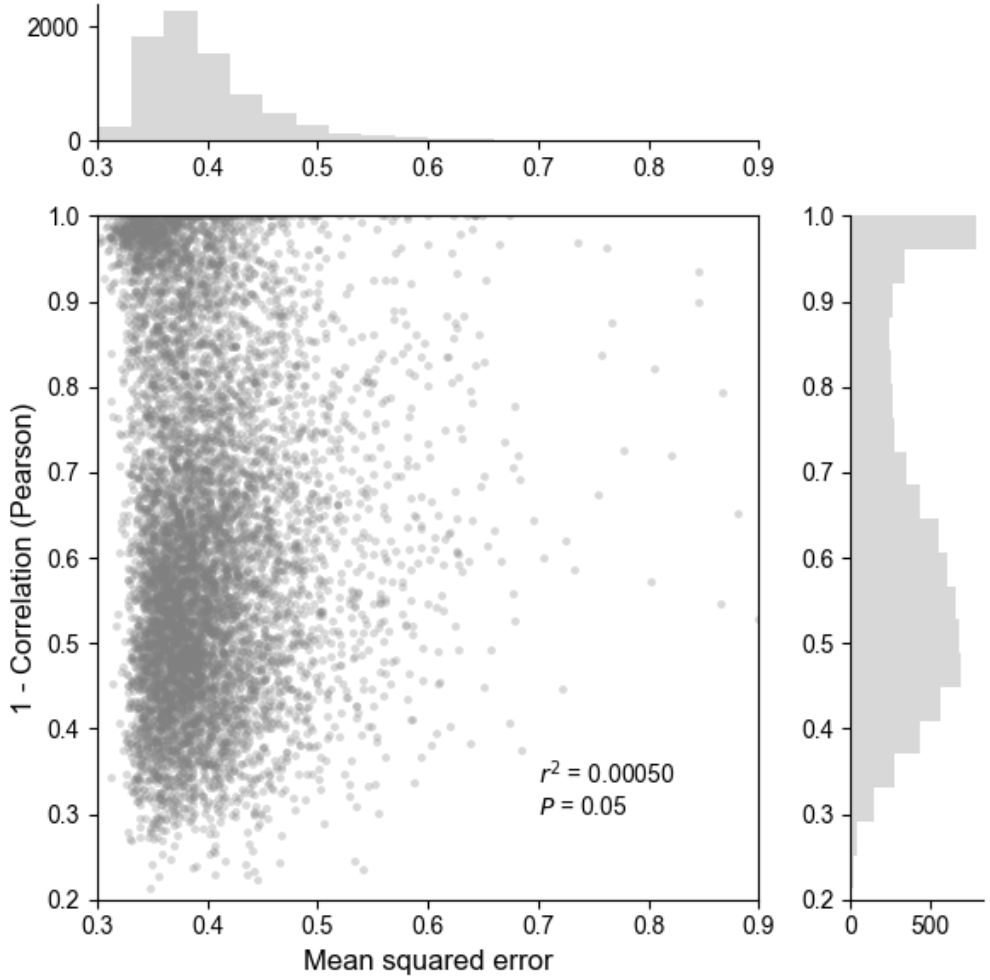
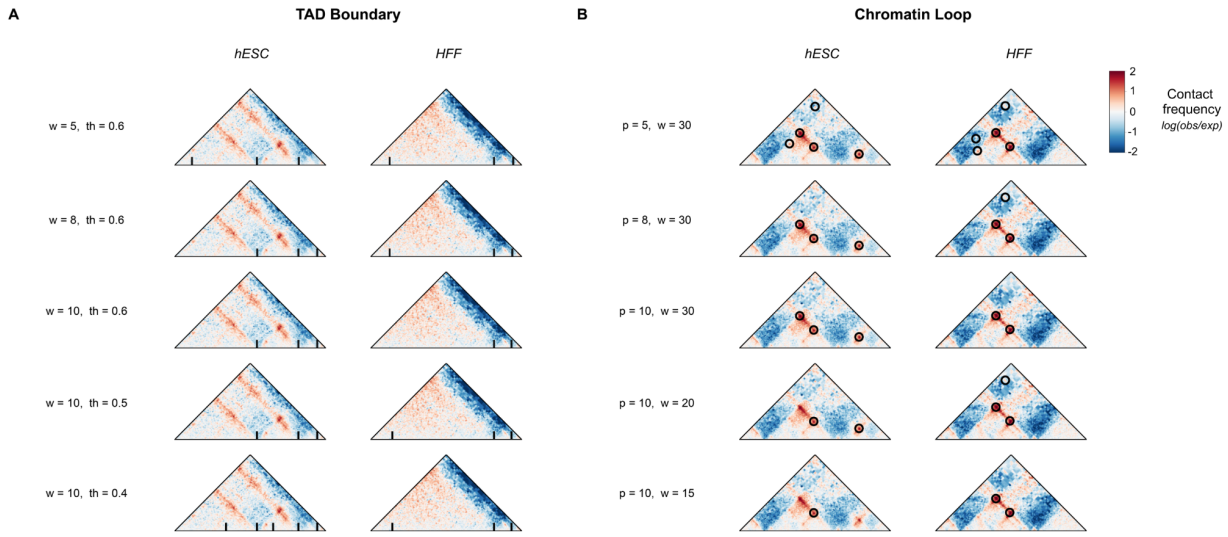
Supplementary Information

