2016

# A compendium of chromatin contact maps reveals spatially active regions in the human genome

Anthony D. Schmitt
*University of California - San Diego*

Ming Hu
*New York University School of Medicine*

Inkyung Jung
*Ludwig Institute for Cancer Research*

Zheng Xu
*University of North Carolina at Chapel Hill*

Yunjiang Qiu
*University of California - San Diego*

**See next page for additional authors**

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs
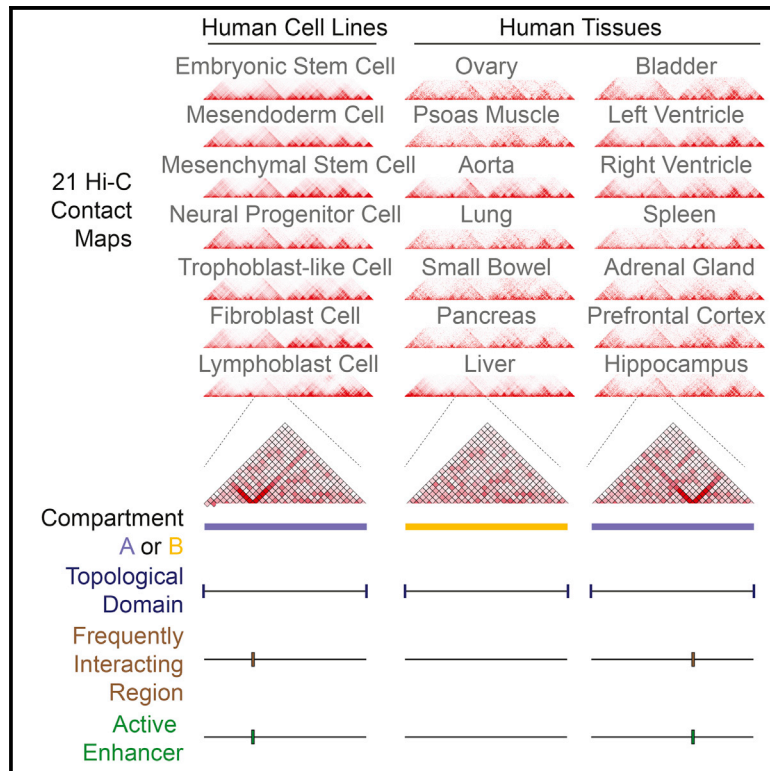
**Authors**

Anthony D. Schmitt, Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L. Tan, Yun Li, Shin Lin, Yiing Lin, Cathy L. Barr, and Bing Ren

# Cell Reports

# A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

## Graphical Abstract

## Authors

Anthony D. Schmitt, Ming Hu, Inkyung Jung, ..., Yiing Lin, Cathy L. Barr, Bing Ren

## Correspondence

hum@ccf.org (M.H.), biren@ucsd.edu (B.R.)

## In Brief

Schmitt et al. analyze Hi-C maps in 21 human cell lines and primary tissues and uncover a class of genome organizational features termed FIREs. FIREs are local interaction hotspots, highly tissue-specific, and correspond to active enhancers. We discuss the implications of our findings for the study of gene regulation and disease. Explore the Cell Press IHEC web portal at http://www.cell.com/consortium/IHEC.

## Highlights

- Integrative analysis of chromatin architecture in a broad set of human tissues

- FIREs are an architectural feature of chromatin organization

- FIREs are enriched for super-enhancers and show tissue-specific chromatin interactions

