FIREcaller: Detecting frequently interacting regions from Hi-C data

Cheynna Crowley a,b, Yuchen Yang a,e, Yunjiang Qiu c,d, Benxia Hu a,e, Armen Abnousi f, Jakub Lipiński g,h, Dariusz Plewczyński g,h, Di Wu b,i, Hyejung Won a,e, Bing Ren c,j,k, Ming Hu f,l, Yun Li a,b,i,*

Abstract

Hi-C experiments have been widely adopted to study chromatin spatial organization, which plays an essential role in genome function. We have recently identified frequently interacting regions (FIREs) and found that they are closely associated with cell-type-specific gene regulation. However, computational tools for detecting FIREs from Hi-C data are still lacking. In this work, we present FIREcaller, a stand-alone, user-friendly R package for detecting FIREs from Hi-C data. FIREcaller takes raw Hi-C contact matrices as input, performs within-sample and cross-sample normalization, and outputs continuous FIRE scores, dichotomous FIREs, and super-FIREs. Applying FIREcaller to Hi-C data from various human tissues, we demonstrate that FIREs and super-FIREs identified, in a tissue-specific manner, are closely related to gene regulation, are enriched for enhancer-promoter (E-P) interactions, tend to overlap with regions exhibiting epigenomic signatures of cis-regulatory roles, and aid the interpretation or GWAS variants. The FIREcaller package is implemented in R and freely available at https://yunliweb.its.unc.edu/FIREcaller.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Chromatin folding in the three-dimensional (3D) space is closely related to genome function [1]. In particular, gene regulation is orchestrated by a collection of cis-regulatory elements, including promoters, enhancers, insulators, and silencers. Alteration of chromatin spatial organization in the human genome can lead to gene dysregulation and consequently, complex diseases including developmental disorders and cancers [2,3].

High-throughput chromatin conformation capture (Hi-C) has been widely used to measure genome-wide chromatin spatial organization since first introduced in 2009 [4-6]. Analyzing Hi-C data has led to the discovery of structural readouts at a cascade of resolutions, including A/B compartments [6], topologically associating domains (TADs) [7], chromatin loops [8], and statistically significant long-range chromatin interactions [9-11]. Among these Hi-C readouts identified in mammalian genomes, TADs and chromatin loops are largely conserved across cell types [12,13], while A/B compartments and long-range chromatin interactions exhibit rather moderate levels of cell-type-specificity [6,7].

As an attempt to identify Hi-C readouts that are better indicative of cell type or tissue-specific chromatin spatial organizations, we have in our previous work [14], identified thousands of frequently interacting regions (FIREs) by studying a compendium of Hi-C datasets across 14 human primary tissues and 7 cell types. We defined FIREs as genomic regions with significantly higher local chromatin interactions than expected under the null hypothesis of random collisions [14].
FIREs are distinct from previously discovered Hi-C structural readouts such as A/B compartments, TADs, and chromatin loops. In general, FIREs tend to reside at the center of TADs, associate with intra-TAD enhancer-promoter (E-P) interactions, and are contained within broader regions of active chromatin [14]. FIREs are tissue and cell-type-specific, and enriched for tissue-specific enhancers and nearby tissue-specifically expressed genes, suggesting their potential relevance to tissue-specific transcription regulatory programs. FIREs are also conserved between human and mouse. In addition, FIREs have been revealed to occur near cell-identity genes and active enhancers [14]. Thus, FIREs have proven valuable in identifying tissue and cell-type-specific regulatory regions, functionally conserved regions such as enhancers shared by human and mouse, and in interpreting non-coding genetic variants associated with human complex diseases and traits [14-16].

Since the discovery of FIREs, we have collaborated with multiple groups to further demonstrate their value in various applications, resulting in multiple recent preprints and publications [16-19]. For example, in an analysis of adult and fetal cortex Hi-C datasets, FIREs and super-FIREs recapitulated key functions of tissue-specificity, such as neurogenesis in fetal cortex and core neuronal functions in adult cortex [19]. In addition, evolutionary analyses revealed that these brain FIRE regions have stronger evidence for positive selection and fewer rare genetic variants [19]. For another example, Gorkin et al. [16] investigated how 3D chromatin conformation in lymphoblastoid cell lines (LCL) varies across 20 individuals. They reported that FIREs are significantly enriched in LCL-specific enhancers, super-enhancers, and immune related biological pathways and disease ontologies, further demonstrating the close relationship between FIREs and cis-regulatory elements [16]. In particular, even with the sample size of ≤ 20 individuals, hundreds of FIRE-QTLs (that is, genetic variants associated with the strength of FIRE) have been reported, suggesting that FIREs show strong evidence of genetic regulation.