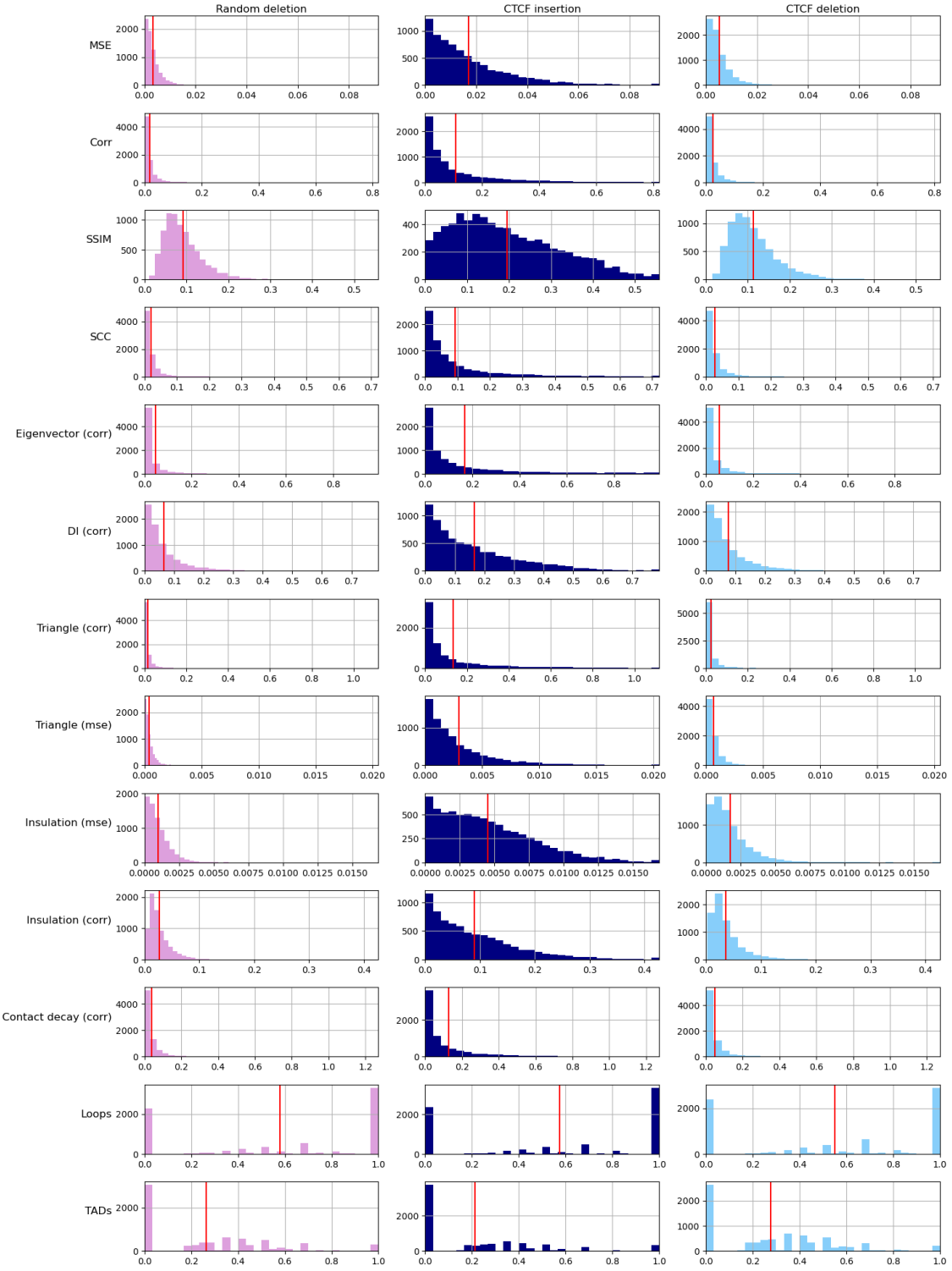
Supplemental Figures



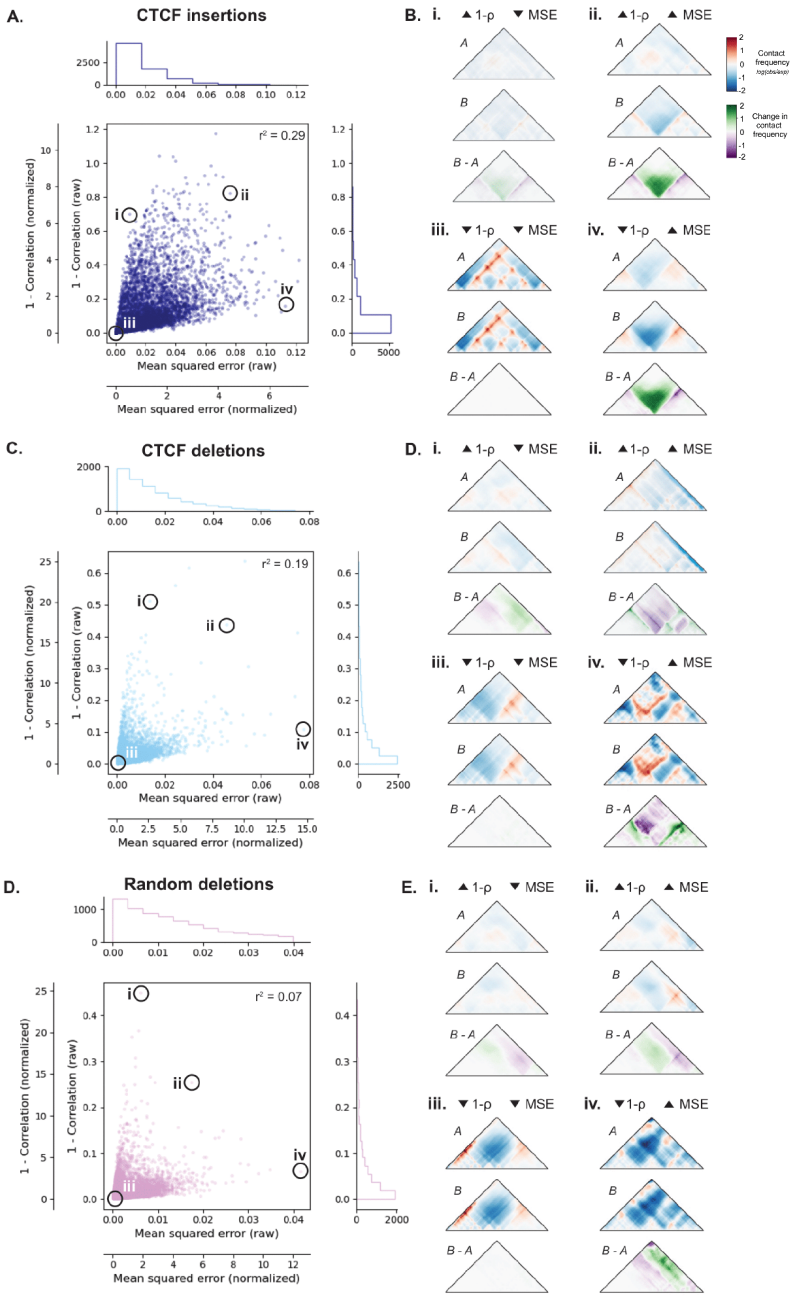
Supplemental Figure 1. Pearson correlation versus mean squared error comparisons of contact maps. Mean squared error (MSE) and Pearson correlation coefficient calculated across the genome on experimental contact maps from embryonic stem cell (ESC) and human foreskin fibroblast (HFF). Each point represents a comparison between maps from HFF and ESC cell types (n = 7840 windows). There is a weak relationship between the Pearson correlation and MSE (r^2 = 0.0005, P = 0.05)



