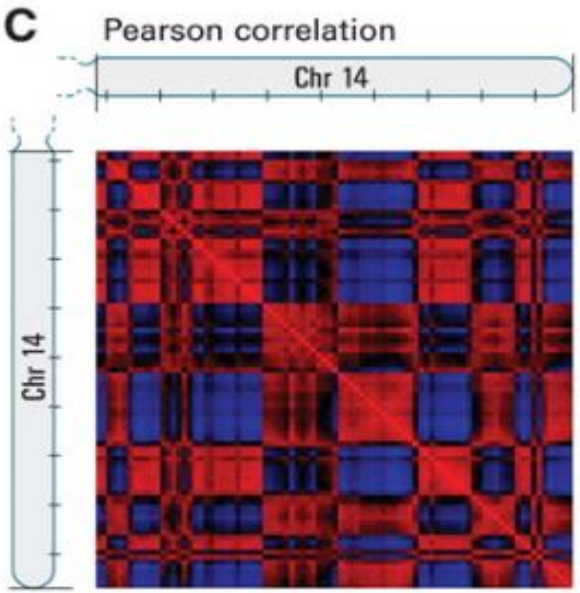
From Structural Feature to Biological Functions

Compartment





Structural Feature

Individual 1 Mb loci could be assigned to one of two long-range contact patterns, which we called compartments A and B, with loci in the same compartment showing more frequent interaction

Biological Functions

Compartment A is highly enriched for open chromatin; compartment B is enriched for closed chromatin.

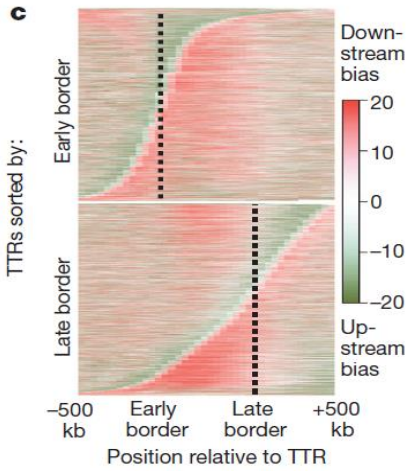
TAD

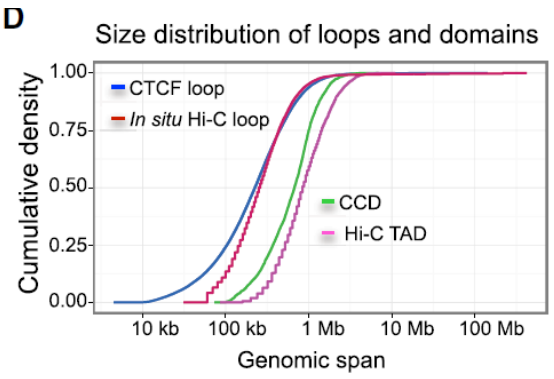
Structural Feature

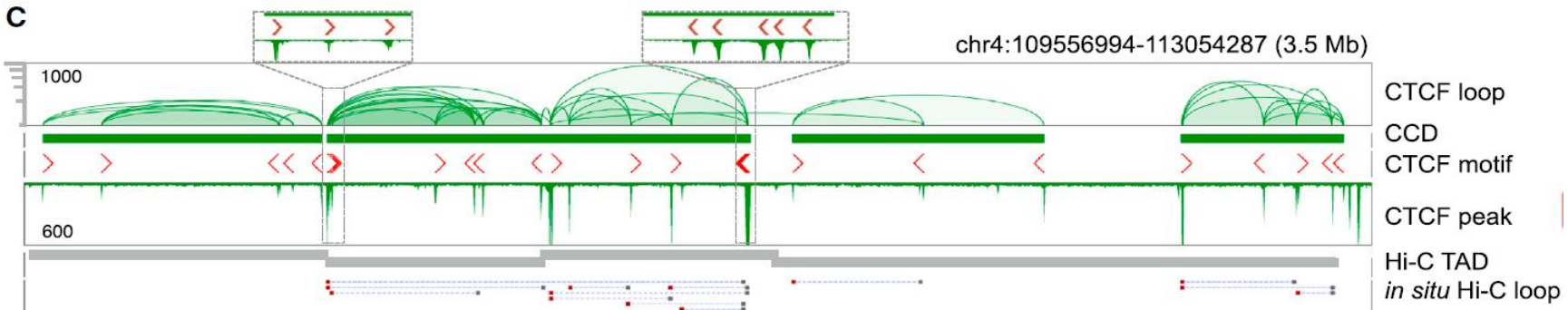
1. intra-domain chromatin interactions are significantly stronger than inter-domain interactions
2. 1M (Dixon et al., 2012)

Biological Functions

1. the genomic positions of TADs appear to be stable across cell types and conserved across species in mammals
2. provides structural basis for chromatin regulation：most identified enhancer-promoter interactions were located in the same TADs
3. TADs resembles chromatin contact domains (CCDs)
4. Topologically associating domains are stable units of replication-timing regulation







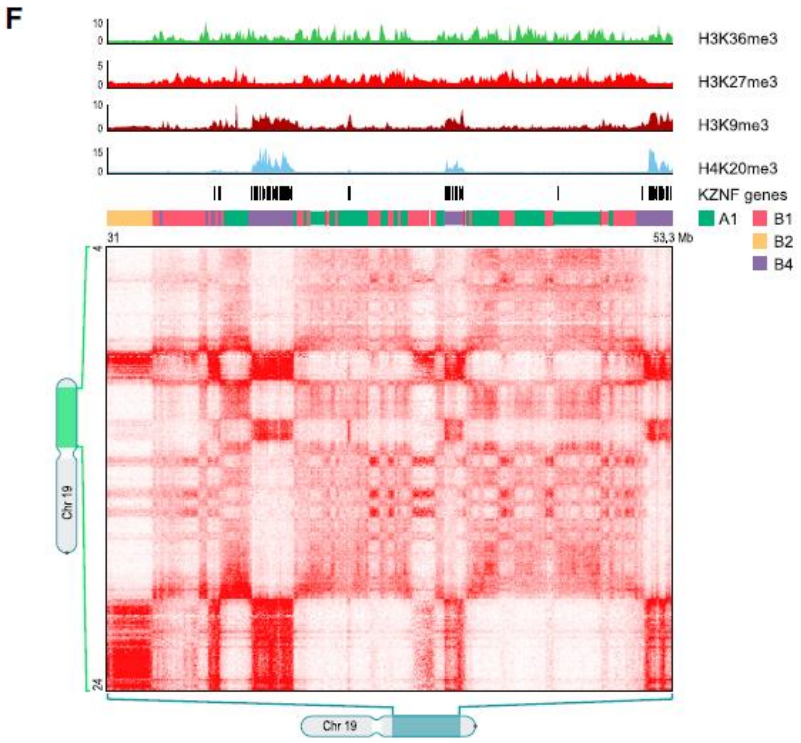
Contact domains and subcompartment

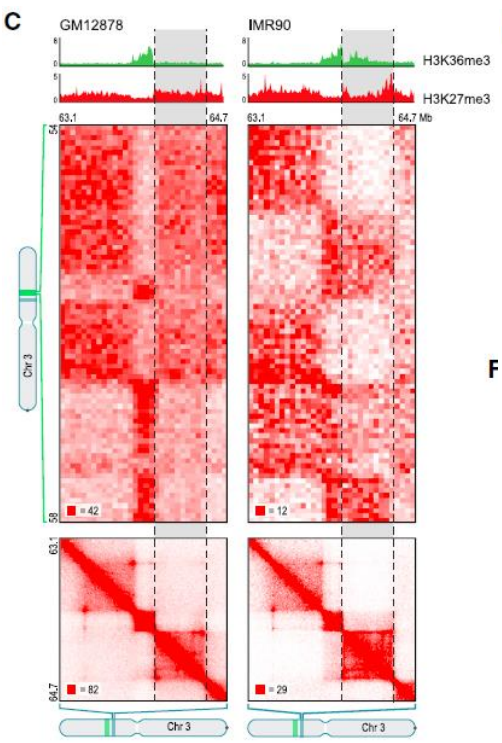
Structural Feature

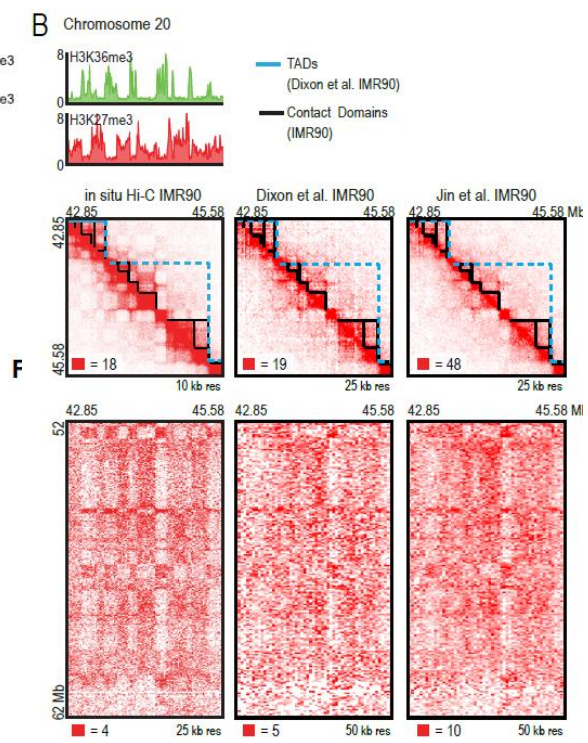
1. intra-domain chromatin interactions are significantly stronger than inter-domain interactions
2. smaller than TADs, about 185kb in Rao et all.

Biological Functions

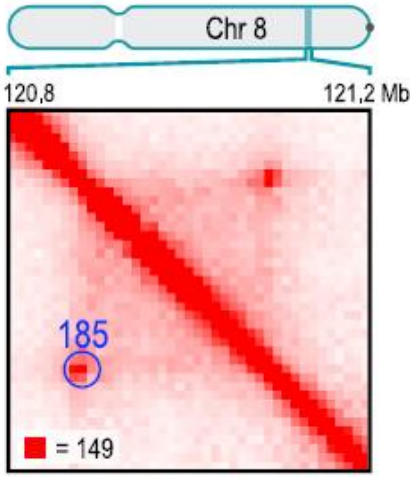
1. Contact Domains Exhibit Consistent Histone Marks Whose Changes Are Associated with Changes in Long-Range Contact Pattern.
2. nearly all the boundaries we observe are associated with either a subcompartment Transition（300kb） or a loop(200kb).



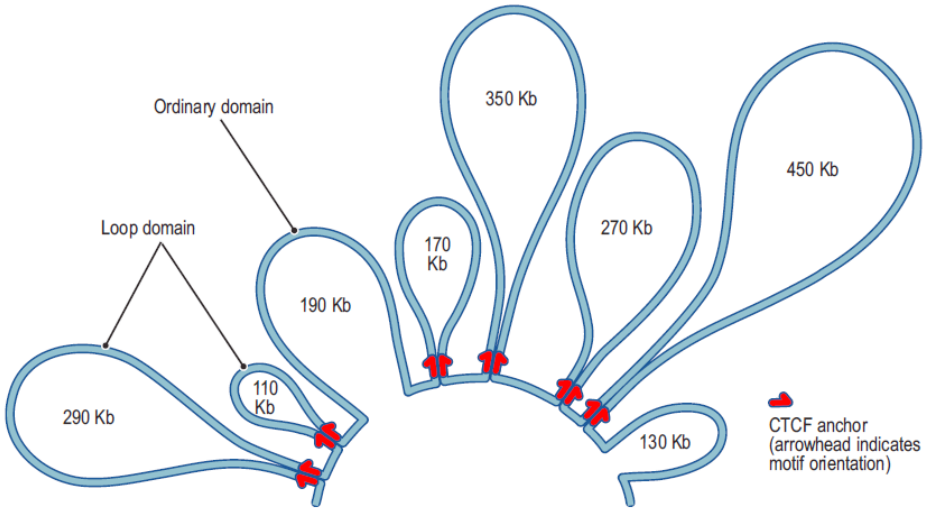




Loop



F



Structural Feature

the peak pixel is enriched as compared to other pixels in its neighborhood.