- FIRE formation is partially dependent on CTCF and the Cohesin complex

## Accession Numbers

GSE87112

CellPress

# A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

[1]Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA
[2]UCSD Biomedical Sciences Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[3]Division of Biostatistics, Department of Population Health, New York University School of Medicine, 650 First Avenue, New York, NY 10016, USA
[4]Departments of Genetics, Biostatistics, and Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA
[5]USCD Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[6]Division of Cardiology, Department of Medicine, University of Washington, 850 Republican Street, Seattle, WA 98108, USA
[7]Department of Surgery, Washington University School of Medicine, 660 S Euclid Ave., Campus Box 8109, St. Louis, MO 63110, USA
[8]Krembil Research Institute University Health Network, The Hospital for Sick Children, The University of Toronto, Krembil Discovery Tower, 60 Leonard Ave. 8KD-412, Toronto, ON M5T 2S8, Canada
[9]Department of Cellular and Molecular Medicine, Moores Cancer Center and Institute of Genome Medicine, UCSD School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA
[10]Quantitative Life Sciences Initiative, University of Nebraska, Lincoln, NE 68583, USA
[11]Department of Statistics, University of Nebraska, Lincoln, NE 68583, USA
[12]Co-first author
[13]Present address: Arima Genomics Inc., 6404 Nancy Ridge Dr., San Diego, CA, 92121, USA
[14]Present address: Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA
[15]Present address: Department of Biological Sciences, KAIST, Daejeon 34141, South Korea
[16]Lead Contact
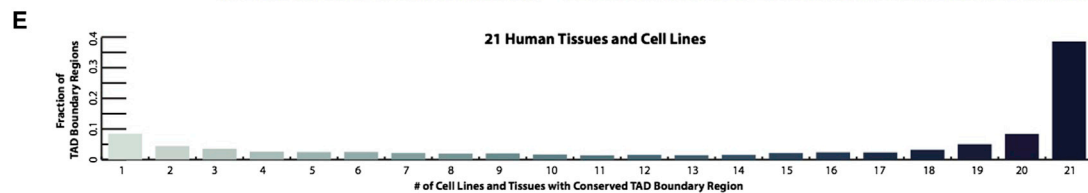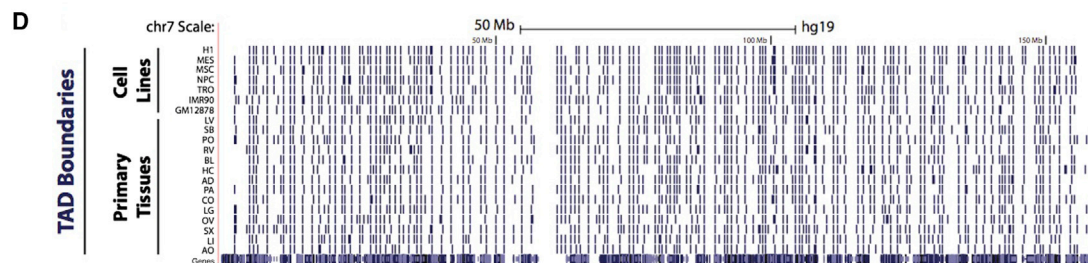*Correspondence: hum@ccf.org (M.H.), biren@ucsd.edu (B.R.)
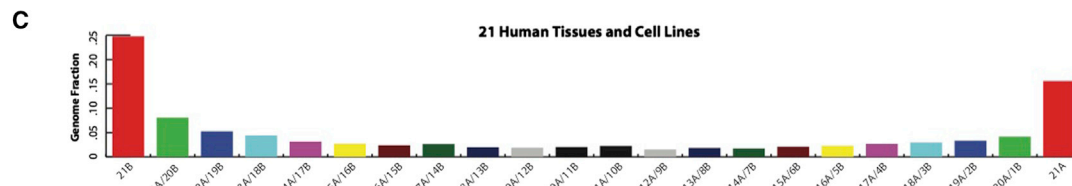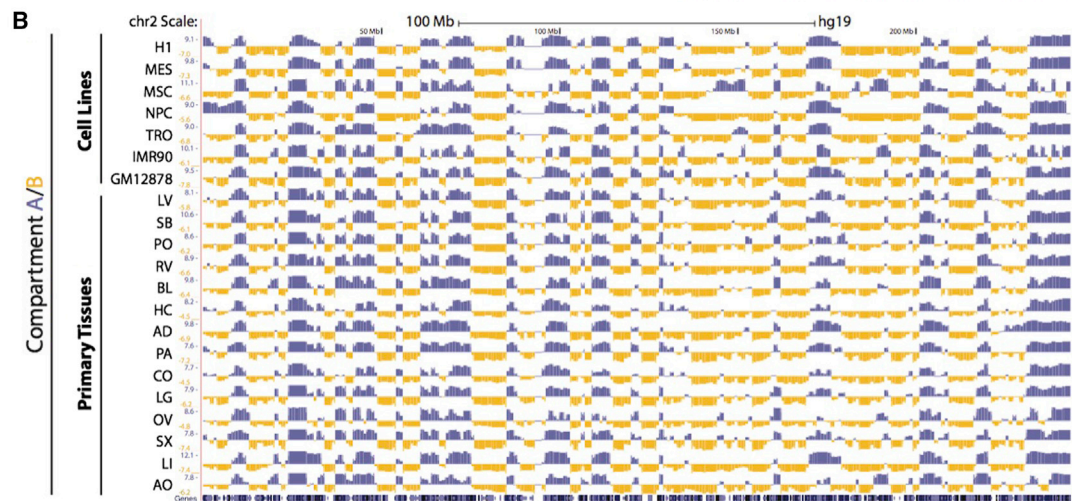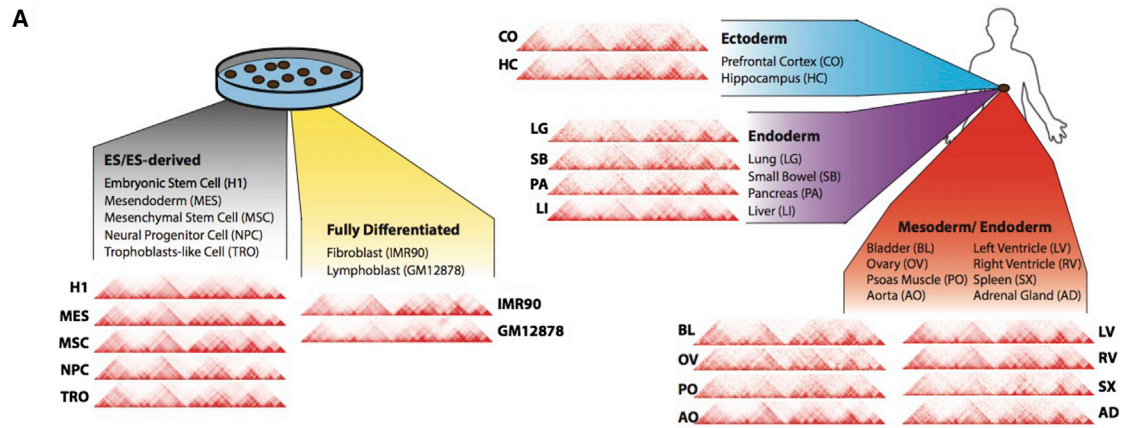http://dx.doi.org/10.1016/j.celrep.2016.10.061