Despite the importance and utilities of FIREs, only in-house pipelines exist for detecting FIREs, limiting the general application of FIRE analysis and the full exploration of cell-type-specific chromatin spatial organization features from Hi-C data. In this work, we describe FIREcaller, a stand-alone, user-friendly R package for detecting FIREs from Hi-C data, as an implementation of the method described in our previous work [14].

2. Materials and methods

2.1. Input matrix

First, FIREcaller takes an \( n \times n \) Hi-C contact matrix \( M \) as the input (Fig. 1), which can be from a gzipped text file, or the widely used .hic or .cooler file. The contact matrix \( M \) is constructed by dividing the genome into consecutive non-overlapping bins of size \( b \) for each chromosome. In our original work [14], \( b \) was fixed at 40 Kb. In this FIREcaller work, we allow \( b \) to be 10 Kb, 20 Kb, or the default 40 Kb. Each entry in the contact matrix \( M_{ij} \), corresponds to the number of reads mapped between bin \( i \) and bin \( j \). The corresponding symmetric \( n \times n \) matrix reflects the number of mapped intra-chromosomal reads between each bin pair [6]. We removed all intra-chromosomal contacts within 15 Kb to filter out reads due to self-ligation.

Recommendations for the resolution of the input matrix depend on the sequencing depth of the input Hi-C data. Specifically, we recommend using a 10 Kb bin resolution for Hi-C data with ~2 billion reads, 20 Kb bin resolution for Hi-C data with 0.5–2 billion reads, and a 40 Kb bin resolution for Hi-C data with < 0.5 billion reads [6,8,20-23] (more details can be found in Supplement Information S1).

2.2. Cis-interaction calculation

Taking the n x n contact matrix as the input, FIREcaller calculates the total number of local cis-interactions for each bin (40 Kb default). Following our previous work [14], we define “local” to be within ~ 200 Kb by default. This threshold is largely driven by empirical evidence that contact domains exert influences on transcription regulation within 200 Kb. For instance, contact domains reported in human GM12878 cells from in-situ Hi-C data are at a median size of 185 Kb [8,20]. In addition, Jin et al. reported a median distance of E-P interactions at 124 Kb [21], Song et al. reported ~ 80% of promoter-interacting regions within 160 Kb [24], and Jung et al. found promoter-centered long-range chromatin interactions with a median distance of 158 Kb [25]. Consistently, an analysis of the dorsolateral prefrontal cortex sample [26] showed E-P interactions at a median distance of 157 Kb, and our study showed adult cortex E-P interactions at a median distance of 190 Kb [19] (Supplement Information S2). On the other hand, multiple cis-regulatory regions have been shown to control their target genes from longer genomic distances [3,19,20,27]. To accommodate these longer-range chromatin interactions, our FIRE-
caller software allows a user-specified upper bound of the cis-interacting regions.

2.3. Bin level filtering

Bins are then filtered based on multiple criteria that may lead to systematic biases, including effective restriction fragment lengths which measures the density of the restriction enzyme cut sites within each bin, GC content, and sequence uniqueness [28,29]. FIREcaller removes bins with 0 mappability, 0 GC content or 0 effective fragment length. It also removes bins for which > 25% of their neighborhood (within 200 Kb, by default) bins have 0 mappability, 0 GC content or 0 effective fragment length. In addition, any bins with a mappability < 90% are removed. Finally, any bins overlapped within the MHC region or the ENCODE blacklist regions [30] are also filtered out (Supplement Information S3).

2.4. Within-sample normalization

FIREcaller then uses the HiCNormCis method [14] to conduct within-sample normalization. HiCNormCis adopts a Poisson regression approach, adjusting for the three major sources of systematic biases: effective fragment length, GC content, and mappability [14].

As a brief summary of the HiCNormCis method, we let \( U_i, F_i, G_{Ci} \) and \( M_i \) represent the total cis-interactions (15–200 Kb, by default), effective fragment length, GC content, and mappability for bin \( i \) respectively. We assume that \( U_i \) follows a Poisson distribution, with mean \( \theta_i \), where \( \log(\theta_i) = \beta_0 + \beta_1 F_i + \beta_2 G_{Ci} + \beta_3 M_i \). After fitting the Poisson regression model, we define the residuals \( R_i \) from the Poisson regression as the normalized cis-interaction for bin \( i \) which are approximately normal (Supplement Information S4).