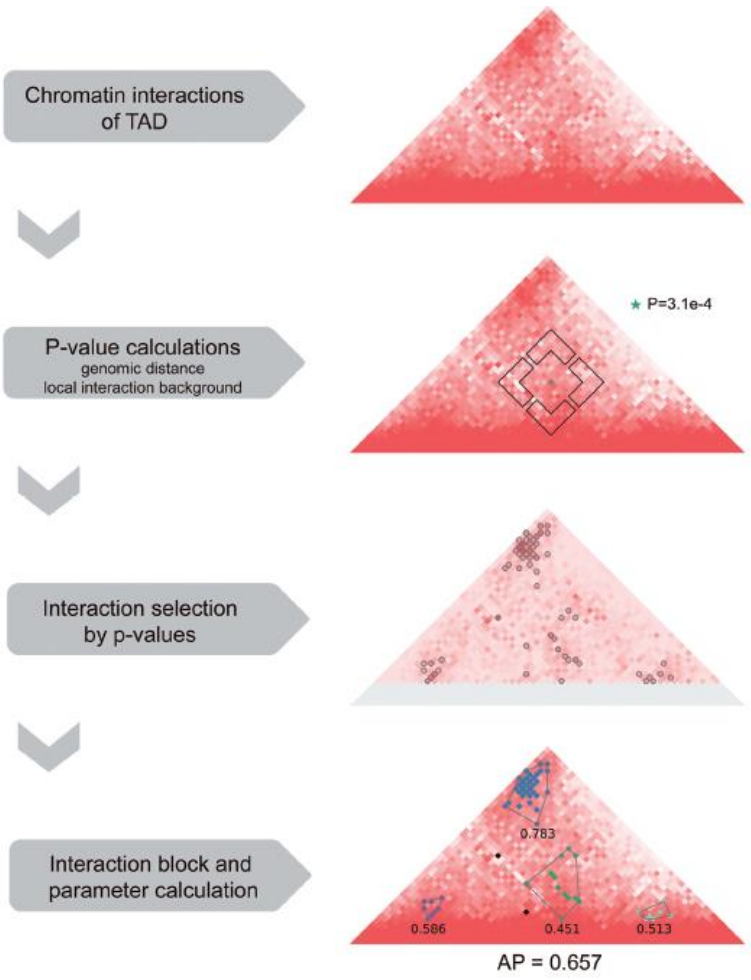
Biological Functions

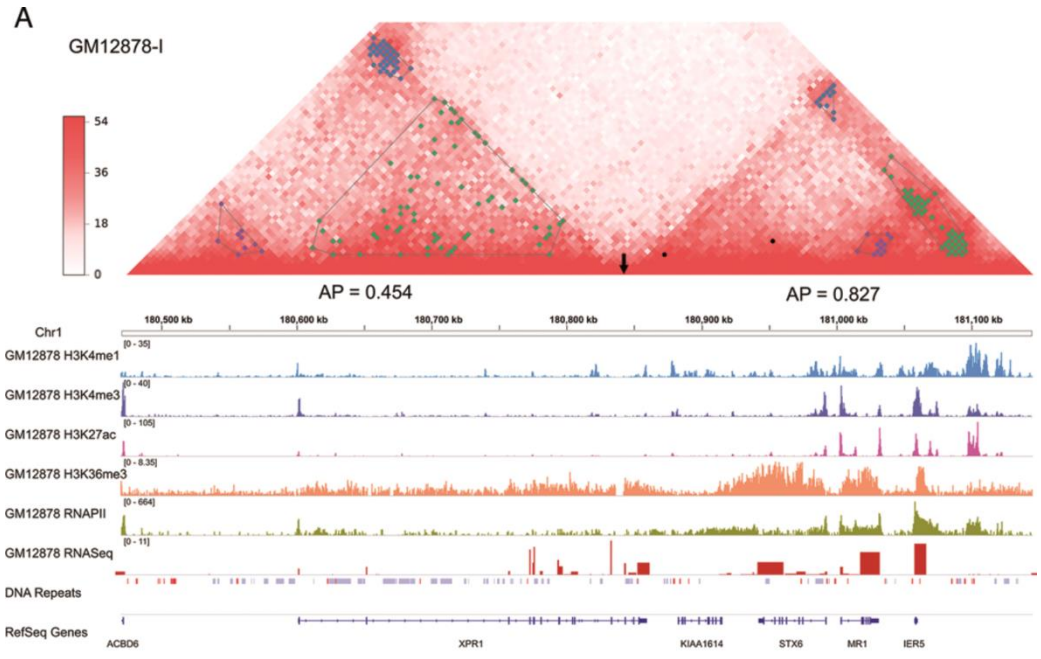
1.Most loops are short (<2 Mb) and strongly conserved across cell types and between human and mouse.
2.Loops Anchored at a Promoter Are Associated with Enhancers and Increased Gene Activation.
3. Loops Frequently Demarcate the Boundaries of Contact Domains
4.CTCF and the cohesin subunits RAD21 and SMC3 associate with loops; each of these proteins is found at over 86% of loop anchors.

Beyond TAD, loop, compartment

aggregation preference (AP)