## SUMMARY

The three-dimensional configuration of DNA is integral to all nuclear processes in eukaryotes, yet our knowledge of the chromosome architecture is still limited. Genome-wide chromosome conformation capture studies have uncovered features of chromatin organization in cultured cells, but genome architecture in human tissues has yet to be explored. Here, we report the most comprehensive survey to date of chromatin organization in human tissues. Through integrative analysis of chromatin contact maps in 21 primary human tissues and cell types, we find topologically associating domains highly conserved in different tissues. We also discover genomic regions that exhibit unusually high levels of local chromatin interactions. These frequently interacting regions (FIREs) are enriched for super-enhancers and are near tissue-specifically expressed genes. They display strong tissue-specificity in local chromatin interactions. Additionally, FIRE formation is partially dependent on CTCF and the Cohesin complex. We further show that FIREs can help annotate the function of non-coding sequence variants.

## INTRODUCTION

Chromosome conformation capture (3C)-based techniques have begun to reveal molecular details of nuclear organization in eukaryotic cells (Dekker et al., 2002; Dixon et al., 2012, 2015; Dostie et al., 2006; Fraser et al., 2015; Jin et al., 2013; Lieberman-Aiden et al., 2009; Rao et al., 2014; Seitan et al., 2013; Simonis et al., 2006; Sofueva et al., 2013; Vietri Rudan et al., 2015; Zuin et al., 2014). It is now clear that each chromosome occupies a separate space in the interphase nucleus, known as a "chromosome territory," which is partitioned into distinct neighborhoods or compartments (Lieberman-Aiden et al., 2009; Meaburn and Misteli, 2007). Within each compartment, topologically associating domains (TADs) constrain chromatin interactions (Dixon et al., 2012, 2016; Nora et al., 2012; Sexton et al., 2012). Within each TAD, chromatin interactions between distal *cis*-regulatory elements occur in a cell-type-dependent manner to allow modulation of promoter activity by enhancers (Dryden et al., 2014; Montavon and Duboule, 2013; Phillips-Cremins et al., 2013; Simonis et al., 2006; Tang et al., 2015). Previous 3D genome analyses have been largely limited to cultured cells and a small collection of primary cell types. By contrast, our knowledge of chromatin organization in human tissues is still scarce. Variation in chromatin interaction patterns among diverse tissue types remains poorly defined, and its functional relationship with gene regulation remains to be characterized. This is a critical shortcoming because diseases pertaining to

CrossMark

A

ES/ES-derived

**Embryonic Stem Cell (H1)**
Mesendoderm (MES)
Mesenchymal Stem Cell (MSC)
Neural Progenitor Cell (NPC)
Trophoblasts-like Cell (TRO)

**Fully Differentiated**
Fibroblast (IMR90)
Lymphoblast (GM12878)

CO
HC
**Ectoderm**
Prefrontal Cortex (CO)
Hippocampus (HC)

LG
SB
PA
LI
**Endoderm**
Lung (LG)
Small Bowel (SB)
Pancreas (PA)
Liver (LI)

**Mesoderm/ Endoderm**
Bladder (BL)      Left Ventricle (LV)
Ovary (OV)        Right Ventricle (RV)
Psoas Muscle (PO) Spleen (SX)
Aorta (AO)        Adrenal Gland (AD)

H1
MES
MSC
NPC
TRO

IMR90
GM12878

BL
OV
PO
AO

LV
RV
SX
AD

B  Compartment A/B — Cell Lines / Primary Tissues

C  21 Human Tissues and Cell Lines

D  TAD Boundaries — Cell Lines / Primary Tissues

E  21 Human Tissues and Cell Lines

*(legend on next page)*

specific organ systems are often not easy to recapitulate in vitro. Therefore, systematic characterization of chromosome architecture across a broad set of well-annotated primary tissues could be of great value for further study of genome function.

Recent studies of chromatin modification landscapes across a large number of human tissues and cell types have greatly improved our understanding of genome function and regulation (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium et al., 2015). The research has revealed that over 12% of the genome possesses cell-type-specific chromatin signatures consistent with them acting as *cis*-regulatory sequences. However, to better understand how these DNA sequences contribute to tissue- and cell-type-specific gene expression patterns, it is necessary to characterize the chromatin architecture in each tissue. Here, we report integrative analysis of chromatin organization maps of 14 human tissues and 7 human cell lines for which complete epigenome datasets have been generated by the Epigenome Roadmap Consortium, ENCODE, or the National Institute of Child Health and Human Development (NICHD) (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium et al., 2015). We developed a computational method to discover the spatially active chromatin segments termed frequently interacting regions (FIREs). We find FIREs are enriched for active enhancer regions, harboring super-enhancers as well as disease-associated variants in the corresponding disease-relevant tissue type. In addition, FIREs are substantially conserved between human and mouse genomes of the same cell type, and their formation depends in part on the Cohesin complex and CTCF. Finally, most FIREs exhibit promiscuous interactions in the local chromatin neighborhood. These observations improve our understanding of the role of dynamic chromatin organization in the regulation of tissue-specific gene expression programs in human cells.

## RESULTS

### Compendium of Chromatin Organization Maps across 21 Human Cell and Tissue Types

We conducted Hi-C analysis on 14 primary human tissues collected from four donors (Figure 1A), for which epigenome datasets had been produced as part of the NIH Epigenome Roadmap project (Roadmap Epigenomics Consortium et al., 2015). We