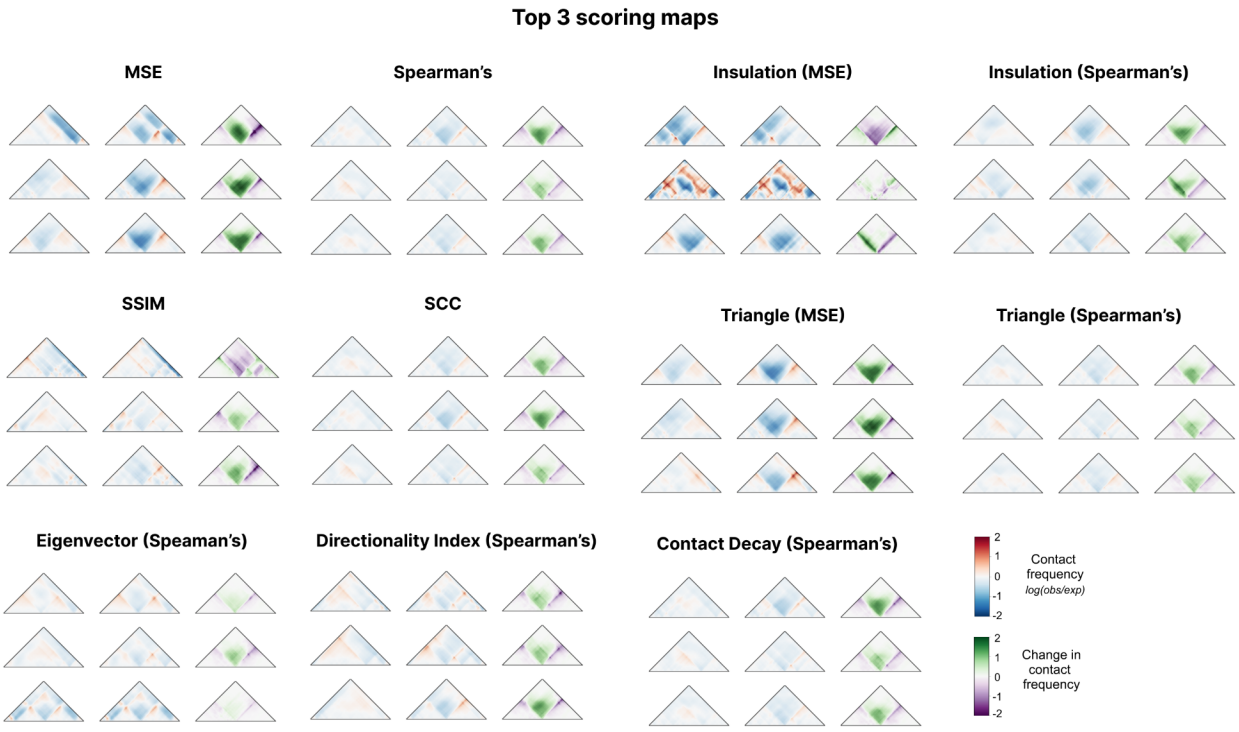
Supplemental Figure 2. Sensitivity of TAD and loop caller on parameter shifts. (**A**) TAD boundaries (highlighted with black bar) called with different sizes of diamond-shaped window (*w*) and thresholds of insulation scores (*th*). (**B**) Chromatin loops (highlighted with black circle) identified using different sizes of center window (*p*) and donut filter (*w*). Example maps used here are the same as in **Fig. 3Bi**.