- quantitatively measure the chromatin interaction patterns of TADs

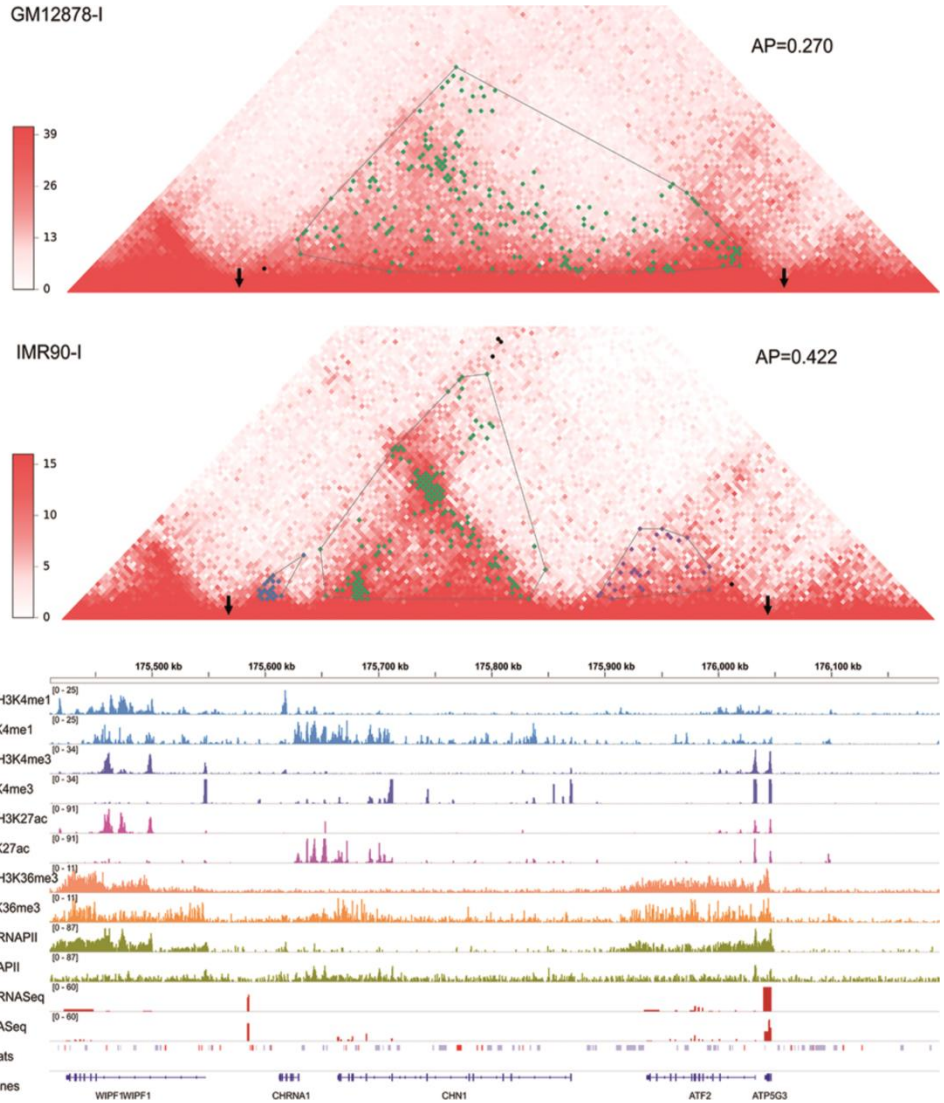




B

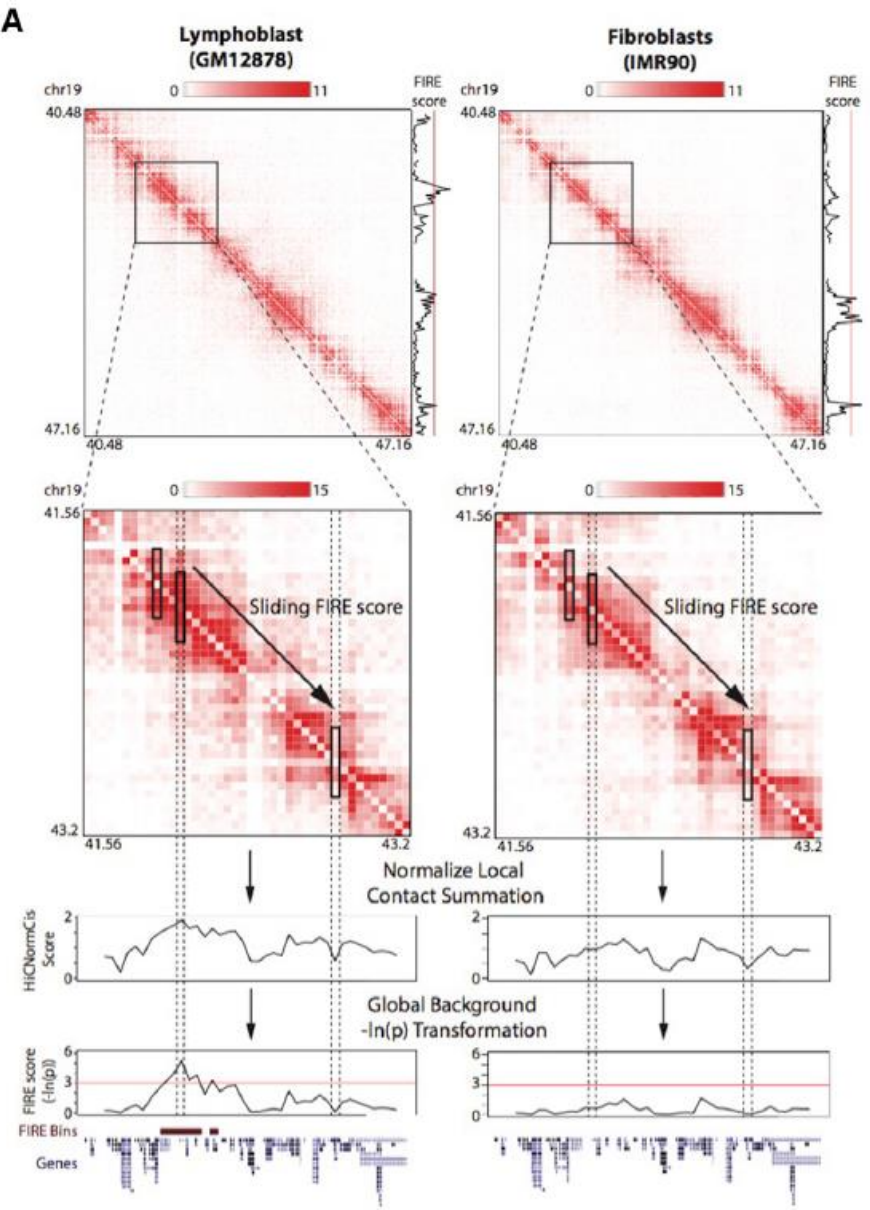
	GC-Content	TSS	SINE	LINE	LaminB1	H3K4me1	H3K4me3	H3K27ac	H3K36me3	RNAPII	RNASeq
GM12878-I (5K)	0.46 (3.00e-15)	0.39 (0)	0.39 (1.00e-11)	-0.26 (2.90e-13)		0.56 (0)	0.54 (0)	0.44 (0)	0.45 (0)	0.45 (0)	0.32 (0)
IMR90-I (5K)	0.5 (0)	0.41 (0)	0.4 (4.50e-13)	-0.29 (1.10e-16)		0.53 (0)	0.54 (0)	0.46 (0)	0.52 (0)	0.34 (0)	0.21 (7.50e-11)
K562-I (5K)	0.54 (1.10e-16)	0.39 (2.10e-15)	0.43 (3.10e-11)	-0.27 (3.30e-11)		0.53 (0)	0.46 (0)	0.47 (0)	0.37 (0)	0.5 (0)	0.32 (1.90e-15)
GM12878-I (10K)	0.46 (8.50e-14)	0.41 (0)	0.37 (3.60e-11)	-0.26 (3.60e-13)		0.55 (0)	0.52 (0)	0.41 (0)	0.5 (0)	0.43 (0)	0.33 (0)
IMR90-I (10K)	0.48 (0)	0.42 (0)	0.38 (2.20e-16)	-0.29 (0)		0.54 (0)	0.52 (0)	0.44 (0)	0.54 (0)	0.37 (0)	0.18 (6.30e-08)
K562-I (10K)	0.55 (1.10e-16)	0.42 (0)	0.43 (1.20e-10)	-0.28 (6.30e-12)		0.57 (0)	0.5 (0)	0.51 (0)	0.44 (0)	0.51 (0)	0.36 (0)
HMEC-I (10K)	0.44 (9.90e-14)	0.38 (0)	0.35 (2.00e-09)	-0.26 (1.60e-11)		0.53 (0)	0.46 (0)	0.42 (0)	0.46 (0)		0.18 (5.40e-05)
HUVEC-I (10K)	0.49 (7.40e-15)	0.42 (0)	0.38 (1.40e-10)	-0.3 (2.10e-15)		0.55 (0)	0.5 (0)	0.48 (0)	0.5 (0)	0.49 (0)	0.29 (1.60e-11)
NHEK-I (10K)	0.37 (6.10e-10)	0.34 (1.10e-15)	0.31 (1.30e-08)	-0.22 (6.30e-09)		0.46 (0)	0.43 (0)	0.38 (0)	0.41 (0)	0.39 (0)	0.16 (2.30e-05)
hESC-T (20K)	0.3 (5.60e-07)	0.28 (4.50e-10)	0.26 (6.40e-08)	-0.19 (9.50e-07)		0.34 (1.70e-12)	0.34 (2.50e-14)	0.36 (4.40e-16)	0.32 (4.00e-12)	0.31 (4.30e-13)	0.14 (1.40e-03)
GM12878-T (20K)	0.23 (1.60e-02)	0.26 (8.80e-05)	0.23 (1.80e-02)	-0.14 (3.90e-02)		0.44 (1.90e-14)	0.4 (6.80e-14)	0.31 (3.30e-07)	0.39 (9.50e-12)	0.34 (1.30e-08)	0.26 (3.80e-06)
IMR90-T (20K)	0.39 (4.00e-06)	0.31 (7.00e-08)	0.29 (2.20e-05)	-0.28 (3.10e-08)		0.58 (0)	0.42 (1.10e-16)	0.47 (0)	0.5 (0)	0.35 (1.10e-13)	0.14 (1.90e-03)
mESC-T (20K)	0.42 (3.70e-10)	0.26 (1.40e-04)	0.31 (5.20e-08)	-0.39 (2.40e-11)	-0.45 (1.20e-15)	0.43 (1.30e-13)	0.34 (6.40e-09)	0.4 (2.70e-15)	0.42 (1.40e-15)	0.44 (1.90e-13)	0.2 (2.30e-04)
Cortex-T (20K)	0.25 (4.10e-04)	0.065 (0.14)	0.11 (2.90e-02)	-0.24 (3.20e-05)		0.34 (1.00e-10)	0.17 (1.90e-03)	0.31 (2.40e-09)		0.23 (1.20e-05)	0.09 (4.80e-02)

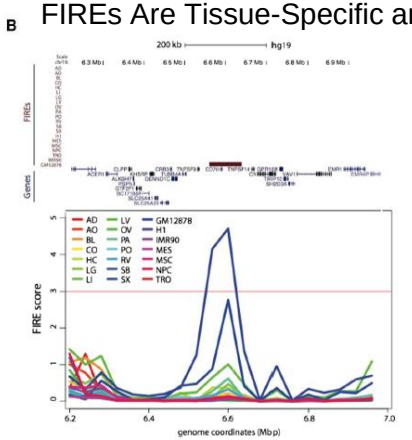
aggregation preference （AP）

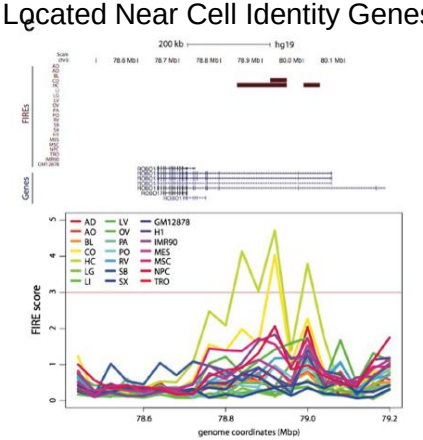


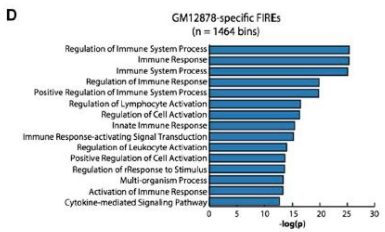
		H3K4me1	H3K4me3	H3K27ac	H3K36me3	RNAPII	RNASeq
GM12878-I IMR90-I	(5K)	2.1e-07	2.0e-07	2.6e-07	4.3e-11	0.21	8.8e-04
		6.4e-10	1.8e-03	9.6e-10	1.3e-03	2.1e-11	6.2e-06
GM12878-I K562-I	(5K)	6.5e-04	1.3e-02	1.4e-05	2.2e-02	1.2e-03	1.4e-03
		7.4e-08	6.7e-06	6.0e-10	7.2e-07	1.2e-08	4.8e-08
IMR90-I K562-I	(5K)	4.7e-08	7.4e-02	1.1e-06	0.28	3.7e-11	7.5e-03
		1.1e-07	1.5e-06	3.1e-05	2.4e-07	8.8e-03	0.42
GM12878-I IMR90-I	(10K)	3.3e-05	2.2e-04	1.4e-02	4.0e-06	0.35	0.11
		1.2e-17	2.7e-08	5.0e-16	2.1e-11	1.3e-04	2.1e-14
GM12878-I K562-I	(10K)	6.1e-09	2.7e-06	2.9e-06	1.7e-03	0.32	4.6e-05
		3.7e-13	1.1e-13	3.1e-09	9.3e-11	5.7e-06	3.4e-07
IMR90-I K562-I	(10K)	2.0e-10	0.46	7.5e-07	0.43	4.3e-15	1.2e-05
		5.9e-16	4.9e-08	2.8e-11	9.2e-12	1.1e-07	0.47
GM12878-I HMEC-I	(10K)	3.7e-02	4.7e-03	4.6e-02	8.7e-02		3.0e-02
		2.6e-11	6.7e-07	8.6e-07	1.7e-08		2.6e-10
GM12878-I HUVEC-I	(10K)	2.6e-02	5.5e-02	1.7e-02	3.0e-02	0.22	5.8e-02
		1.5e-07	5.7e-04	1.9e-06	8.7e-09	5.6e-03	3.2e-09
GM12878-I NHEK-I	(10K)	0.17	0.28	1.9e-02	0.21	0.44	7.1e-02
		1.1e-03	1.4e-03	5.6e-04	1.9e-03	2.2e-04	1.3e-08
IMR90-I HMEC-I	(10K)	1.5e-04	0.16	1.1e-04	0.22		1.1e-04
		2.2e-14	7.0e-07	6.2e-04	1.9e-07		0.27
IMR90-I HUVEC-I	(10K)	4.1e-03	0.28	6.2e-03	1.3e-02	3.1e-06	3.2e-04
		1.2e-10	1.1e-09	7.2e-05	5.7e-11	0.26	2.2e-02
IMR90-I NHEK-I	(10K)	2.0e-02	0.22	1.5e-03	0.37	6.3e-07	1.5e-03
		1.4e-14	6.9e-10	2.9e-10	8.4e-13	0.20	9.7e-02
K562-I HMEC-I	(10K)	7.1e-06	3.8e-07	1.3e-06	5.2e-07		8.0e-02
		2.0e-05	2.7e-02	5.1e-07	1.7e-03		1.2e-03
K562-I HUVEC-I	(10K)	2.2e-07	1.5e-08	4.3e-09	7.6e-09	2.6e-04	2.0e-02
		8.3e-12	1.3e-07	7.7e-11	5.8e-05	2.4e-08	1.6e-06
K562-I NHEK-I	(10K)	4.9e-04	7.7e-04	2.4e-04	4.1e-04	1.6e-02	0.20
		5.0e-08	6.2e-02	1.0e-05	0.16	5.0e-07	2.5e-02
HMEC-I HUVEC-I	(10K)	3.7e-06	2.7e-02	7.4e-04	7.5e-05		2.6e-02
		3.6e-03	8.9e-03	2.9e-06	7.9e-03		3.0e-04
HMEC-I NHEK-I	(10K)	0.22	0.44	4.3e-02	0.15		4.3e-03
		0.25	0.27	0.23	2.0e-02		0.34
HUVEC-I NHEK-I	(10K)	2.0e-02	0.24	2.1e-03	1.2e-02	1.9e-02	2.1e-03
		3.7e-03	0.15	6.0e-03	0.48	6.9e-02	8.4e-02
hESC-T GM12878-T	(20K)	1.1e-05	1.1e-02	8.0e-04	7.2e-04	1.1e-02	7.9e-03
		8.7e-02	0.29	2.8e-03	0.43	0.18	1.1e-03
hESC-T IMR90-T	(20K)	1.0e-08	9.1e-07	2.6e-06	2.1e-10	1.6e-02	0.12
		0.38	6.6e-02	9.4e-02	0.35	0.42	3.7e-04
GM12878-T IMR90-T	(20K)	2.9e-03	3.3e-04	6.8e-02	6.9e-04	0.25	0.21
		2.9e-05	8.7e-05	1.1e-05	3.7e-04	9.8e-02	1.4e-04
mESC-T Cortex-T	(20K)	1.3e-02	7.1e-05	2.0e-05		0.30	8.2e-02
		0.45	2.4e-02	5.5e-02		0.13	0.25

