SUPPLEMENTARY MATERIALS

The process of using dynamic programming

The dimension of Hi-C interaction is N and the number of domains is K. For region R, the estimate $\widehat{m}_R$ of mean m_R is calculated as the empirical mean of Hi-C interaction frequency in the right upper corner of the matrix:

$$\{(i,j): 1 \le i \le N/4, (3N/4+1) \le j \le N\}$$

We can decompose R into $R_1, \ldots, R_k$,

$$R_K = \{(i,j): b1_k \le i \le b2_k, 1 \le j \le b1_{k-2} - 1\}$$

$$S_R = \sum_{k=3}^{N} S_{R_k} = \sum_{k=3}^{N} \left(m_{R_k} - \hat{m}_R\right)^2$$

We want to identify positions of D_k to optimize the following objective:

$$O(N,K) = \min \sum_{k=1}^{2} S_{D_k} + S_{A_2} + P_2 + \sum_{k=3}^{K} \left(S_{D_k} + S_{A_k} + S_{R_k} - P_k\right)$$

When K > 2:

We define the gain function E_k:

$$E_k = S_{D_k} + S_{A_k} + S_{R_k} - P_k$$

Then we want to minimize this objective which can be optimized by dynamic programming:

$$O(b2_k, k) = \min O(b2_{k-1}, k-1) + E_k$$

When K = 1:

$$O(N,1) = S_{D_1}$$

When K = 2:

$$O(N,2) = \min O(b2_1, 1) + E_{k2}$$

The definitions of bin, unorganized region, boundary, border and the same domain (Figure **S1**)**:**

A chromosome was divided into ***bins*** of equal size and Hi-C interaction matrix represents the interactions between every two bins. For Hi-C domains, we considered the region between two adjacent domains larger than 10 bins as ***unorganized region***, otherwise as ***boundary*** (Figure **S1**). When analyzing the enrichment of CTCF and histone modification marks around boundaries, we calculated the average number of peaks within +/- 20 KB of the boundary. This definition is same as defined by Dixon et al.

For all kinds of domains, ***borders*** of a domain are defined as the two end points of a domain. If the size of region between two domains is 0, we consider these two domains share one border. The borders we defined here are used to illustrate the change-points of the domains which mean the changes of chromatin structure (Figure **S1**). If each border of a domain was correspondingly within one bin distance to another domain, these two domains were considered as ***the same domain*** when two different domain sets were compared.

Aligning Hi-C domain borders to 5C or arrowhead domain borders

For borders of 5C or arrowhead domains, we first round their chromosome coordinate to the multiple of the size of a bin. Then we aligned the Hi-C domain borders to these new chromosome coordinates to get the overlap.

Data source

Hi-C data from two different cell lines (human IMR90 cell line and mouse ES cells), and 5C data from mouse ES cells were obtained from literature [7, 8, 22, 23] (please see Table S1 for data source), respectively.

Data preprocessing

For ChIP-seq data, peaks were called using MACS with default parameters [37]. UCSC Liftover tool [38] was used to convert human genome version hg19 to hg18. The reads were mapped using BWA [42] with default parameters and the reads multiple mapped or mapping quality less than 10 were removed. We removed the contact pairs of genomic distance less than 20 kilo-bases that were considered as self-ligation. Every chromosome was divided into bins with size of 40 kilo-bases. We normalized the Hi-C data as previously described by Yaffe and Tanay [43].

In human IMR90 and mouse ES, for the domain calling using HiCseg (using model of non-block diagonal with Gaussian), DI (default parameters) and CHDF method, the minimum domain size is 3 bins (40 kb per bin). For each method, domains with over 90% bins which did not contain Hi-C interaction data were filtered out.