FIREcaller fits a Poisson regression model by default. Users can also fit a negative binomial regression model. In practice, both Poisson regression and negative binomial regression model achieve similar effect of bias removal, while Poisson regression is computationally more efficient (Supplement Information S5).

Our FIREcaller package also allows users to directly input a normalized contact map, for example, data normalized by a different normalization pipeline, via the “normalized” option. By default, normalized = FALSE, if switched to TRUE, FIREcaller will bypass this within-sample normalization step.

2.5. Across-sample normalization

If the user provided multiple Hi-C datasets, FIREcaller uses the R function normalize.quantiles in the “preprocessCore” package to perform quantile normalization of the normalized cis-interactions across samples [31].

2.6. Identifying FIREs

FIREcaller then converts the normalized cis-interactions into Z-scores, calculates one-sided p-values based on the standard normal distribution, and classifies bins with \( p < 0.05 \) as FIREs. The output file contains, for each bin, the normalized cis-interactions, the \(-\ln(p\text{-value})\) (i.e., the continuous FIRE score), and the dichotomized FIRE or non-FIRE classification.

2.7. Detecting Super-FIREs

FIREcaller also identifies contiguous FIREs, termed as super-FIRE (Fig. 2). FIREcaller first concatenates all contiguous FIRE bins by summing their \(-\ln(p\text{-value})\) (i.e., the continuous FIRE score) to quantify the overall or cumulative amount of chromatin interactions. The summed continuous FIRE scores from contiguous FIREs (which we term as super-FIRE score) are then evaluated against their rank from least interactive to most interactive, where FIREcaller determines the inflection point where the slope of the tangent line is one. Super-FIREs are defined as contiguous FIRE regions beyond the inflection point (Fig. 2B). This method is adapted from the Ranking of Super-Enhancer (ROSE) algorithm [32], which was originally proposed for the identification of super-enhancers.

2.8. Identification of differential FIREs

Similar to TADcompare to identify differential TADs [33], FIREcaller allows users to identify differential FIREs between different experimental conditions (e.g., tissues, cell lines, treatments, or developmental stages), when each condition contains at least two replicates. FIREcaller first calculates the normalized cis-interactions for each replicate, and then applies the R package “limma” to perform differential FIRE analysis. FIRE bins with fold change > 2 (in terms of the average normalized cis-interactions between conditions) and Benjamini-Hochberg adjusted \( p\text{-value} < 0.05 \) are selected as differential FIREs.

2.9. Visualizing FIREs and super-FIREs

To visualize FIREs and super-FIREs with other epigenetic data such as TAD boundaries, ChIP-seq peaks and the locations of typical enhancers and super-enhancers, FIREcaller generates a circos plot using the “circlize” package in R [34] (Supplement Information S10).

3. Results

To further demonstrate the utility of FIREcaller in terms of connecting the 3D genome structure and function, we first visualized FIREs of Hi-C datasets in Schmitt et al [14] using a virtual 4C plot (Section 3.1) in HUGIn [35], then presented novel FIRE results in fetal [36] and adult brain tissue [37] and integrated with gene expression data (Section 3.2), followed by the joint analysis of E-P interactions, and histone modifications (Section 3.2 - 3.4), as well as differential FIRE analysis (Section 3.5).

3.1. An illustrative example

We used the Hi-C data from human hippocampus tissue in our previous study [14] to showcase the utility of FIREcaller. Fig. 3 shows an illustrative example of a 400 Kb super-FIRE (merged from 10 consecutive bins, and marked by the yellow horizontal bar in the “FIREs” track), which overlaps with two hippocampus super-enhancers (indicated by the two orange horizontal bars in the “Enhancers” track). Notably, this super-FIRE contains a schizophrenia-associated GWAS SNP rs9960767 (black vertical line) [38], and largely overlaps with gene TCF4 (chr18: 52,889,562–53,332,018; pink horizontal bar depicted at the top with the color of the bar reflecting the log10 expression of the gene), which plays an important role in neurodevelopment [39]. Since rs9960767 resides within a super-FIRE with highly frequent local chromatin interactions, we hypothesize that chromatin spatial organization may play an important role in gene regulation in this region, elucidating potential mechanism by which rs9960767 affects the risk of schizophrenia.
3.2. Integrative analysis of FIREs with gene expression in human brain tissues

To study the relationship between FIREs and tissue-specifically expressed genes, we applied FIREcaller to Hi-C data from fetal [36] and adult [37] cortical tissues, and identified 3,925 fetal FIREs and 3,926 adult FIREs. Among them, 2,407 FIREs are fetal-specific and 2,408 FIREs are adult-specific (the remaining 1,518 FIREs are shared).