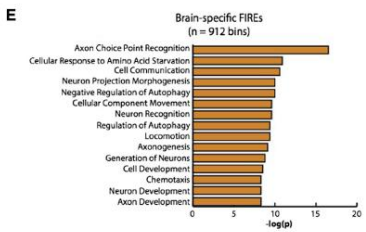
FIRE



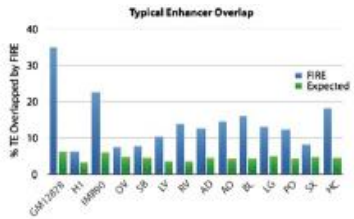


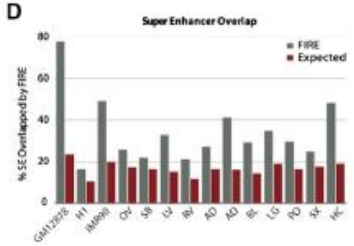




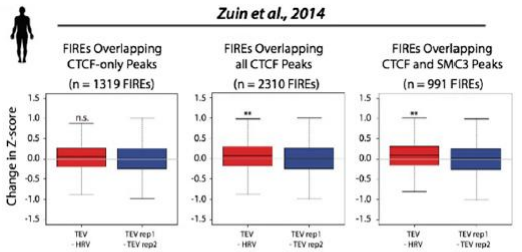


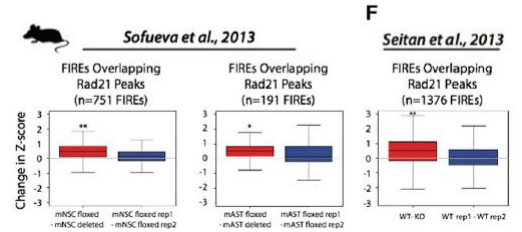
FIREs Are Enriched for Active Enhancers and Super-Enhancers





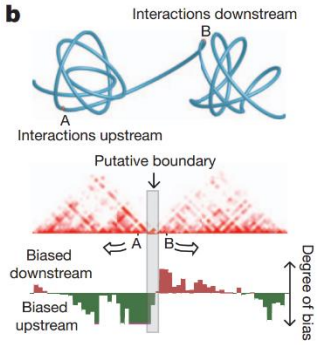
CTCF and Cohesin Complex Contribute to Establishment of FIREs



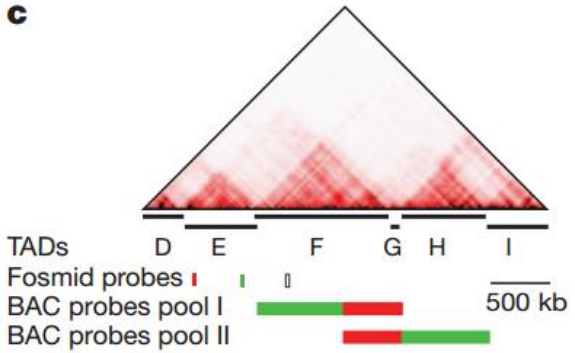


background

- **拓扑结构域（TAD）**
  - **TAD概念的提出**

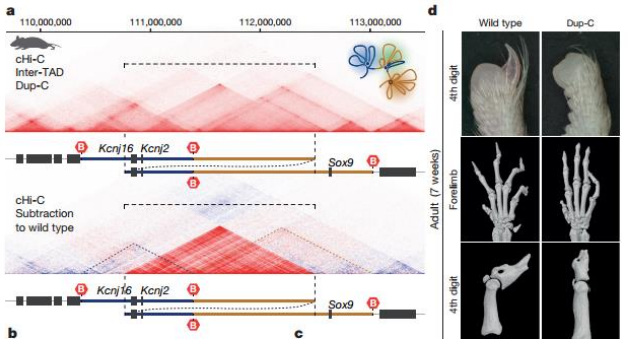


Dixon et al. Nature 2012 (**Bing Ren Lab**)



Nora et al. Nature 2012 (**Edith Heard lab & Job Dekker Lab**)

  - **TAD与基因调控、疾病形成等密切相关**
  - **划分TAD的计算方法**
    - DI， 隐马模型， Nature 2012
    - Artmus， 动态规划， Algorithms Mol. Biol. 2014
    - HicSeg， 动态规划， Bioinformatics 2014
    - CHDF， 动态规划， QB 2015
    - Insulation score,
    - Spectral， 谱分解， Bioinformatics 2016
    - TopDom， 一维信号谱， NAR 2016
    - **TADtree, Bioinformatics 2016**
    - **IC-finder, NAR,2017**
    - **Arrowhead cell,2014**



Franke et al. Nature 2016

content

- 1. HiCTAD method
- 2. Comparative analysis of domain boundary detecting methods
- 3. Application
- 4. problems

Figure1. HiCTAD method

- **构造相对绝缘谱**

$$S(i,w) = \frac{U(i,w) + D(i,w) - B(i,w)}{U(i,w) + D(i,w) + B(i,w)}$$

给定位置i，可选取不同大小的划窗w

- **计算平均相对绝缘谱及其峰值**

$$\mathrm{m}S(i) = \frac{1}{T+1}\sum_{w} S(i,w);\ \ w = 3, \ldots, 3+T$$

$$\mathrm{L} = \{\, i \mid \mathrm{mS}\ \text{信号的峰值点}\ i \,\}$$

- **寻找局部峰值**

  二次下包络

- **选取cut off**

  GSEA-like method

  CTCF motif file

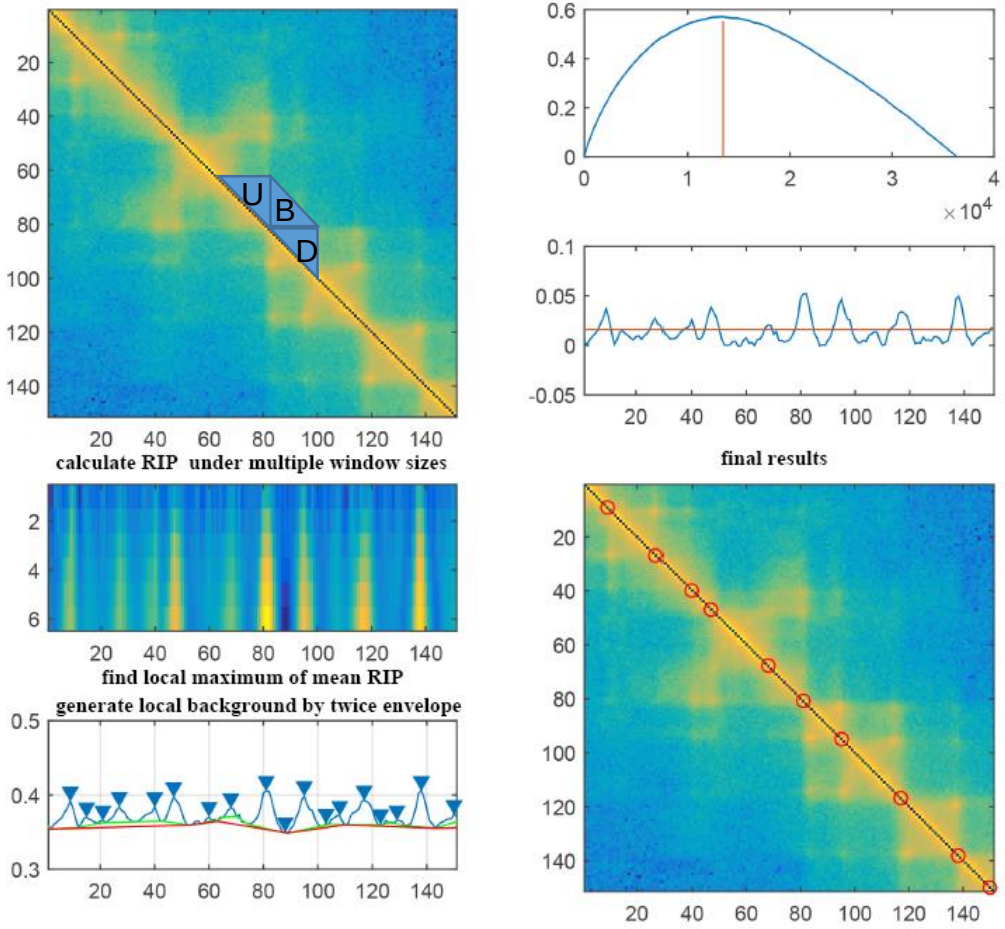
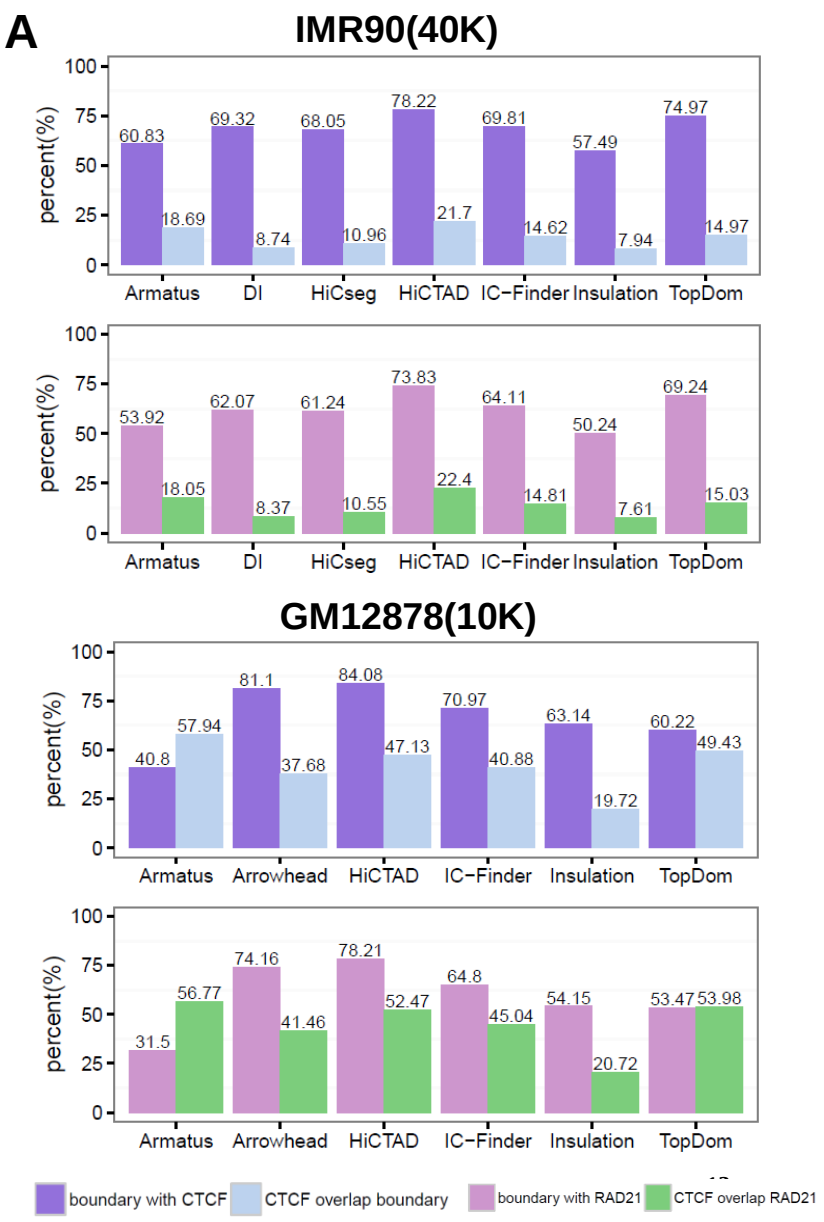


Figure2. Architectural and regulatary elements enrich on HiCTAD detected boundary