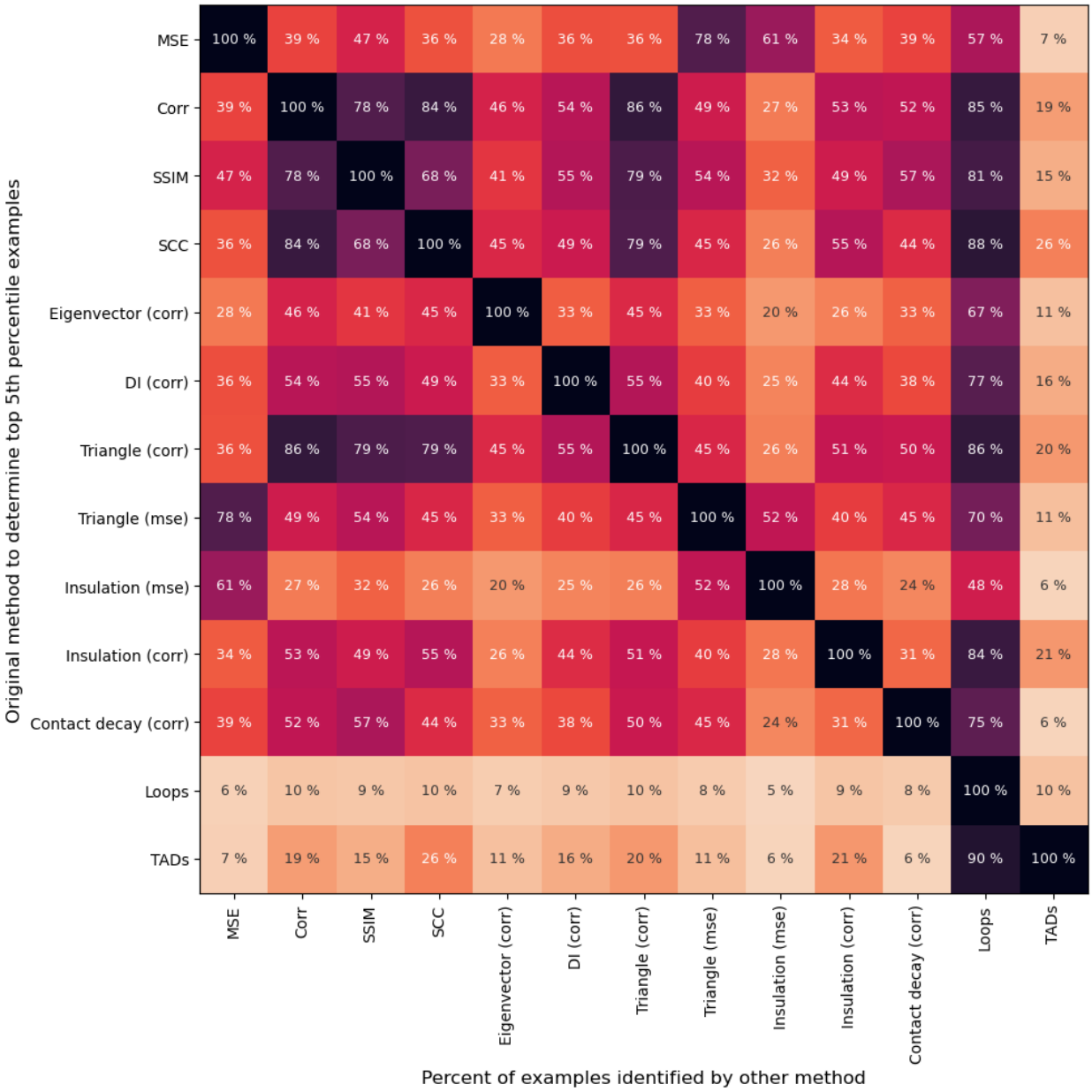
Supplemental Figure 3. Score distributions of random deletions, CTCF deletions, and CTCF insertions. Each disruption score method (rows) produces a different range and mean (red line) across scores produced. Histograms show the raw scores comparing maps produced by 7500 random 100 bp deletions (left), 7500 CTCF insertions (middle), and 7500 CTCF deletions (right). To enable comparisons between the different scores, the main text figures report scores standardized to the mean disruption produced by a random 100 bp deletion. Thus, both an MSE-based disruption score and correlation-based disruption score describe that the maps are twice as different as the average 100 bp deletion.



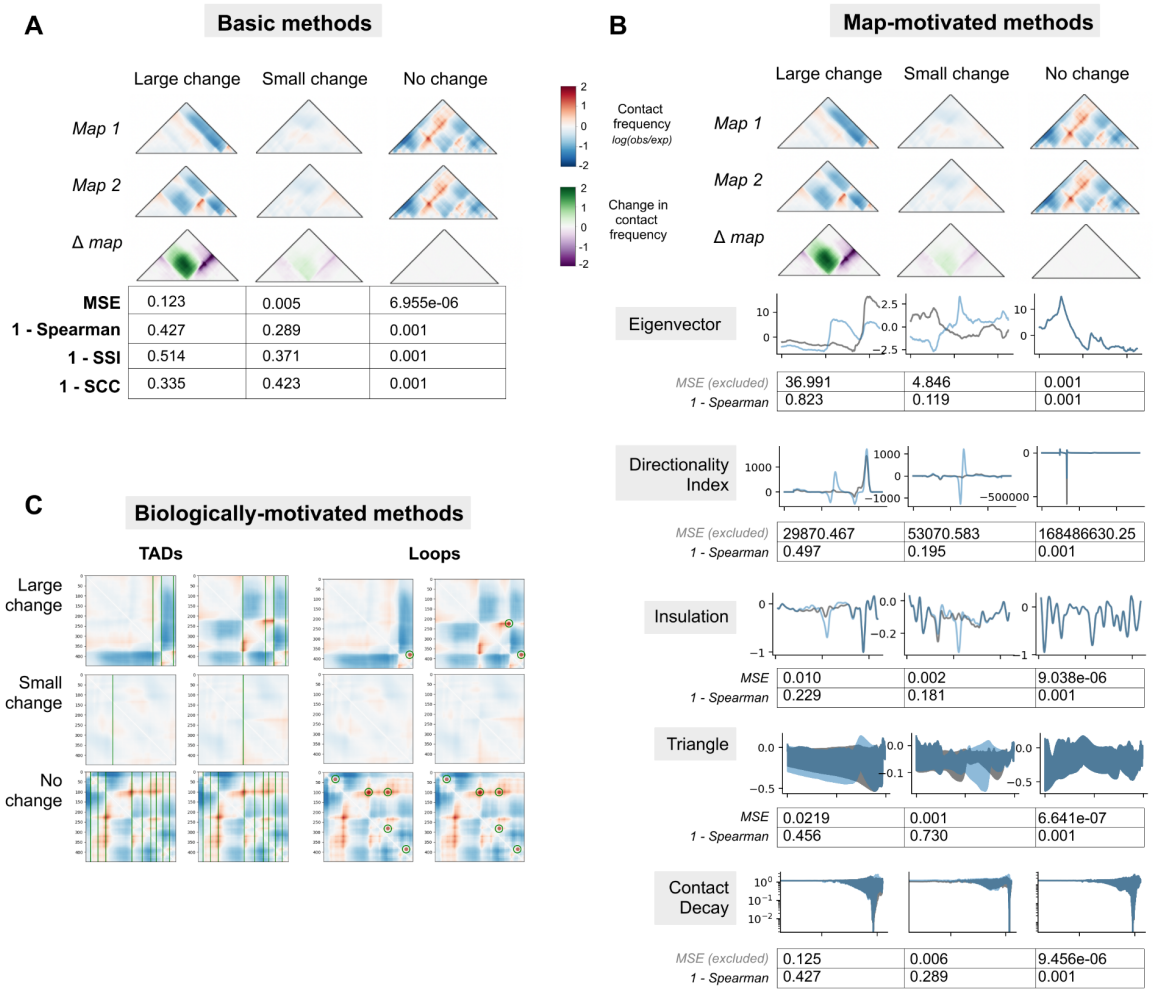
Supplemental Figure 4. Basic methods to compare contact frequency maps rank maps differently on *in silico* perturbations. Mean squared error (MSE) vs Spearman's correlation (ρ) scores across an *in silico* screen of 7,500 map pairs with and without CTCF insertions (A-B), CTCF deletions (C-D) and random 100 bp deletions (E-F), similar to **Fig. 2.** We plot 1 - ρ such that higher values for both methods reflect increasing differences between maps. MSE versus Spearman's correlation are plotted where each point represents a comparison between a reference and perturbed map (**Fig. 4a**). Normalized scores are divided by the mean of the distribution of random deletions. Across all three perturbations, there is a weak relationship between the two disruption scores (r^2 = 0.29, 0.19, and 0.07 for CTCF insertions, CTCF deletions and random deletions, respectively). The relationship is strongest for CTCF insertions, for which scores are highest, followed by CTCF deletions, which have the next highest scores. Yet perturbations with at least one high score are not always concordantly scored. Examples of extreme scores for each perturbation are shown in B, D, and F in panels i through iv, illustrating that perturbations with high MSE and low 1 - ρ are consistently maps with high contrast, while low MSE and high 1 - ρ perturbations are maps with overall low contrast.