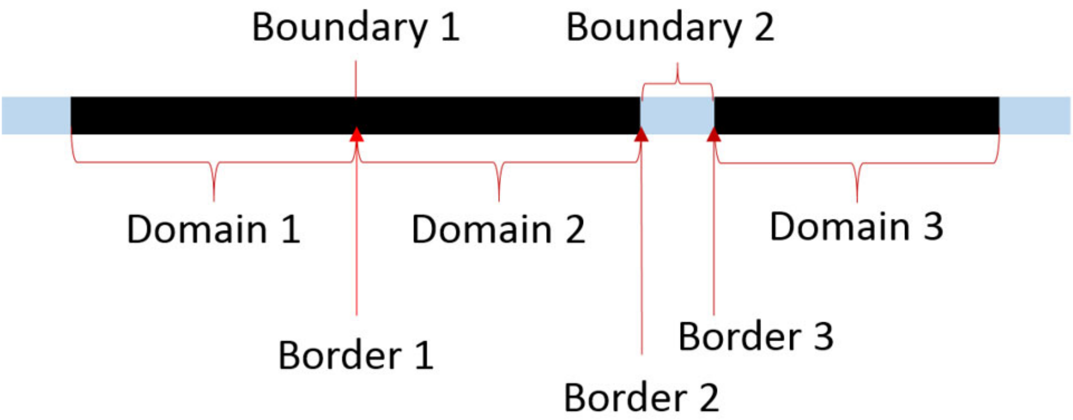

Figure S1. The schematic graph for the definition of domain, boundary and border.

Chromatin (non-domain, shallow blue bar) and domains (black bars) are shown here, respectively.

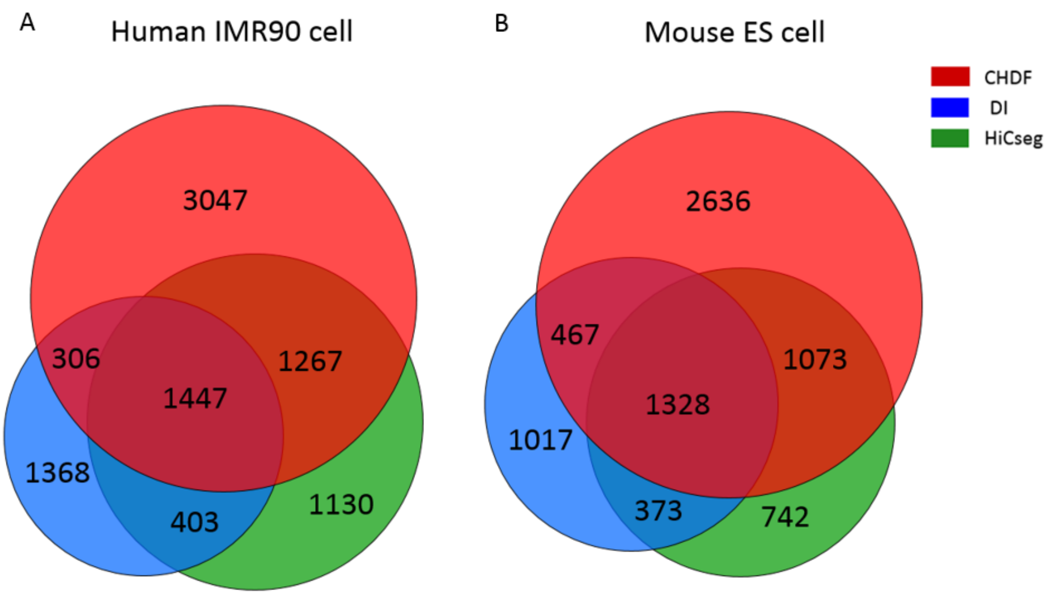

Figure S2. Domain borders overlapped among the three methods

(A) 6067 domain borders in human IMR90 cell line, (B) 5504 domain borders in mouse ES cells were identified by CHDF (red), which included many identified by either DI (blue) or HiCseg (green).