Architectural protein enrichment of different methods



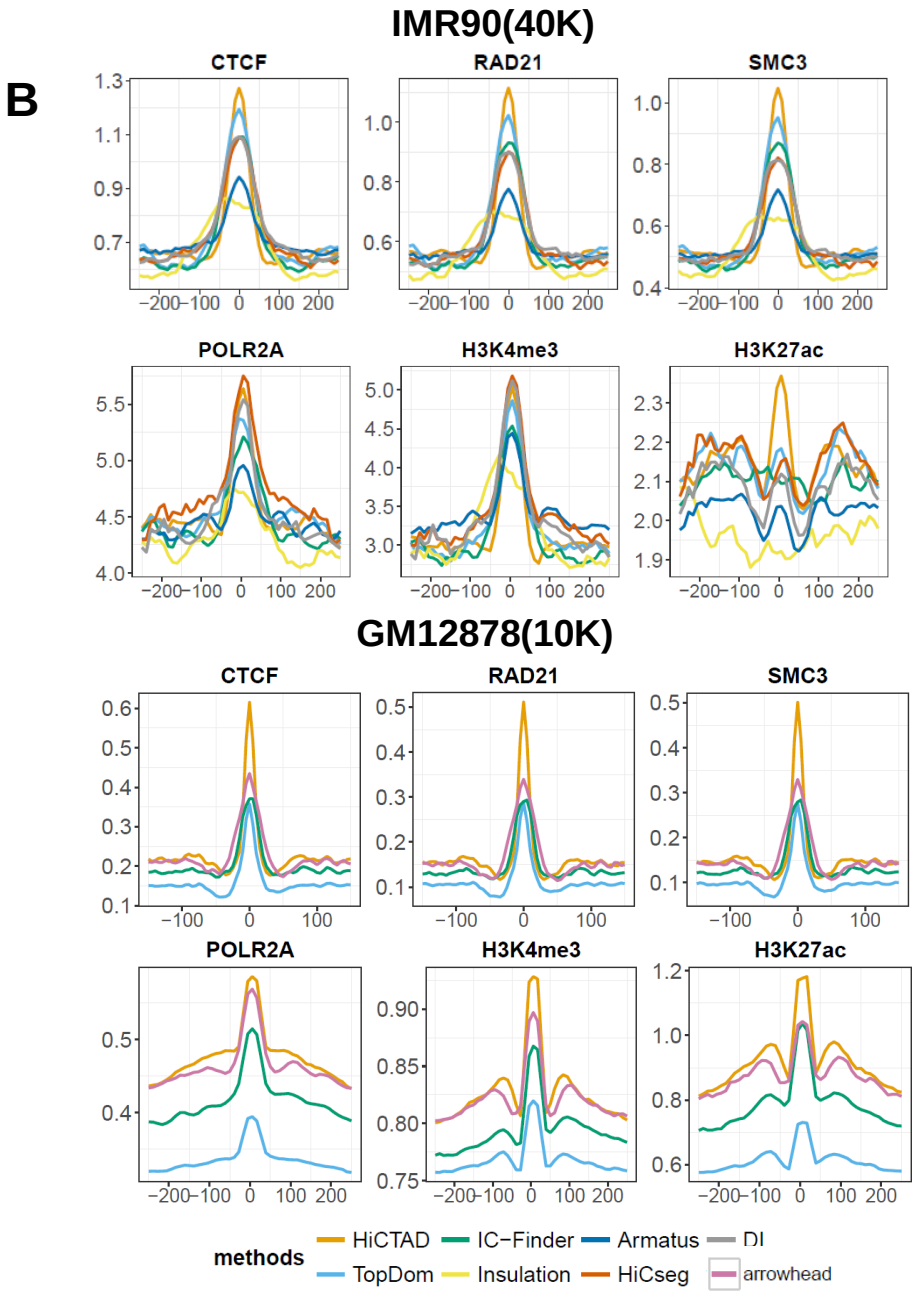


Figure3. HiCTAD can detect finer structure

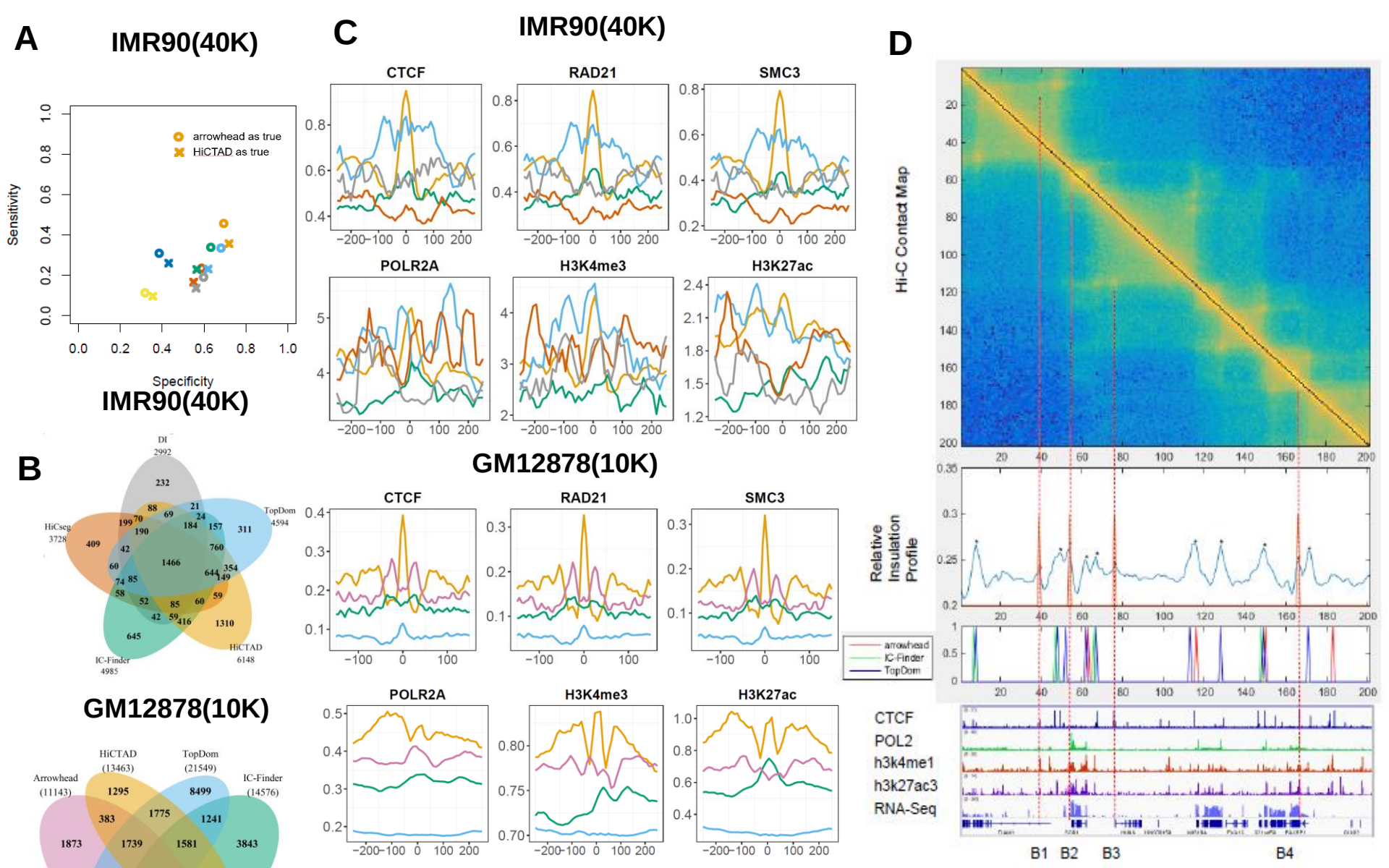
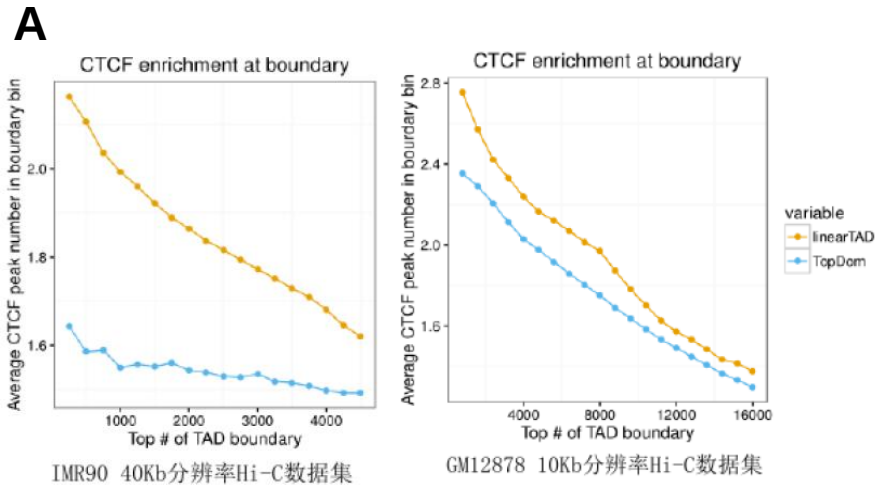
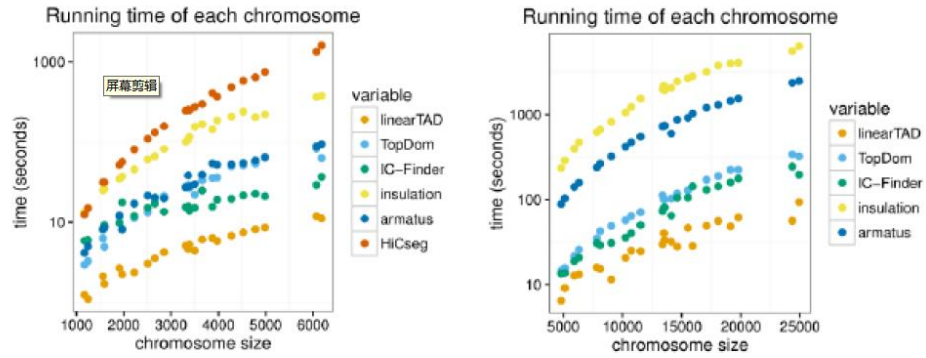
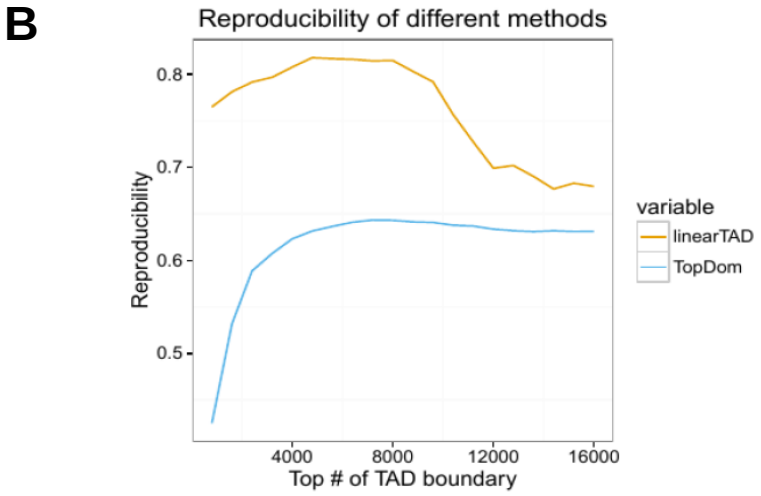