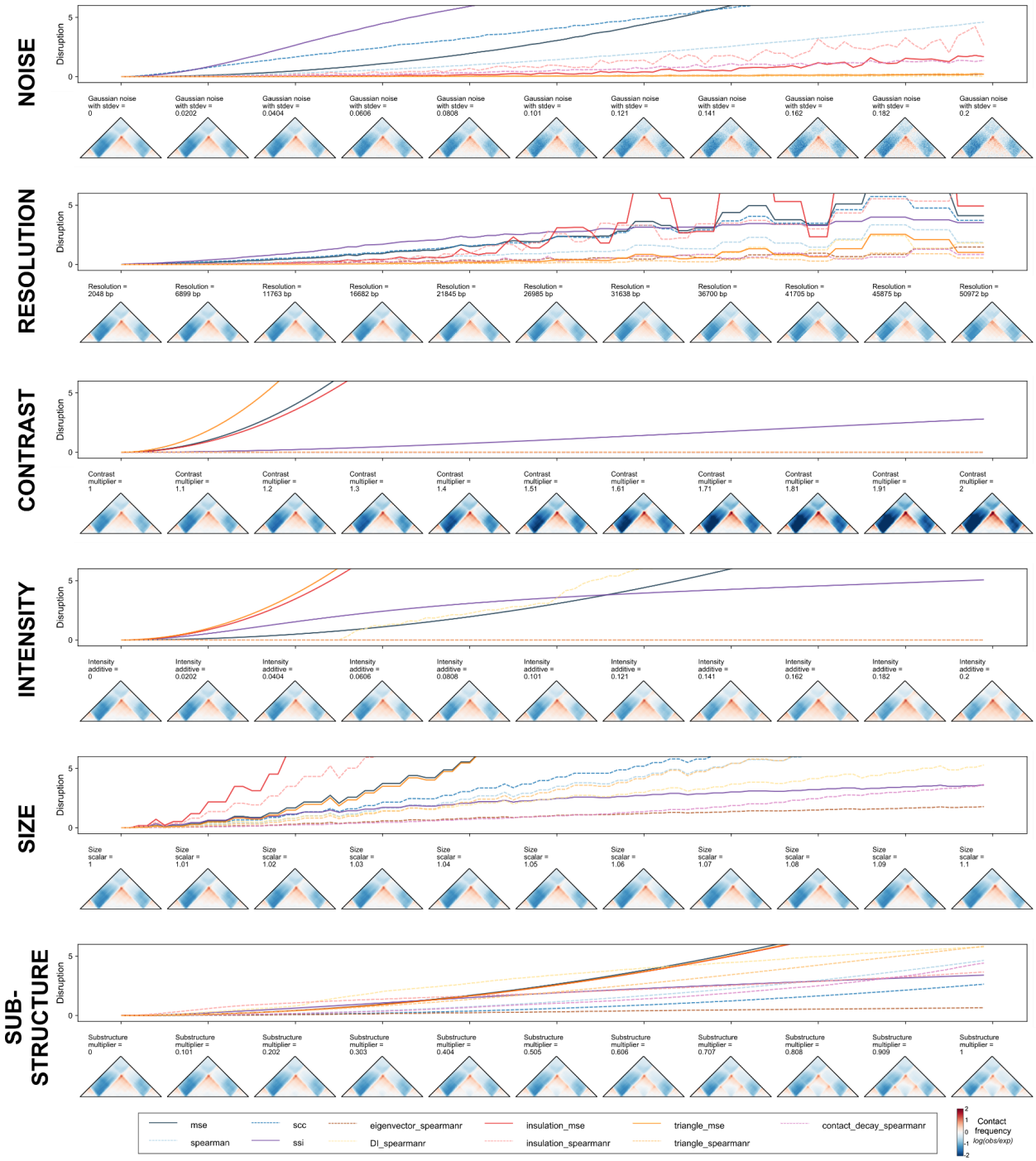
Supplemental Figure 5. The three most disruptive map pairs of each scoring method. For each example row, the unperturbed map is shown on the left, the perturbed map is shown in the middle, and the difference between the two maps is shown on the right. The top three disruptive maps were chosen across the *in silico* screen.