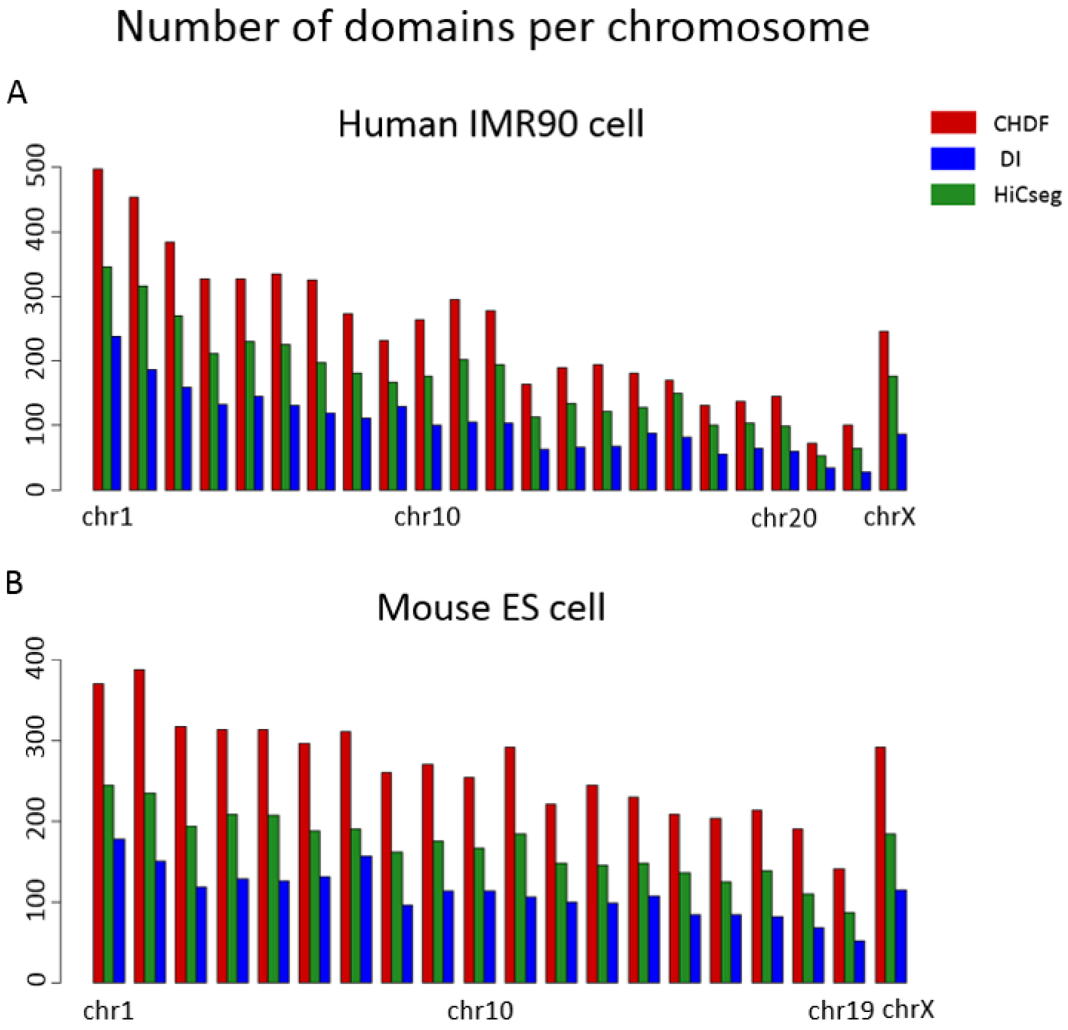

Figure S3. Number distribution of domains in every chromosome identified by three methods.

(A), human IMR90 cell line; (B), mouse ES cells.