Figure4. HiCTAD is robust and fast







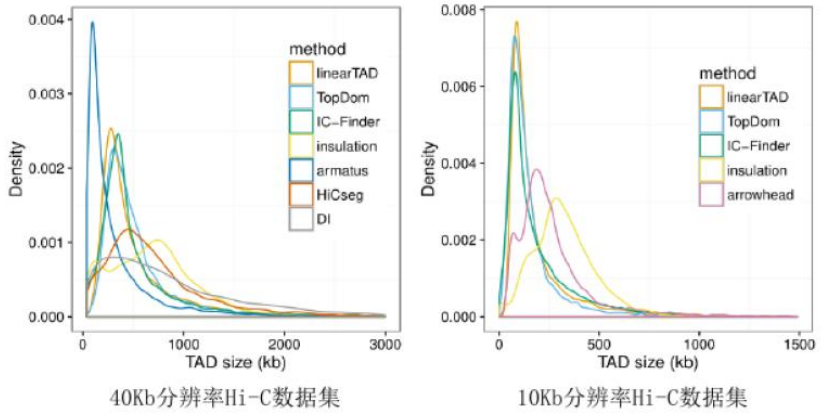
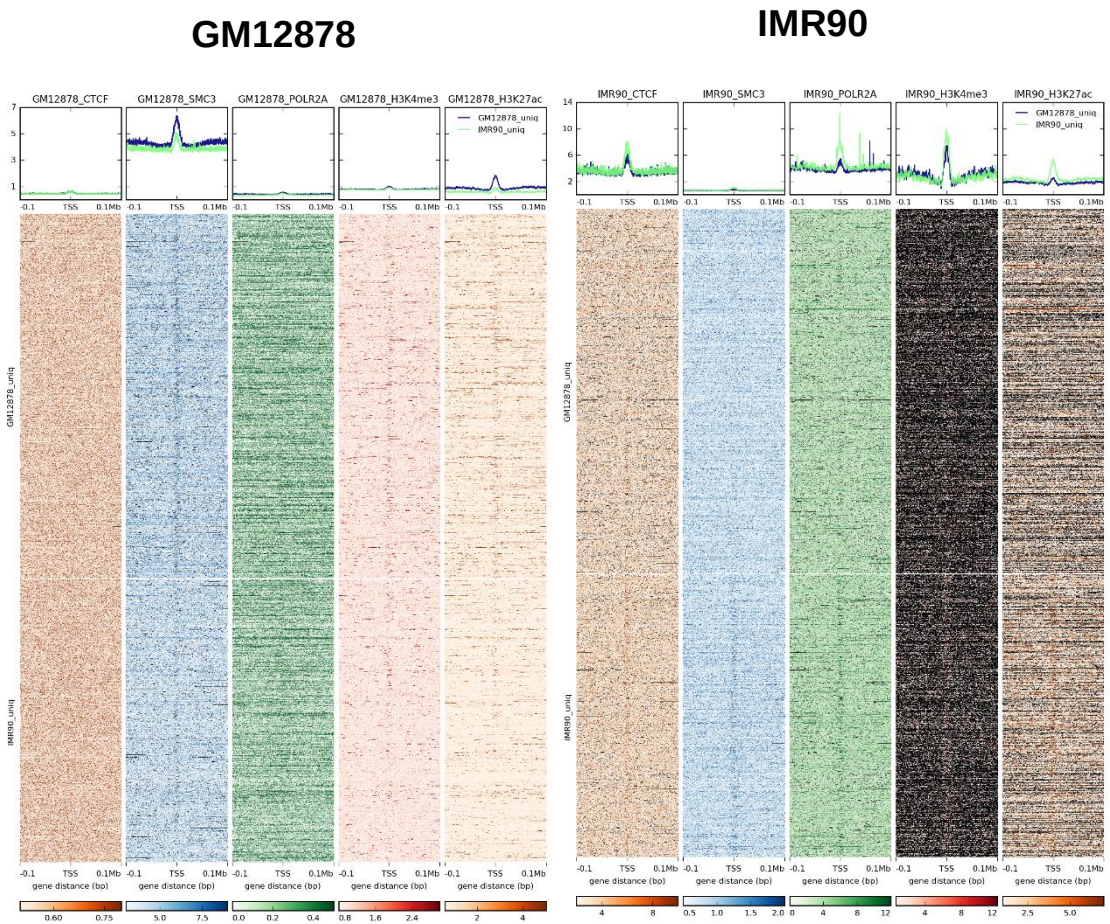


图 4.11　不同方法检测的拓扑结构域的大小

Figure5. HiCTAD facilitate differentical domain boundary detection



cell-specific boundary enrich cell-specific H3K27ac and SMC3,POL2A signal

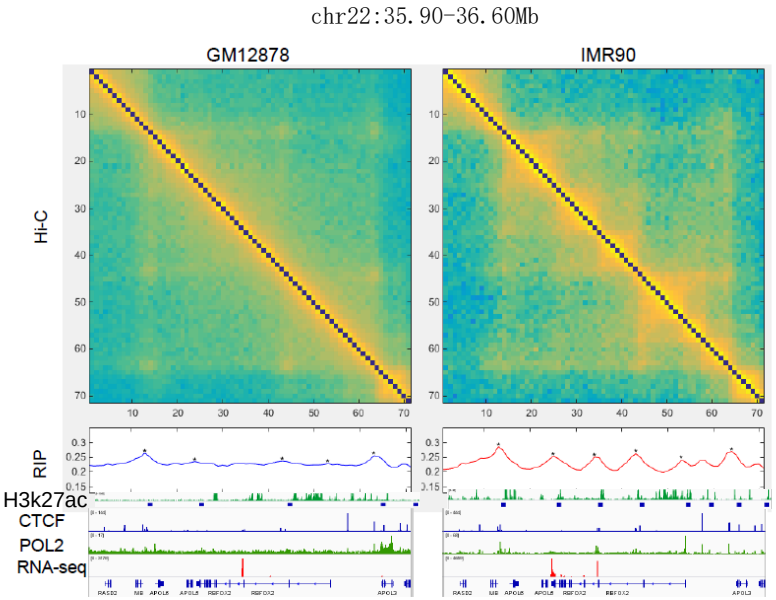


图 4.14　差异的拓扑结构域与 RBFOX2 基因的差异表达

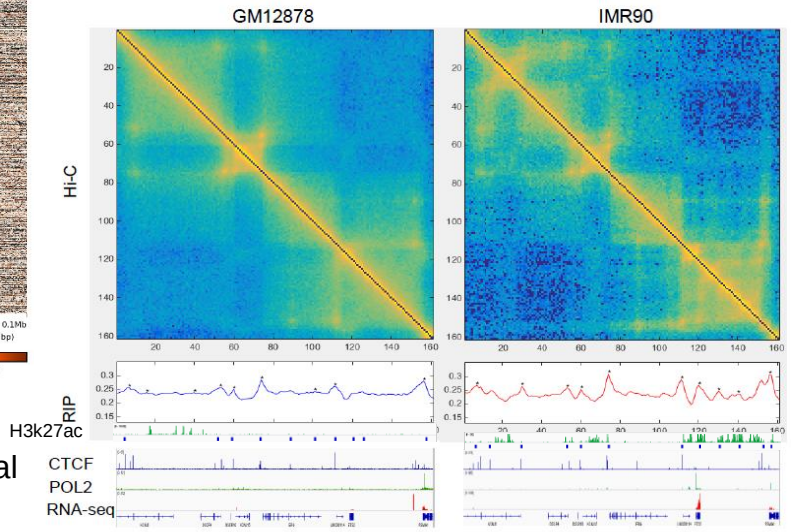
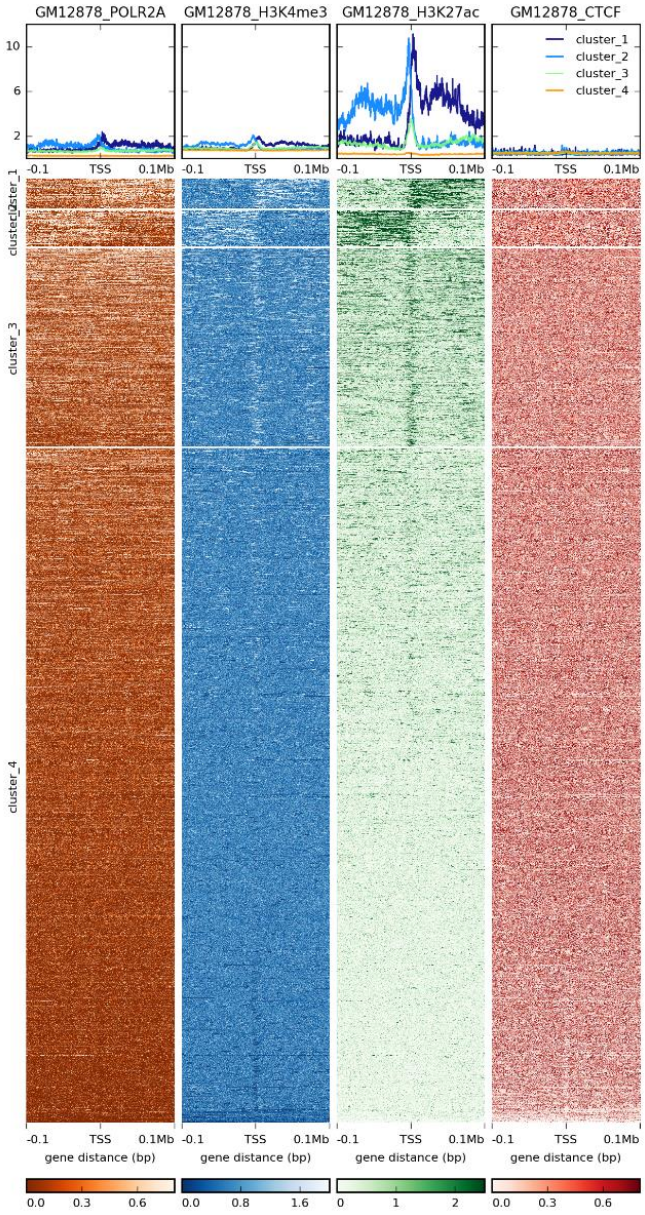


图 4.15　差异的拓扑结构域与 ETS2 基因的差异表达

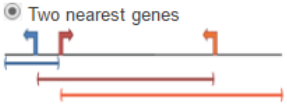
Figure5. HiCTAD facilitate differentical domain boundary detection



a subcompartment Transition or a loop

Figure5. HiCTAD facilitate differentical domain boundary detection

Great analysis



within 1000.0 kb

Gene regulatory domain definition: Each gene is assigned a regulatory domain that extends in both directions to the nearest gene's TSS but no more than the maximum extension in one direction.

GM12878

GO Biological Process

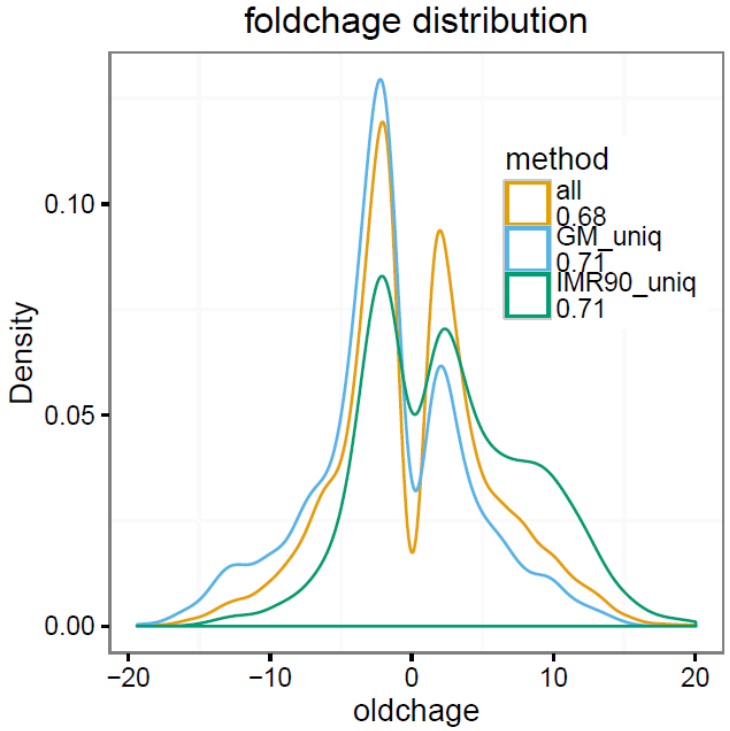
Term Name	Binom Rank	Binom Raw P-Value	*Binom FDR Q-Val*
interferon-gamma-mediated signaling pathway	64	1.8157e-6	2.9619e-4
negative regulation of B cell activation	93	1.0716e-5	1.2029e-3
nose development	237	5.2360e-4	2.3065e-2
mast cell activation	255	7.8711e-4	3.2225e-2
membrane raft organization	264	8.9193e-4	3.5272e-2