combined the resulting datasets with those previously generated by us for seven cultured cell types using a common experimental protocol that was reported separately (Dixon et al., 2012, 2015; Jin et al., 2013; Selvaraj et al., 2013). The combined datasets were processed using a common data processing pipeline, after merging data from biological replicates deemed as reproducible (Figures S1A–S1E). Collectively, we analyzed >8.6 billion unique contacts, out of which >2.5 billion were long-range (>15 kb) intra-chromosomal contacts, with 809M unique contacts and 254M long-range *cis* contacts per cell line and 214M unique contacts and 53M long-range *cis* contacts per tissue type (Table S1). We first analyzed compartment A/B patterns in each tissue/cell type (Figure 1B; Table S2). As previously reported for cultured human cells (Dixon et al., 2015), we observed substantial compartment A/B switching across primary tissues (Figures 1B and 1C), finding that 59.6% of the genome is dynamically compartmentalized in different tissues and cell types. These data also underscore the significant degree of compartment conservation across the genome, revealing that as much as 40.4% of the genome is invariant, which is a statistically significant degree of invariant genome compartmentalization (chi-square test p value < 2.2e−16) (Figure S1F).

TADs have been reported to be stable across different cell types and experimental conditions and conserved in related species (Dixon et al., 2012, 2015; Rao et al., 2014; Zuin et al., 2014). To investigate the degree of TAD boundary conservation in primary human tissues, we applied the insulation score method (Crane et al., 2015), which is robust in sequencing depth (Figures S1G–S1I) to identify TAD boundaries at 40-kb bin resolution (Table S3). We identified a total of 3,010 distinct TAD boundaries in 21 samples (14 tissues and 7 cell lines). Upon careful inspection of a broad panel of genetic loci (Figures 1A and 1D) as well as systematic comparison across samples (Figures 1D and 1E), we find that TAD boundaries are indeed highly conserved across different cell lines and tissues. These results are highly significant, considering that, by chance, only 1.7% of TAD boundaries are expected to share for all (chi-square test p value < 2.2e−16).

### Identification of Frequently Interacting Regions in the Human Genome

As a means to investigate conserved and tissue-specific chromatin interactions, we first used Fit-Hi-C (Ay et al., 2014) to

---

**Figure 1. Global Features of 3D Genome Organization in 7 Cell Lines and 14 Adult Tissues**

(A) Illustration of the primary 21 Hi-C datasets analyzed, depicting the cell (left panel) or tissue (right panel) origin of the samples as well as the germ layer origin for tissues (right panel). Hi-C interaction patterns across an 11.68-Mb region (chr12:82,840,000–94,520,000) are shown for all 7 cell lines and 14 tissues at 40-kb bin resolution.

(B) Genome browser snapshot showing compartment A/B patterns (PC1 value) across chromosome 2 in 21 samples, with 7 cell lines at the top and 14 primary adult tissues on the bottom. Compartment A/B patterns are at 1-Mb bin resolution. Positive PC1 in blue corresponds to compartment A, and negative PC1 in yellow corresponds to compartment B.

(C) Bar plots showing the degree of conservation of A/B compartment labels of 21 human cell lines and adult tissues. The y axis is the fraction of the genome conserved by the 22 possible combinations of compartment A/B designations. The label below each bar represents the composition of the compartment designations. For example, "16A/5B" represents the genomic region where 16 samples exhibit a compartment A label and the other five samples exhibit a compartment B label.

(D) Genome browser snapshot showing topological domain boundaries across chromosome 7 in 21 samples, with 7 cell lines at the top and 14 primary adult tissues on the bottom. Boundaries are identified at 40-kb bin resolution.