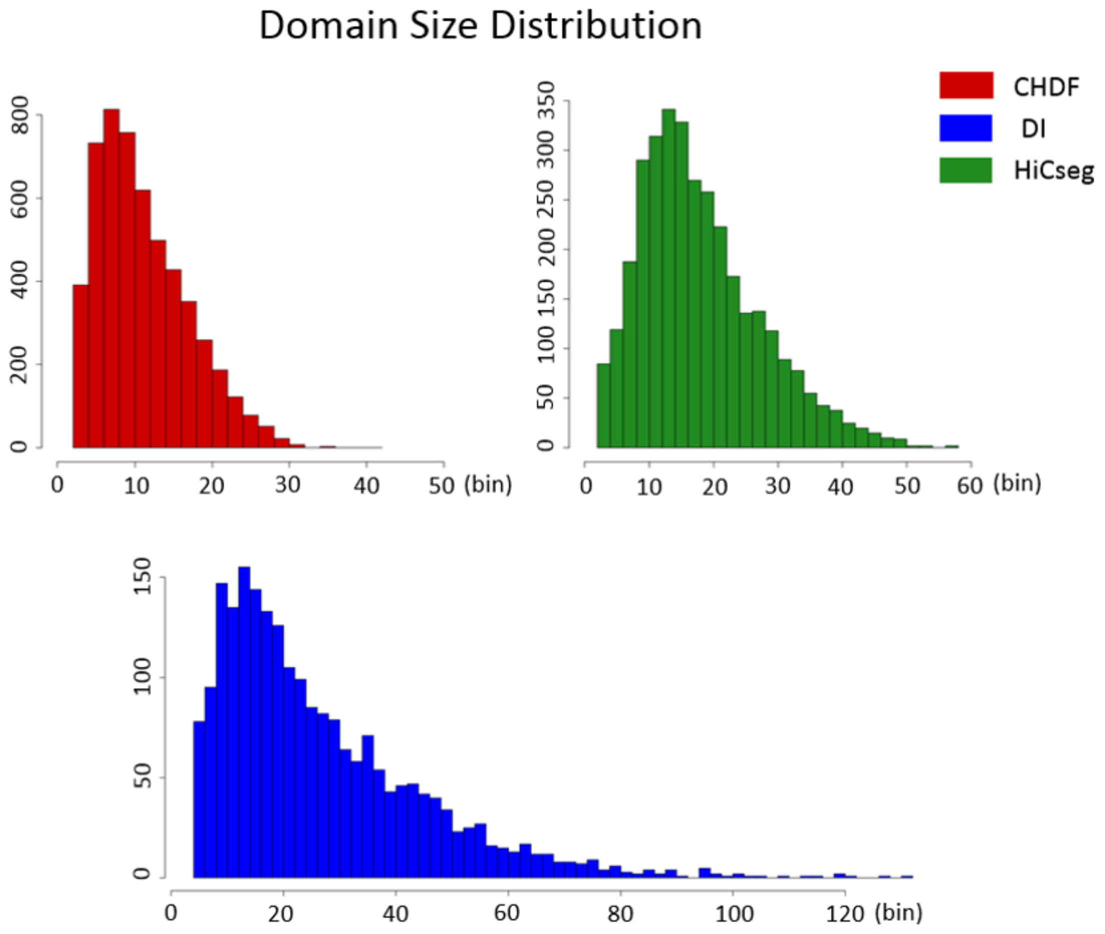

Figure S4. Size distribution of domains in mouse ES cells identified by three methods

(A), CHDF (red); (B), HiCseg (green); (C), DI (blue)).