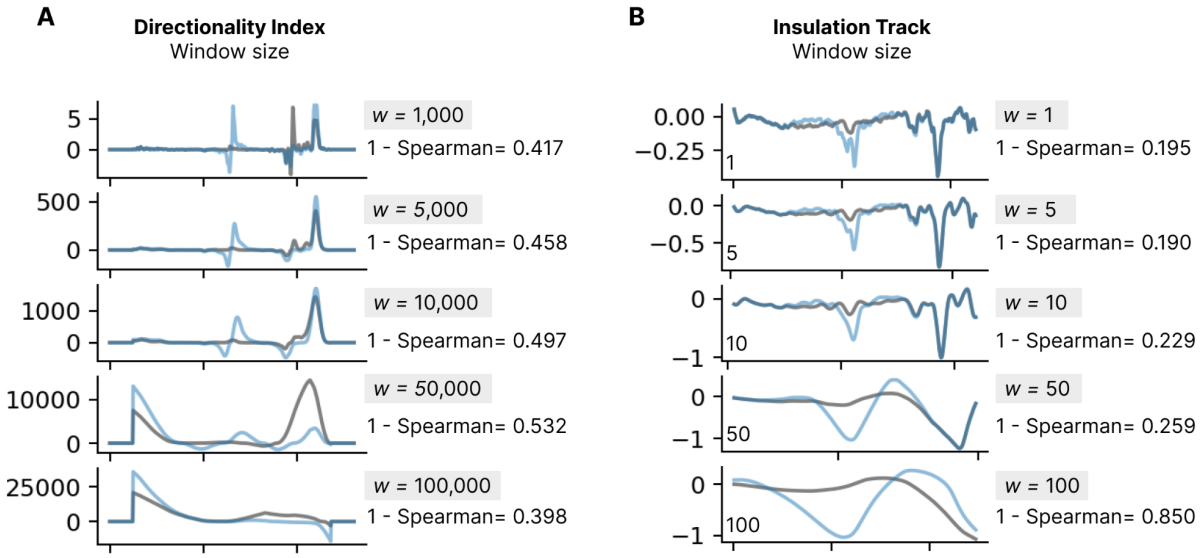
Supplemental Figure 6. Overlap of the most disruptive map pairs identified by each scoring method. Each cell in the heatmap represents the percentage of map pairs that are above the 5% cutoff for the method in row and above the 5% cutoff for the method in the column. Darker colors indicate higher concordance for the top scoring loci. The heatmap is symmetric except for Loops and TADs. The imbalance of these two methods is caused by multiple map pairs that have scores equal to the 5th percentile, which results from methods producing low counts of discrete values.



Supplemental Figure 7. Scoring metrics on contact map pairs with large, small, and minimal changes. (**A**) Basic method scoring results across three example loci with a large, small, and minimal change upon CTCF motif insertion. (**B**) Map-motivated scoring results across three example loci. Raw tracks are shown for each measurement and the MSE and Spearman's correlation between the tracks are shown below. (**C**) Feature-informed scoring examples across three example loci with a no change, a minimal change, and a large change to folding.



Supplemental Figure 8. Changes of disruption scores with gradual increases in perturbations. Each subpanel shows the changes of disruption scores (top row) and contact maps (bottom row) against the incremental changes in a technical or biological variation. The colors of the scoring metric are the same as seen in **Fig. 5**.



Supplemental Figure 9. Sensitivity of directionality index and insulation tracks on parameter shifts. Directionality index (**A**) and insulation (**B**) tracks across a range of input window size choices, as well as the resulting Spearman's correlation between the two tracks. A window size of 10 Mb was used for both approaches to produce the *in silico* scoring results in the **Results** section.

Supplemental Tables

Supplementary Table 1: Method summary table

Method (abbreviation)	Description	Parameters	Time per 1k calls (s)
Mean squared error (MSE)	Measures the average squared difference between two flattened contact matrices.	NA	0.24
Spearman's correlation coefficient (corr)	Assesses the correlation between the intensity of corresponding pixels in two maps.	NA	5.31
Stratum-adjusted correlation coefficient (SCC)	Addresses bias due distance-dependence patterns in Hi-C data by stratifying based on their genomic distance, calculating a Pearson correlation coefficient for each stratum and aggregating the weighted stratum-specific correlation coefficients with weights derived from the generalized Cochran–Mantel–Haenszel statistic.	NA	2.96
Structural similarity index measure (SSIM)	Quantifies the perceived change in structural information of two images by incorporating three terms that correspond to structure, luminescence, and contrast	NA	2.37
Eigenvector difference (eigenvector)	Calculated from the first eigenvector that corresponds to each contact frequency map. Eigenvector tracks are compared between maps using spearman's rank correlation.	NA	41.51
Directionality index difference (DI)	A measure of contact frequency bias towards either upstream sequence or downstream sequence at a locus. An inflection of DI values from negative to positive and vice versa indicates a potential chromatin boundary.	(1) size of the focal bin where upstream and downstream contact frequency is compared (2). size of upstream and downstream bins	3.98
Insulation difference (insulation)	Identifies TAD boundary-like regions by comparing the frequency of within-region contacts upstream and downstream of some point to the inter-region contacts between regions A and B	(1). size of central window (region where insulation score is calculated, x_size) (2). size of upstream/downstream windows used (a_size)	10.98
Contact decay difference (contact decay)	Measures chromatin interaction as a function of genomic distance. Interaction frequency across the contact map is ranked by genomic distance between all pairs of contact, resulting in a track of distance vs interaction frequency. As distance increases, the probability of contact between loci decreases.	(1). mean vs median	7.17
Triangle profile (triangle)	Tries to leverage our understanding that contacts are represented by different subsets of triangles within the larger map. Compares the average contact intensity within all sub-triangles of two contact maps.	NA	4284.15
TAD calling (TADs)	TAD boundaries are called by finding the local minima of the insulation profile, which is calculated using a diamond-shaped window-based method. Specifically, a square is slid along each diagonal bin of the matrix and the averaged contact frequency within each window is calculated and called as insulation score.	(1) window size (w) (2). threshold of boundary strength (3). upper bound of distance when two TAD boundaries are considered one	8.16
Loop calling (Loops)	Loops are identified by comparing with their local background. For each bin in the upper triangle window of the matrix, we first check whether it is a local maximum and then calculate the mean signals of center window surrounding the bin and also the mean signals in a donut-shape neighborhood, a lower-left neighborhood, vertical and horizontal neighborhoods around the pixel.	(1). center window size (p) (2). window size (w) (3). threshold of the ratio of center window to donut and lower-left filter (4). threshold of the ratio of center window to vertical filter (5). threshold of the ratio of center window to horizontal filter (6). upper bound of bin distance where two loops are considered as same one.	104.18