We then overlapped FIREs with gene promoters and found that the dynamics of FIREs across brain developmental stages are closely associated with gene regulation dynamics during brain development (Fig. 4). Specifically, we examined expression levels of genes whose promoters (defined as ±500 bp of transcription start site [TSS]) overlap with fetal brain-specific FIREs and are expressed in fetal brain, similarly genes whose promoter overlap with adult brain-specific FIREs and are expressed in adult brain. Gene expression data in both fetal and adult brain cortex are from two of our recent studies [36,37]. These criteria resulted in 707 and 882 genes in fetal and adult brain, respectively. Among them, 412 genes are fetal brain specific, 587 are adult brain specific, and 295 genes are shared (Table 1).

For the 587 genes overlapped with adult brain-specific FIREs, the mean gene expression levels, measured by \( \log_{2}(\text{FPKM}) \), are \(-0.052\) and \(0.190\) in fetal and adult brain cortex, respectively. These 587 genes are significantly up-regulated in adult brain (paired t-test \(p\)-value \(= 1.3 \times 10^{-10}\) Fig. 4; Table S5). Meanwhile, for the 412 genes overlapped with fetal brain-specific FIREs, the mean
gene expression levels are 0.551 and 0.209 in fetal and adult brain cortex, respectively. These 412 genes are significantly up-regulated in fetal brain (paired t-test p-value = 7.8 × 10⁻¹³) (Fig. 4; Table S5).

By contrast, for the 295 genes overlapped with FIREs shared between fetal and adult cortex, the mean gene expression levels are 0.328 and 0.312 in fetal and adult brain cortex, respectively. These 295 genes show no significant difference in their expression levels between fetal and adult brain (paired t-test p-value = 0.79).

Similarly, genes not overlapping with any FIREs exhibit no significant expression differences in fetal and adult brains either (paired t-test p-value = 0.85) (Fig. 4). For genes overlapped with “permuted-FIREs”, there is no significant difference in expression levels between fetal and adult brain (paired t-test p-value = 0.84) (Fig. 4).

### 3.3. Integrative analysis of FIREs and E-P interactions

We used Hi-C data from left ventricle and liver tissues from Schmitt et al study [14], and applied Fit-Hi-C [40] to call significant chromatin interactions at 40 Kb bin resolution. We only considered bin pairs within 2 Mb distance. Next, we used H3K27ac ChIP-seq peaks [41] in left ventricle and liver tissues to define active enhancers, and used 500 bp upstream / downstream of TSS to define promoters. A 40 Kb bin pair is defined as an E-P interaction if one bin contains a promoter, and the other bin contains an active enhancer. In total, at an FDR < 1%, we identified 41,401 and 30,569 E-P interactions in left ventricle and liver, respectively. Among them, 29,096 are left ventricle-specific, and 18,264 liver-specific.

We then applied FIREcaller at 40 Kb resolution, and identified 3,643 FIREs in left ventricle and 3,642 FIREs in liver, with 1,186 FIREs shared between these two tissues. We found that FIREs are enriched for E-P interactions compared to non-FIREs for both liver and left ventricle (liver: odds ratio [OR] = 7.2, Fisher’s exact test p-value < 2.2 × 10⁻¹⁶; left ventricle: OR = 4.0, p-value < 2.2 × 10⁻¹⁶). Comparing between the two tissues, we observed that left

| Tissue-specific FIREs and shared FIREs, and overlapping genes. |
|---------------------------------------------------------------|
| # FIREs | # FIREs overlapping with a gene | # of genes overlapping FIREs |
|----------|---------------------------------|-----------------------------|
| Adult-specific | 2,408                           | 488                         | 587                         |
| Fetal-specific | 2,407                        | 338                         | 412                         |
| Shared    | 1,518                           | 258                         | 295                         |

Table 1
ventricle-specific E-P interactions are highly enriched in left ventricle-specific FIREs and liver-specific E-P interactions highly enriched in liver-specific FIREs (OR = 3.8, \( p \)-value < 2.2 \( \times \) 10\(^{-16} \); Table 2). Our results demonstrate that the tissue-specificity of FIREs is closely associated with the tissue-specificity of E-P interactions [14].