(E) Bar plots showing the degree of topological domain boundary conservation across 21 human cell lines and tissues. For each putative boundary region, we tallied how many samples have a boundary within that region (see Supplemental Experimental Procedures). Shown here is a total fraction of TAD boundary regions, whereby the y axis is the fraction of TAD boundaries conserved at least a certain number of samples, as categorized along the x axis.
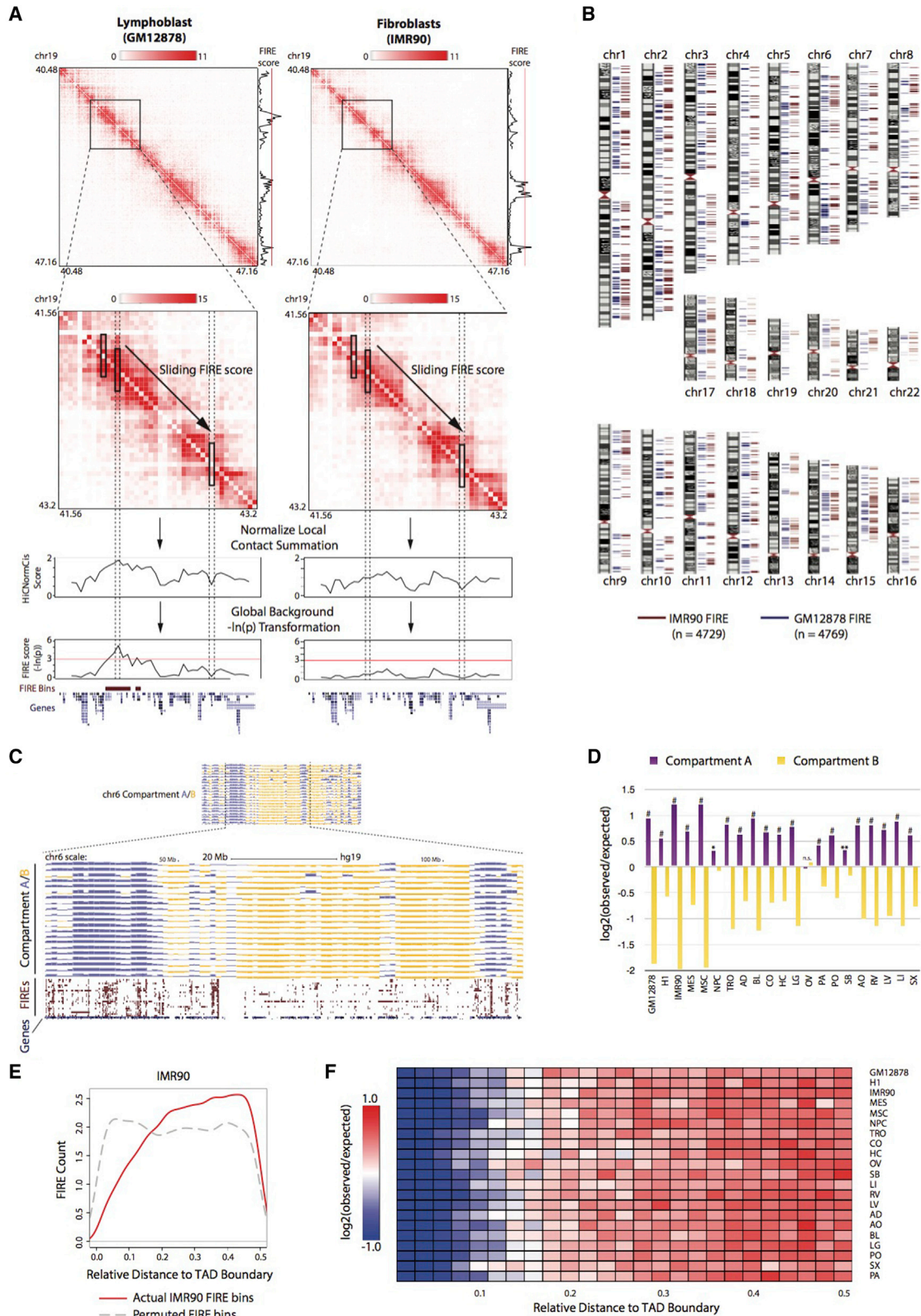
(legend on next page)