GO Cellular Component

phagocytic vesicle	17	8.5928e-7	6.3940e-5
phagocytic vesicle membrane	24	3.4675e-6	1.8276e-4

IMR90

Term Name	Binom Rank	Binom Raw P-Value	*Binom FDR Q-Val*
collagen fibril organization	83	2.2004e-6	2.7678e-4
positive regulation of transforming growth factor beta receptor signaling pathway	134	1.1194e-5	8.7211e-4
regulation of peroxisome proliferator activated receptor signaling pathway	164	2.9672e-5	1.8888e-3
regulation of transcription from RNA polymerase II promoter in response to oxidative stress	183	5.7852e-5	3.3004e-3
negative regulation of endothelial cell proliferation	186	5.8960e-5	3.3094e-3
regulation of fibroblast migration	233	1.5954e-4	7.1484e-3
mesenchymal-epithelial cell signaling	391	1.2599e-3	3.3639e-2
atrioventricular canal development	405	1.4524e-3	3.7439e-2
fibrillar collagen	41	4.0637e-4	1.2538e-2



Structural variation leads to huge expression variation

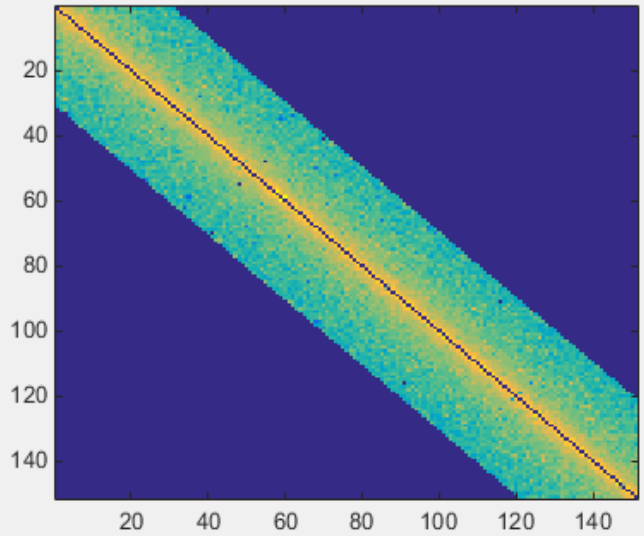
Improvement：1. a more elaborate method to define structural variation related gene
2.Noise removal

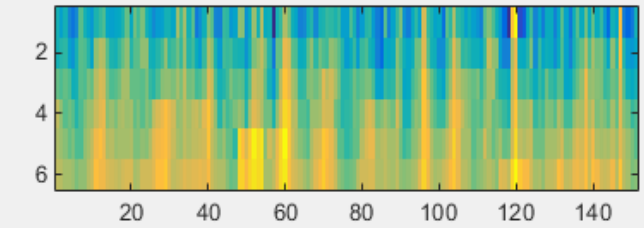
Future work

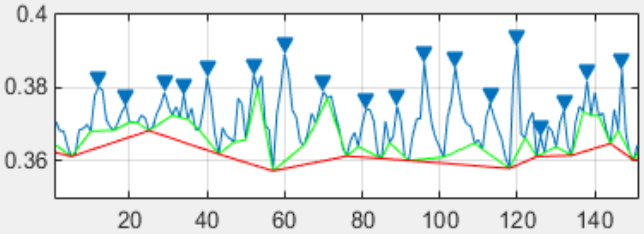
- 1. CTCF motif free cut off strategy
- 2. hierarchical domain detection(distinguish between TAD boundary and sub-TAD boundary)
- 3. application of HiCTAD method

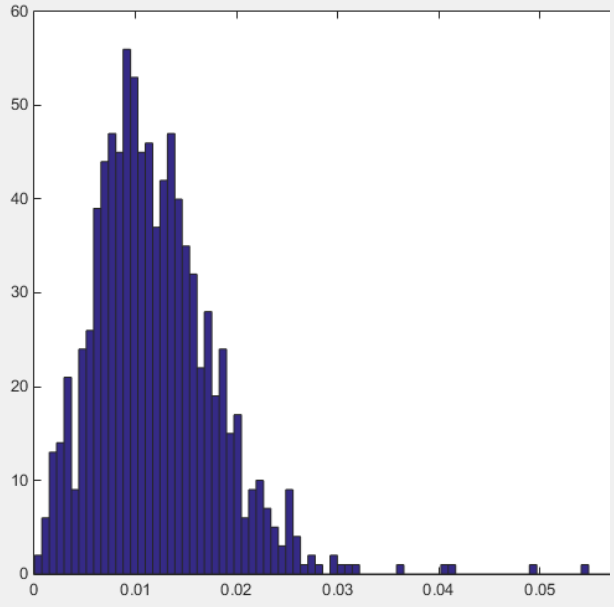
Future work1. CTCF motif free cut off strategy