### 3.4. Integrative analysis of FIREs and ChIP-seq peaks.

Next, we evaluated the relationship between FIREs and histone modifications in cortex samples [26,37,41]. We found that H3K4me3 and H3K27ac ChIP-seq peaks are both enriched at FIRE regions (Fig. 5).

### 3.5. Differential FIREs between GM12878 and H1 cells

We used FIREcaller to identify differential FIREs between GM12878 cells [8] and H1 embryonic stem cells [14], where Hi-C data for each cell type consists of two biological replicates. We identified 4,140 differential FIREs, where 2,346 FIREs are significantly more interactive in GM12878 and 1,794 more interactive in H1.

Next, we tested whether the differential FIREs are enriched for typical enhancers or super-enhancers [41] in the corresponding cell types. As expected (Fig. 6), FIREs more interactive in H1 are significantly more likely to overlap H1 typical enhancers (OR = 1.74; Fisher’s exact test \( p \)-value = 1.03 \( \times \) 10\(^{-4} \)) and super-enhancers (OR = 1.94; \( p \)-value = 0.04). Similarly FIREs more interactive in GM12878 are significantly more likely to overlap GM12878 typical enhancers (OR = 78.37; Fisher’s exact test \( p \)-value < 2.2 \( \times \) 10\(^{-16} \)), and super-enhancers (OR = 78.92; \( p \)-value < 2.2 \( \times \) 10\(^{-16} \)). We note that the odd ratios for these two cell lines differ rather drastically, which is driven by the fact that H1 FIREs are significantly, but not as strongly enriched in H1 enhancers, compared to GM12878. These results are consistent with those reported in the original Schmitt et al paper [14] where ~35% GM12878 FIREs overlapped with GM12878 typical enhancers, whereas only ~6% H1 FIREs overlapped with H1 typical enhancers (Schmitt et al Fig. 4C). Similar patterns were observed for super-enhancers (Schmitt et al Fig. 4D).

### 4. Discussion

In this paper, we present FIREcaller, a user-friendly R package to identify FIREs from Hi-C data. We demonstrate its utilities through applications to multiple Hi-C datasets and integrative analyses with E-P interactions, histone modifications and gene expression. We confirmed that FIREs are tissue/cell-type-specific, enriched of tissue/cell-type-specific enhancers, and are near tissue/cell type-specific E-P interactions are highly enriched in left ventricle-specific FIREs and liver-specific E-P interactions highly enriched in liver-specific FIREs (OR = 3.8, \( p \)-value < 2.2 \( \times \) 10\(^{-16} \); Table 2). Our results demonstrate that the tissue-specificity of FIREs is closely associated with the tissue-specificity of E-P interactions [14].

### Table 2

|                       | Left Ventricle-Specific E-P | Liver-Specific E-P |
|-----------------------|-----------------------------|-------------------|
| Left Ventricle-Specific FIRE | 1,093                      | 416               |
| Liver-Specific FIRE    | 951                         | 1,392             |

Fig. 5. H3K4me3 and H3K27ac ChIP-seq peaks are enriched at FIREs. X-axis is the distance from a bin, with the bins grouped into FIRE bins and non-FIRE bins. Y axis is fold enrichment quantified by MACS [42] when applied to the corresponding histone ChIP-seq data.

Fig. 6. Relationship between differential FIREs and cell-type-specific enhancers in GM12878 and H1 cells. The size of the dots corresponds to the OR and the color of the dots corresponds to the \( p \)-value.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.12.026.