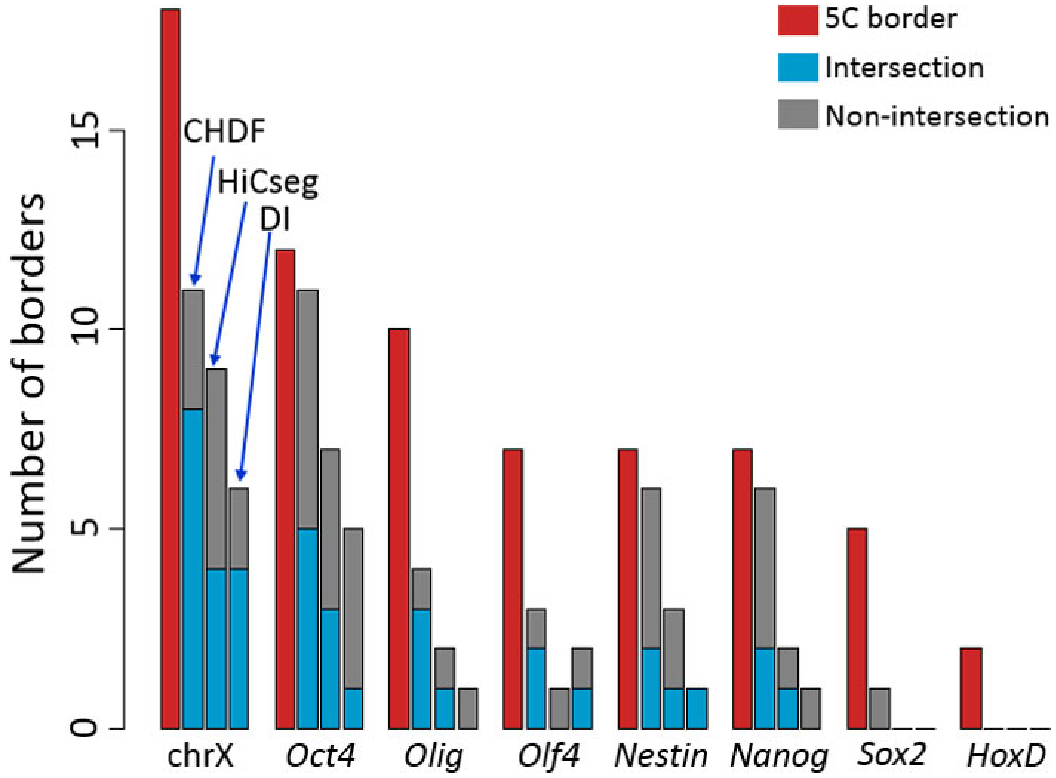

Figure S5. Statistics of Hi-C domain borders overlapped with 5C domain borders in eight genomic regions. For each region, first bar represented number of 5C borders and the followed three bars respectively represented number of domain borders called by three methods.

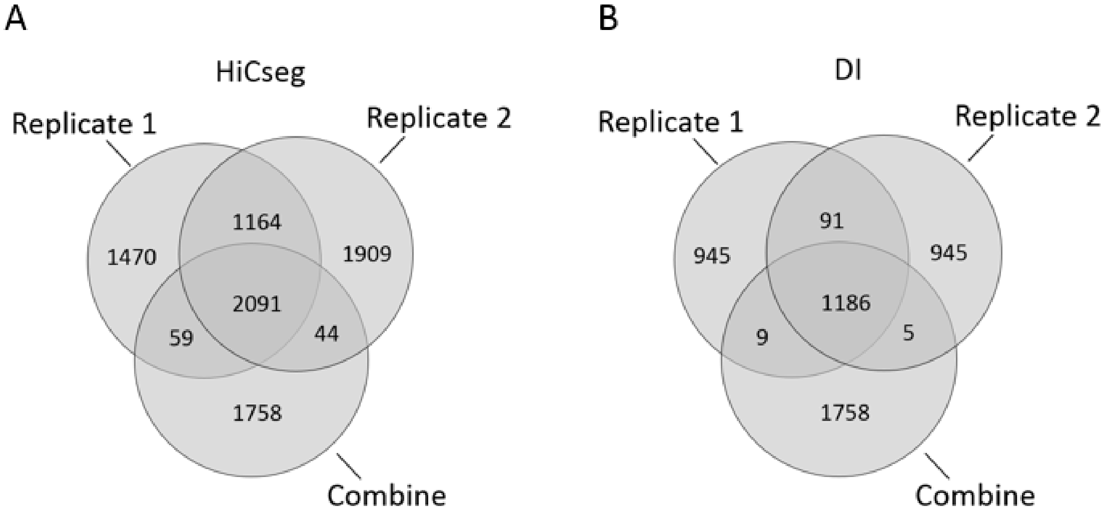

Figure S6. Number of domains called by HiCseg (A) and DI (B) in two replicate and combined datasets. The intersection represented the same domains called in different Hi-C datasets.

Table S1. Data used in this study (some datasets were not deposited in the Gene Expression Omnibus (GEO)).

Dataset	**Accession**
IMR90 CTCF [39]	
mESC CTCF [40]	GSE36027
mESC H3K4me3, H3K4me1, H3k27ac, H3k9ac [40]	GSE31039
IMR90 H3K4me3, H3K4me1, H3k27ac, H3k9ac [41]	GSE16256
Hi-C, J1 mouse ES cells and IMR90 fibroblasts [8]	GSE35156
5C [22]	GSE36203
IMR90 arrowhead domain [13]	GSE63525
Lymphoblast cis-eQTL [26,27,28,32,33]	GSE17080 , GSE19480, GSE6536
Fibroblasts and T-cell cis-eQTL [27]	GSE17080
Monocyte cis-eQTL [29]	
Cortex cis-eQTL [30]	GSE8919
Liver cis-eQTL [31,34]	GSE9588, GSE6536