identify significant chromatin interactions at various significance thresholds (Table S4). However, Fit-Hi-C, like other peak-calling methods (Jin et al., 2013; Rao et al., 2014; Xu et al., 2015, 2016), is sensitive to sequencing depth, and therefore we found considerable variation in total chromatin contacts between samples, precluding any statistically rigorous comparative peak-calling analysis across tissues. However, upon closer examination of the chromatin contacts near the contact matrix diagonal (± 200 kb from the matrix diagonal), we noticed that some regions exhibit unusually high levels of local contact frequency in a tissue-type-dependent manner (Figure 2A). We therefore developed a computational approach to normalize and compare local interaction frequencies across all 21 tissues and cell types. Specifically, we developed a Poisson-regression-based normalization approach (termed as "HiCNormCis") to normalize the total raw local (15–200 kb) *cis* contacts for each 40-kb bin genome-wide (Figure S2A; Supplemental Experimental Procedures). This method removes bias from three sources known to affect Hi-C data, including effective restriction fragment lengths, GC content, and sequence mappability (Hu et al., 2012; Yaffe and Tanay, 2011). Compared to other normalization approaches, such as HiCNorm (Hu et al., 2012), vanilla coverage (Lieberman-Aiden et al., 2009), and iterative correction and eigenvector decomposition (ICE) (Imakaev et al., 2012), HiCNormCis achieved the best performance for bias removal (Figure S2B). Lastly, we used a Gaussian distribution to approximate the normalized total local *cis* contacts (Figure S2C), and converted HiCNormCis output values to −ln(p value), which we define as the final "FIRE score." FIREs (also termed "FIRE bins") are therefore defined as bins with a one-sided p value less than 0.05, corresponding to −ln(p value) greater than 3 (Figure 2A). We found that our FIRE scores were highly reproducible (Figures S2D and S2E), and robust to sequencing depth (Figures S2A and S2F), choice of restriction enzymes in Hi-C library preparation (Figures S2G and S2H), as well as choice of experimental protocols, such as dilution Hi-C or in situ Hi-C (Figure S2I).

We first identified FIREs in GM12878 and IMR90 cells (Figures 2A and 2B). Global analysis of FIREs revealed a dispersed distribution along the genome (Figure 2B). We next determined FIREs

in the remainder of tissues and cell lines (Tables S5 and S6) after removing local genomic feature biases (Figure S2J). We then explored how FIREs are positioned in relation to A or B compartments as well as in relation to TAD boundaries (not chromatin "loops"). Careful inspection of FIRE positioning and genome-wide enrichment analyses indicated that FIREs are enriched in compartment A and depleted in compartment B (Figures 2C and 2D; Table S7). We also examined the FIRE distribution within TADs, and found that FIREs are depleted near TAD boundaries and enriched within TADs and toward the TAD center (Figures 2E and 2F).

### FIREs, Chromatin Loops, and Insulated Neighborhoods

We further analyzed FIREs at 5-kb resolution using previously published in situ Hi-C data in IMR90 and GM12878 (Rao et al., 2014), and compared FIRE positioning relative to the smaller (~185 kb) chromatin "loops." As expected, FIREs are significantly enriched for chromatin loop anchors (chi-square test p value < 2.2e−16); however, ~90% of FIREs are within loops, and these FIREs demonstrate unique properties to be discussed in the following sections. Our data indicate that FIREs are hotspots of local chromatin interactions that are distinct from compartments, TADs, and chromatin loops (Rao et al., 2014), which are generally anchored by convergent CTCF binding. By contrast, most FIREs are located within TADs and chromatin loops, indicating they represent specific loci "within the loop" at higher resolution. Similarly, FIREs are likely distinct from insulated neighborhoods due to the high positional overlap between the CTCF-mediated "chromatin loops" and "insulated neighborhoods" (Ji et al., 2016). Our analysis of FIREs and insulated neighborhoods at 40-kb resolution in H1 cells indicates that insulated neighborhoods are also enriched for FIREs (chi-square test p value = 5.32e−15), but >70% of insulated neighborhoods do not contain a FIRE (Figure S3D) (also discussed more below).

### FIREs Are Tissue-Specific and Located Near Cell Identity Genes

To characterize the tissue-specificity of FIREs, we combined all 21 datasets (7 cell lines and 14 tissues), and performed a