References

[1] Dekker J. Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: integrating chromatin interaction data. Nat Rev Genet 2013;14(6):390–403. https://doi.org/10.1038/nrg3454.
[2] Krüger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol 2016;17(12):771–82. https://doi.org/10.1038/nrm.2016.138.
[3] Y. Li M. Hu Y. Shen Gene regulation in the 3D genome 27 R2 2018 2028 R233 10.1093/hmg/ddy164.
[4] Tjong H et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. Proc Natl Acad Sci USA 2016;113(12): E1663–72. https://doi.org/10.1073/pnas.1521273113.
[5] Yardimci GG et al. Measuring the reproducibility and quality of Hi-C data. Genome Biol 2019;20(1). https://doi.org/10.1186/s12859-019-1658-7.
[6] Lieberman-Aiden E et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 2009;326(5950):289.
[7] Dixon J, Gorkin D, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol Cell 2016;62(5):688–690. https://doi.org/10.1016/j.molcel.2016.05.016.
[8] Rao SSP et al. A three-dimensional map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159(7):1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
[9] Kaul A, Bhattacharyya S, Ay F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiCZ. Nat Protoc 2020;15(3):991–1012. https://doi.org/10.1038/s41596-019-02773-0.
[10] Xu Z et al. A hidden Markov random field-based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data. Bioinformatics 2016;32(3):650–6. https://doi.org/10.1093/bioinformatics/btw650.
[11] Xu Z et al. FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. Bioinformatics 2016;32 (17):2609-26. https://doi.org/10.1093/bioinformatics/btw240.
[12] Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485(7398):376-80. https://doi.org/10.1038/nature11027.
[13] Rao SF et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 2014;159(7):1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
[14] Schmitt AD et al. A Compendium of Chromatin Contact Maps Reveal Spatiotemporal Active Regions in the Human Genome. Cell reports 2016;17(8):2042–59.
[15] Burgess DJ. Epigenomics: Deciphering non-coding variation with 3D epigenomics. Nat Rev Genet 2016;18(1):1-4.
[16] Gorkin DU et al. Common DNA sequence variation influences 3-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 2016;17(12):1056–64. https://doi.org/10.1038/nrg.2016.138.
[17] Halvorsen M et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-15707-w.
[18] Giusti-Rodriguez P et al. Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. bioRxiv 2019;604330.
[19] Yu M, Ren B. The Three-Dimensional Organization of Mammalian Genomes. Annu. Rev. Cell Dev. Commun. 2017;33(1):205–80. https://doi.org/10.1146/annurev-cellbio-100616-060531.
[20] Jin F et al. A high-resolution map of three-dimensional chromatin interactome in human cells. Nature 2013;503(7475):290–4.
[21] Schmitt AD, Hu M, Ren B. Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol 2016;17:743.
[22] Lajoie BR, Dekker J, Kaplan N. The Hitchhiker’s Guide to Hi-C Analysis: Practical guidelines. Methods (San Diego, Calif.) 2015;72:65–75.
[23] Song M et al. Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. Nat Genet 2019;51(8):1252–62. https://doi.org/10.1038/s41588-019-0472-1.
[24] Jung I et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat Genet 2019;51(10):1442–9. https://doi.org/10.1038/s41588-019-0494-8.
[25] Li M et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science 2018;362(6420):eaat7615.
[26] Fulco CP et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science 2016;354(6313):769–73. https://doi.org/10.1126/science.aag2445.
[27] Hu M et al. HiCNorm: removing biases in Hi-C data via Poisson regression. Bioinformatics 2012;28(23):3131–3.
Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 2011;43(11):1059–65. https://doi.org/10.1038/ng.947.

Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep 2019;9(1). https://doi.org/10.1038/s41598-019-43839-z.

Bolstad BM et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19(2):185–93. https://doi.org/10.1093/bioinformatics/19.2.185.

Whyte W et al. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell 2013;153(2):307–19. https://doi.org/10.1016/j.cell.2013.03.035.

Cresswell KG, Dozmorov MG. TADCompare: An R Package for Differential and Temporal Analysis of Topologically Associated Domains. Front Genet 2020;11:158.

Gu Z et al. circlize Implements and enhances circular visualization in R. Bioinformatics 2014;30(19):2811–2.

Martin JS et al. HUGIn: Hi-C Unifying Genomic Interrogator. Bioinformatics 2017.

Won H et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 2016;538(7626):523–7. https://doi.org/10.1038/nature19847.

D. Wang et al. Comprehensive functional genomic resource and integrative model for the human brain Science (New York, N.Y.) 362 6420 2018 p. eaat8464

Stefansson H et al. Common variants conferring risk of schizophrenia. Nature 2009;460(7256):744–7. https://doi.org/10.1038/nature08186.

Forrest MP et al. The emerging roles of TCF4 in disease and development. Trends Mol Med 2014;20(6):322–31. https://doi.org/10.1016/j.molmed.2014.01.010.

Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res 2014;24(6):999–1011. https://doi.org/10.1101/gr.160374.113.

Kundaje A et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–30. https://doi.org/10.1038/nature14248.

Zhang Y et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.