Supplemental Text

Basic methods

Mean Squared Error

The mean squared error (MSE) measures the average squared difference between two flattened contact matrices, such that

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Because MSE is a measure of absolute difference, it consistently prioritizes the greatest changes in intensity between contact maps. MSE has been widely adopted across machine learning as a loss function for consistent performance and ease of use [1–4] . Large changes between maps score highly, while visually smaller or localized changes produce lower MSE values. However, maps with differences in read count or normalization intensity will produce high MSE, despite little change in structure. For this reason, technical artifacts may dominate top map rankings scored by MSE. MSE will also deprioritize maps with large structural changes and low overall contact intensity. 2D map features will not be individually captured since the matrices are collapsed to 1D vectors.

Spearman's Rank Correlation Coefficient

The Spearman's rank correlation coefficient (ρ) assesses the correlation between the intensity of corresponding pixels in two maps by quantifying how well the relationship between the corresponding pixels can be described using a monotonic function. If the rank of intensity of all pixels in two contact maps are the same, the correlation is 1. If there is no relationship between the rank of pixel intensity between maps, the correlation is 0. Spearman's Rank Correlation coefficient can be described as:

$$\rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)},$$

where the number of points in the data set is represented by n, and d^2 is the squared difference in the ranks of a single coordinate y_i between the two maps, which is summed over all points.

Correlation coefficients have been used extensively to compare contact maps[1,4–6]. Large-scale structural changes have high scores with correlation, because the ranks of each pixel in the maps are very different. This approach works well even when the contact intensity is low because magnitude of the values is not considered when converting to rank. However, Spearman's correlation is low even when the contact intensity is negligible (e.g. at an extreme, random noise will generate a very low correlation). The method does not pick up on small or focal changes in intensity, nor does it prioritize large-scale changes in intensity that do not change the map structure– the rank will stay the same even if the magnitude of the values change. Because matrices are flattened before calculating correlation, correlation also ignores the physical relationships between pixels of the map.

Stratum-adjusted Correlation Coefficient (SCC)

Contact frequency in Hi-C maps is known to exhibit a distance-dependent decay. The high similarity of the dependence pattern might bias the correlation between Hi-C maps, thus causing high, spurious correlations. SCC addresses this distance-dependence effect by stratifying Hi-C data based on genomic distance, calculating a Pearson correlation coefficient for each stratum and aggregating the weighted stratum-specific correlation coefficients with weights derived from the generalized

Cochran–Mantel–Haenszel (CMH) statistic[7]. SCC values range from -1 to 1 and share a similar interpretation as standard correlations. The equation to calculate SCC can be written as:

$$\rho_s = \sum_k w_k p_k$$

SCC was first implemented for Hi-C map comparison by Yang et al. in the R package HiCRep [7]. By including distance-aware weights, SCC is able to measure the overall reproducibility of the Hi-C matrices better than standard correlations and is resistant to decreased resolution. However, SCC is less likely to identify small changes in TAD substructures compared to some other methods surveyed here (see **Table 1**).

Structural similarity index measure (SSIM)

Structural similarity index quantifies the perceived change in structural information of two images by incorporating three terms:

Luminescence: $l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$ Contrast: $c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$ Structure: $s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$

Where the integrated SSIM score is equal to:

$$\mathrm{SSIM}(x,y) = \left[l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma\right]$$

SSIM is well-suited for identifying structural changes, and unlike correlation and MSE measures, is not biased by map contrast values. For this reason it has been incorporated into Hi-C map comparison methods previously[8]. However, SSIM is sometimes very sensitive to small changes relative to larger-scale changes that may appear more pronounced to the human eye. SSIM is also sensitive to the order of the input. It should be applied to the matrix as a whole (not vector-by-vector) as it is designed to account for neighboring values. NaN values must be interpolated or masked to zero.

Map-informed methods

Eigenvector difference

This method is inspired by genomic compartments, which are called by calculating the first eigenvector from Hi-C contact maps and assigning each genomic region to its sign [6]. Similarly, eigenvector difference is calculated from the first eigenvector that corresponds to each contact frequency map, creating a vector annotated at each bin for both maps. These vectors are then compared using spearman's rank correlation. Because the components can have different signs that are arbitrarily assigned, MSE is not used for this method as it is sensitive to these signs and would result in falsely high scores when the maps are assigned opposite signs.

Directionality Index (DI)

The Directionality Index (DI) is a measure of contact frequency bias towards either upstream sequence or downstream sequence at some DNA locus. An inflection of DI values from negative to positive and vice versa indicates a potential chromatin boundary, where DI can be calculated by:

$$DI = \frac{B-A}{|B-A|} * \left(\frac{(A-E)^2}{E} + \frac{(B-E)^2}{E}\right)$$

Where A is the number of reads (or average normalized frequency value) that map from a given locus to upstream bins, B is that value for downstream bins, and E is the expectation under the null hypothesis, equal to $\frac{(A+B)}{2}$.

DI was first proposed by Dixon et al. in 2012 [9]. It depends on two parameters: the size of the focal bin whose relative upstream and downstream contact frequency is being compared, and the size of the upstream and downstream bins (40 kb and 2 Mb in the original publication). To create a composite DI disruption score for a variant, DI is calculated for each locus in the region of both maps and compared using MSE or correlation. This composite score is subject to the caveats of the chosen comparison method.