Construct null hypothesis model for **peakscore**

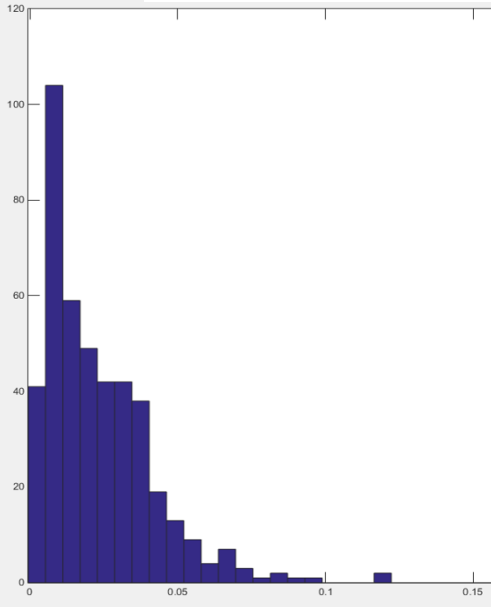






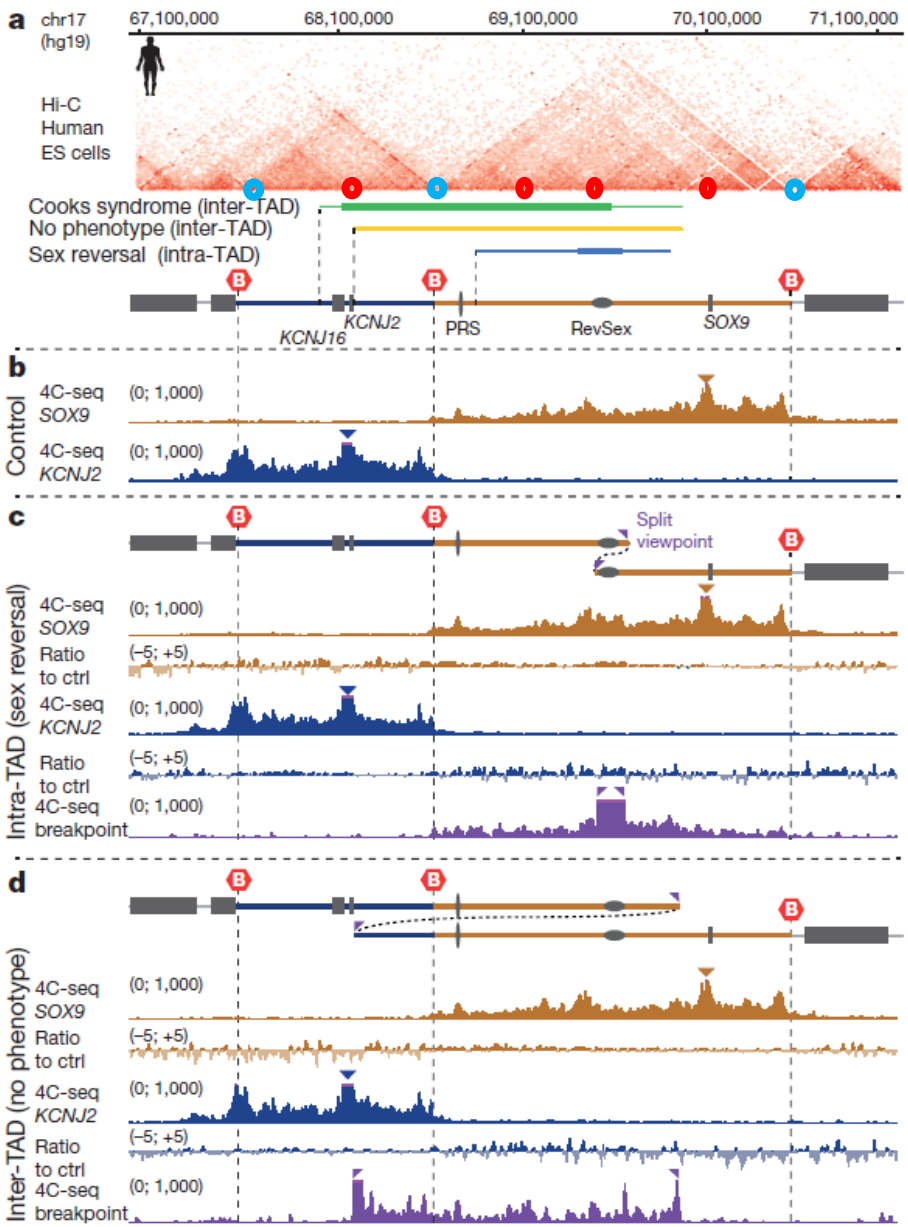


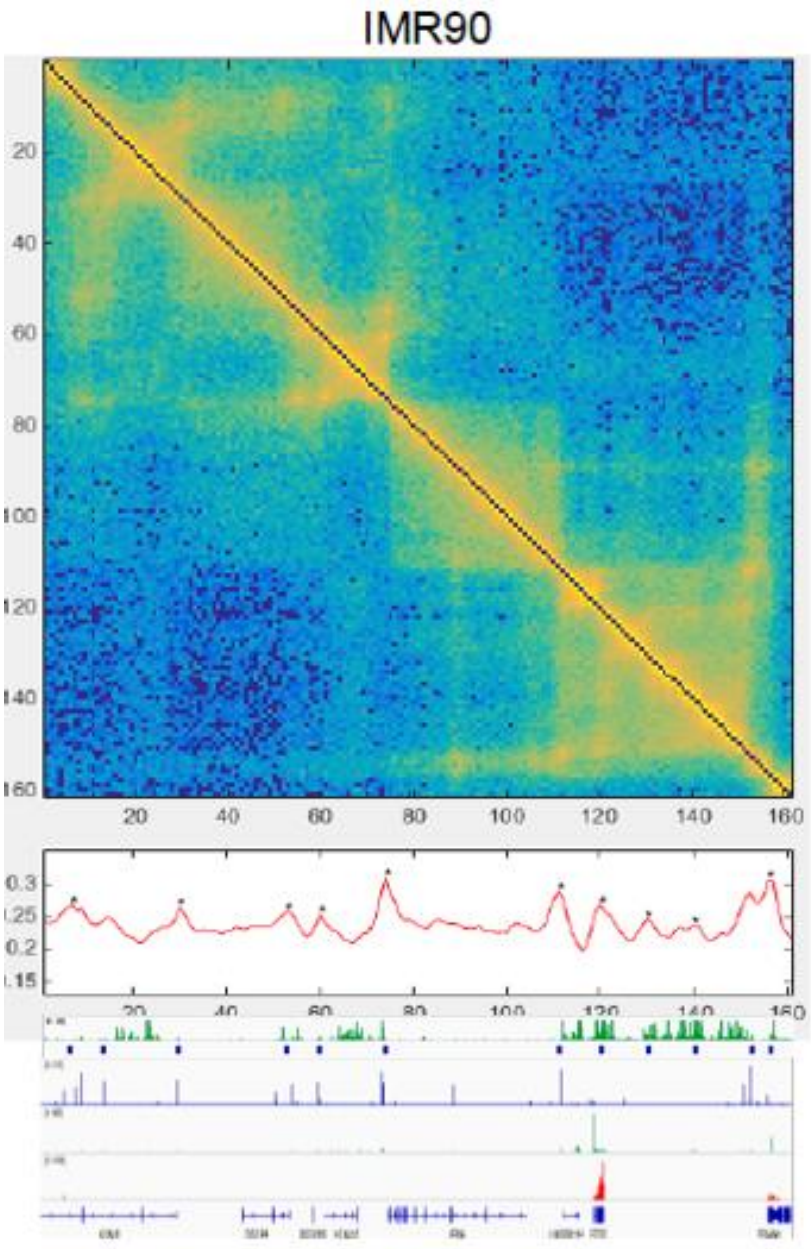
Random **peakscore** distribution



real **peakscore** distribution

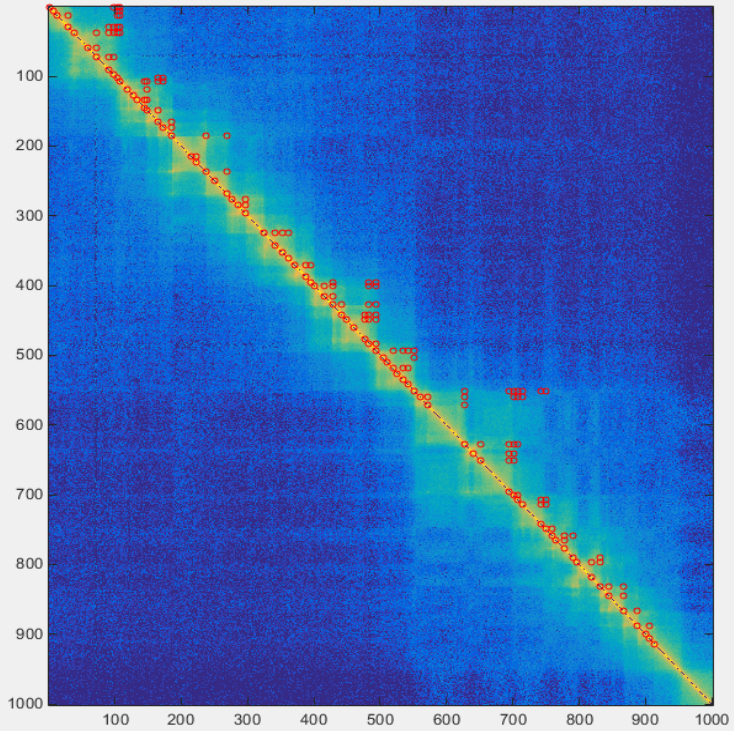
Future work2. hierarchical domain detection

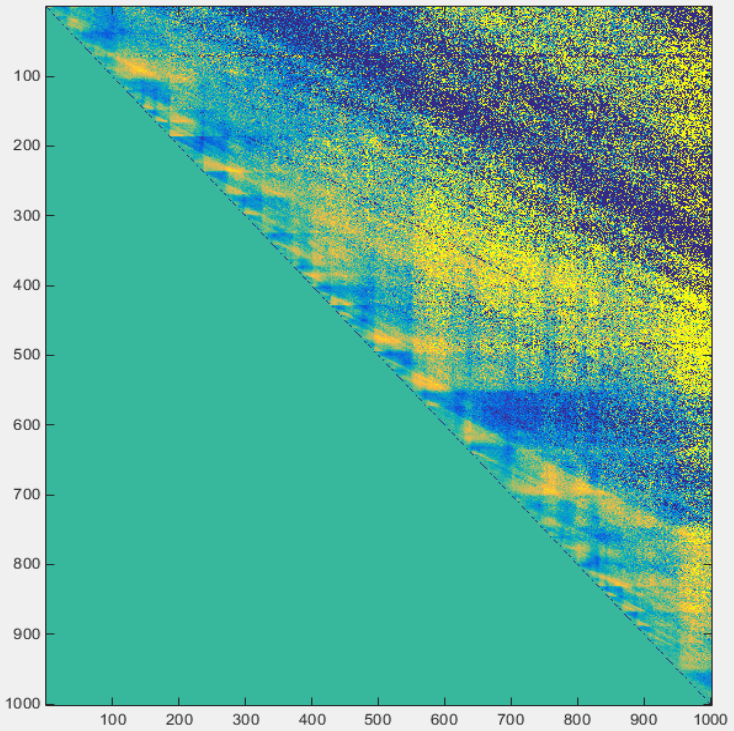




Future work2.hierarchical domain detection

帮助我们确定元件功能行使的区域

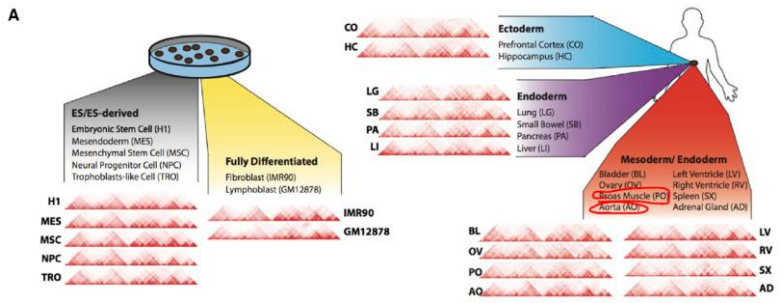




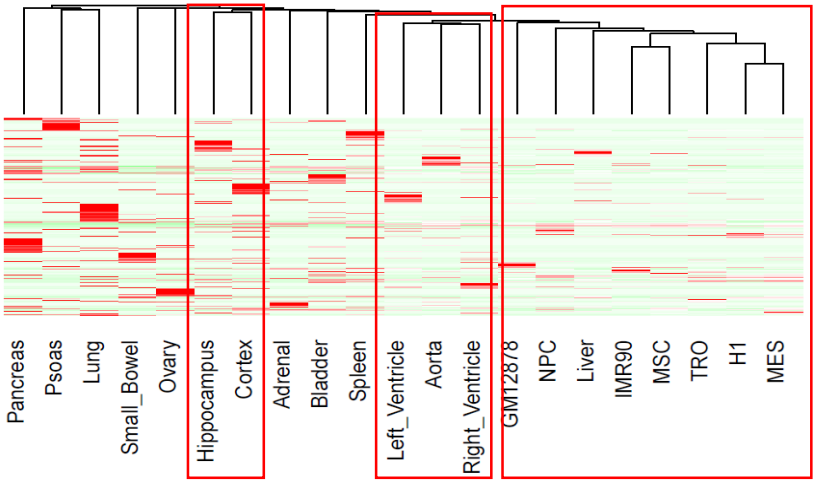
(i) almost all entries in $U_{a,b}$ are negative, and almost all entries of $L_{a,b}$ are positive.
(ii) when the sum of the entries in $U_{a,b}$ is subtracted from the sum of the values in $L_{a,b}$, the resulting value is larg
(relative to a random model)
(iii) the variance of the entries in $U_{a,b}$ and $L_{a,b}$ were both small (relative to a random model).

怎样基于boundary 自动的找domain
怎么去验证domain detection 的正确性

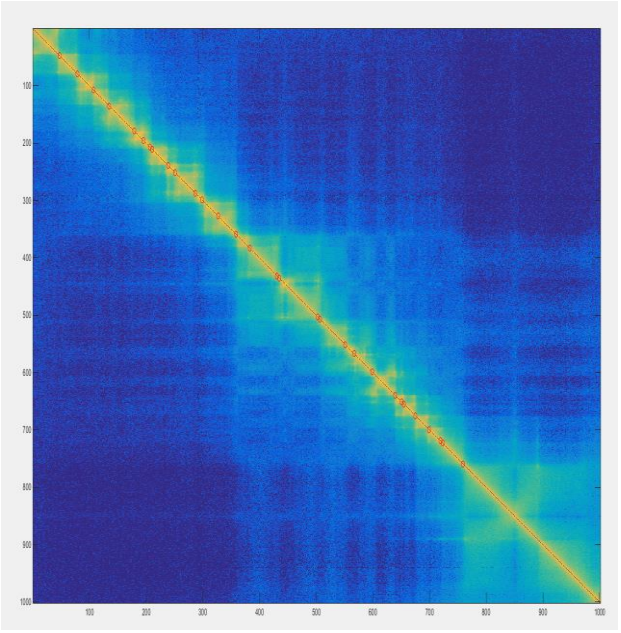
Future work3. conserved boundaries and cell-specific boundaries



3263 conserved boundary
34932 notconserved_TAD.txt
11887 uniq bound



Chr22:4000:5000（GM12878）



1.复制时间域
2.boundary CTCF 特征 (conserved boundary CCD)
or 对boundary 进行分类
3.特殊位置在不同细胞的差别：端粒

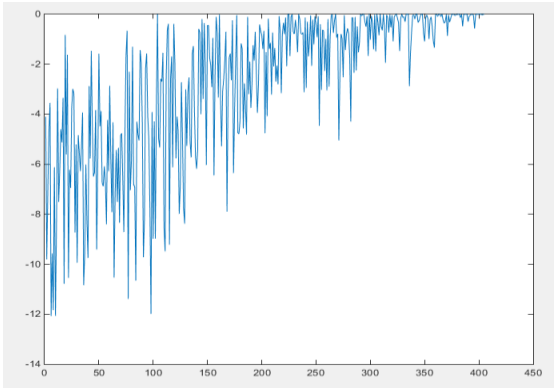
Thank You

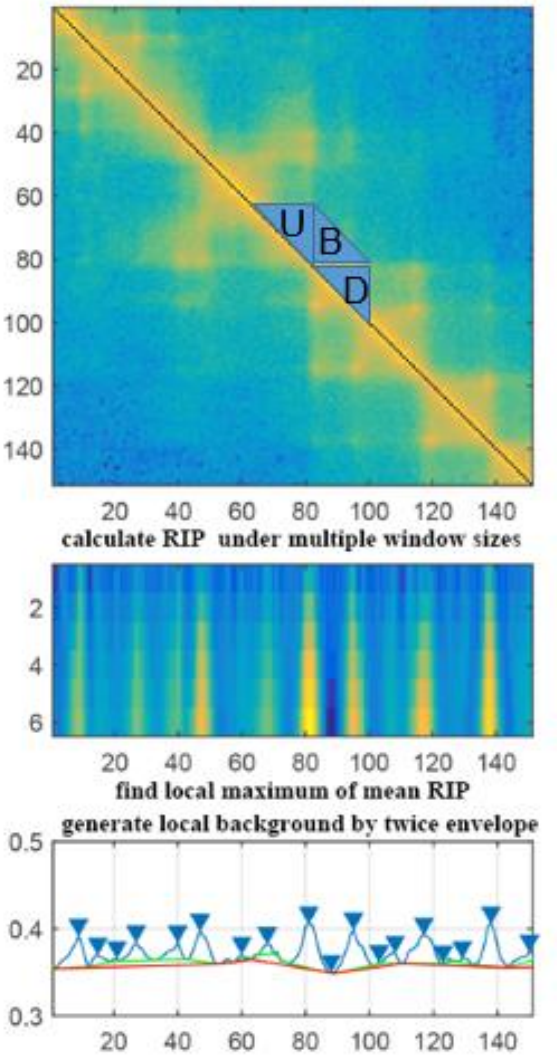
appendix

Problem1. CTCF-independent method

怎样将p-value和peak score合理结合在一起

- 1. TopDom test





- 2. Golobal background z-score