Insulation score

Also known as the ratio score or boundary score, this method seeks to identify TAD boundary-like regions by comparing the frequency of within-region contacts upstream (A) and downstream (B) of some point X to the inter-region contacts between regions A and B [10,11]. The higher the ratio is in a given region, the more likely this region is to be a TAD boundary. The metric is calculated as follows:

$$Insulation = \frac{max(mean(A), mean(B))}{mean(X)},$$

where X quantifies the frequency of local contacts within the central region spanning 20 kilobases, and A and B represent contact frequency in the regions upstream and downstream of X, respectively, spanning 200 kilobases each.

Correlation or MSE can be applied to the insulation tracks of two conditions for a scalar disruption score. The magnitude of the insulation score is dependent on differences in contact intensity between the two maps at each bin, and therefore is sensitive to global change in contact frequency. Variants in regions of DNA with wider ranges of contact intensity (high contrast) may have inflated insulation scores relative to other regions. This method can potentially be improved by adjusting the following parameters: the size of central window X, i.e. the region for which the insulation score is being calculated (default: 20kb), and the size of upstream/downstream windows A and B (default: 200kb)

Contact probability decay

Contact decay, or the P(S) curve, measures chromatin interaction as a function of genomic distance [12,13]. Interaction frequency across the contact map is ranked by genomic distance between all pairs of contact, resulting in a track of distance vs interaction frequency. As distance increases, the probability of contact between loci decreases. Decay curves may be calculated at a given resolution such that the chromosome is divided into n = L/r bins, where L is the chromosome length and r is resolution. Across an *n* x *n* contact map, the contact frequency of each entry $A_{i,j}$ is ordered by the distance between loci, i-j. A steeper decay in contact frequency indicates a greater distance between further loci, while a shallow contact decay

suggests more interaction between distant loci. Contact decay measures a global signal of relative interaction increase or decrease, but will not be sensitive to local structural changes to contact matrices.

Triangle Method

Basic methods (correlations, MSE) all ignore the physical relationships between pixels of the map when they are flattened into vectors. They are therefore over-simplified characterizations of the relationships between maps. This method tries to leverage our understanding that contacts are represented by different subsets of triangles within the larger map to address this gap. The triangle-based method compares the average contact intensity within all sub-triangles of two contact maps. In comparison to MSE and correlations, the flattened representation of the map is the average contact intensity of all sub-triangles instead of just each pixel on its own. The flattened representations can then be compared with either MSE or a correlation method.

The performance of this method depends on the correlation or MSE used over the sub-triangles (see their individual pros and cons). The advantage over those basic methods is that triangle comparison is more feature-informed to capture relevant contact relationships. Because there are so many more smaller triangles than larger triangles, this method likely prioritizes more local changes; however, one could weight the triangles or subset to only the larger or smaller sub-triangles to prioritize only larger or smaller scale interactions. One caveat is that this method is significantly slower than other methods, but speed can be improved by creating lower resolution maps before computing.

Feature-informed methods

TADs

TAD boundaries are called by finding the local minima of the insulation profile, which is calculated using a diamond-shaped window-based method proposed by Crane et al. [10]. Specifically, a square (a W x W diamond-shaped window) is slid along each diagonal bin of the matrix and the averaged contact frequency within each window is calculated and called as insulation score. Bins with a low insulation score indicate a high insulatory effect, thus the bins reaching the local minima are identified as candidate TAD boundaries. The boundary strength is calculated for each local minima using peak prominence and candidates with strength above a threshold are referred to as TAD boundaries. The scores for the bins at the end of the diagonal and within the window size are not calculated. The overlap, gain, and loss of TAD boundaries between two Hi-C matrices are reported to show their consistency and changes. The boundary locations within a set resolution r are considered the same. This method could be further improved by changing the following parameters: window size (w), threshold of boundary strength, and upper bound of distance when two TAD boundaries are considered the one.

Loops

Chromatin loops are the positions where a pair of loci showing closer proximity compared to loci lying between them, corresponding to pixels with higher contact frequency than the ones in their neighborhood. We identify loops by comparing regions with their local background, as in HiCCUPS[14]. Specifically, for each bin in the upper triangle window of the matrix, we first check whether it is a local maximum (across neighborhood window size w) and then calculate the mean signal of center window (window size p) surrounding the bin as well as the mean signal in a donut-shape neighborhood, a lower-left neighborhood, vertical and horizontal neighborhoods around the pixel. The bins enriched above its neighborhood with ratios of mean signals of the center window to the neighborhoods higher than certain thresholds are considered as candidate loops. The bins at the corners are not considered. Loops that are the same, gained, lost between two Hi-C matrices are identified. The loops that are located within a window of size r of one another are treated as the same. This method can potentially be improved by adjusting the following

parameters: the center window size (p), window size (w), threshold of the ratio of center window to donut and lower-left filter, threshold of the ratio of center window to vertical filter, threshold of the ratio of center window to horizontal filter, and the upper bound of bin distance where two loops are considered as same one.

Supplemental References

1. Fudenberg, G., Kelley, D. R. & Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods* **17**, 1111–1117 (2020).
2. Schwessinger, R. *et al.* DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat. Methods* **17**, 1118–1124 (2020).
3. Yang, R., Das, A., Gao, V. R., Karbalayghareh, A. & Noble, W. S. Epiphany: predicting hi-c contact maps from 1d epigenomic signals. *bioRxiv* (2021).
4. Tan, J. *et al.* Cell-type-specific prediction of 3D chromatin organization enables high-throughput in silico genetic screening. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01612-8.
5. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
7. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
8. Galan, S. *et al.* CHESS enables quantitative comparison of chromatin contact data and automatic feature extraction. *Nat. Genet.* **52**, 1247–1255 (2020).
9. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
10. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
11. Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
12. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
13. Zhou, J. *et al.* Robust single-cell Hi-C clustering by convolution- and random-walk–based imputation. *Proceedings of the National Academy of Sciences* **116**, 14011–14018 (2019).
14. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).