**Figure 2. Identification and Positional Enrichment of Frequently Interacting Regions**

(A) Illustrative examples showing the FIRE score methodology. Hi-C contact maps from a 6.68-Mb region (chr19:40,480,000–47,160,000) are shown for GM12878 and IMR90 cells at 40-kb bin resolution (top). To the right of the contact maps are line plots showing the fully processed FIRE score for each 40-kb bin. A red line is drawn at the significance cutoff. The second row of contact maps illustrates FIRE scores in a sub-matrix (chr19:41,560,000–43,200,000) of the above contact maps (black box). Line plots directly below show the intermediate stage in the FIRE score calculation, which is the output from HiCNormCis (see Supplemental Experimental Procedures). Genome-wide HiCNormCis normalized counts are then *Z* score transformed and converted to a −ln(p value) scale to obtain the final FIRE score (bottom line plots). Dashed columns highlight two 40-kb bins, one showing a FIRE peak in GM12878 cells, but not in IMR90 cells, and the other showing a low FIRE score in both cell types.
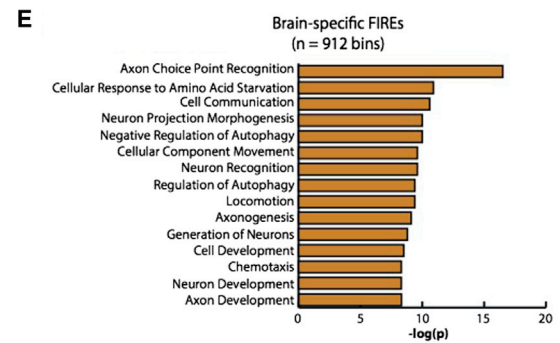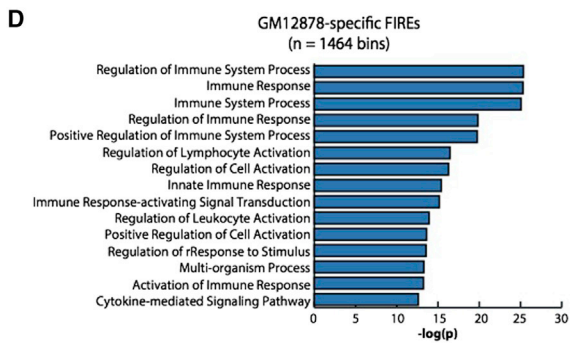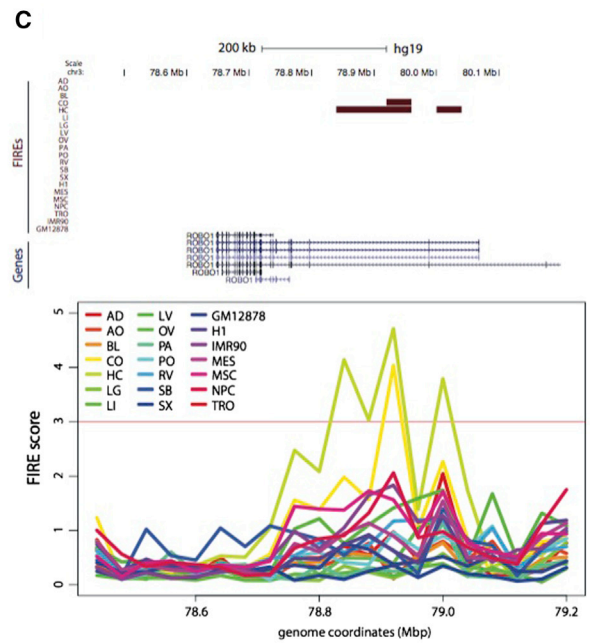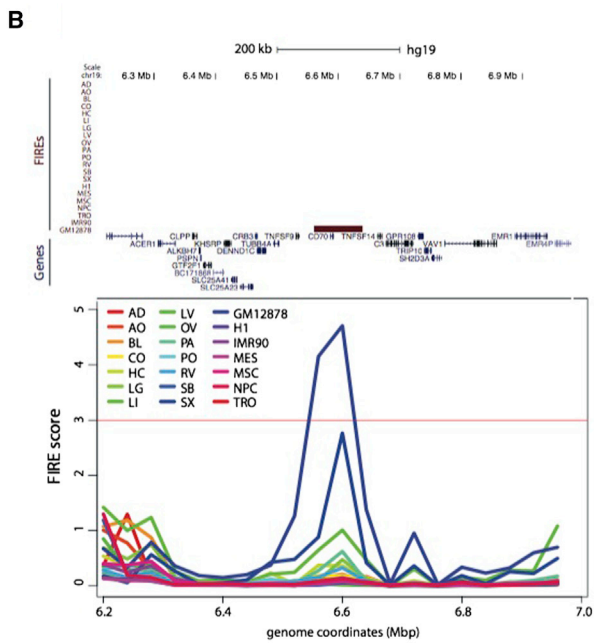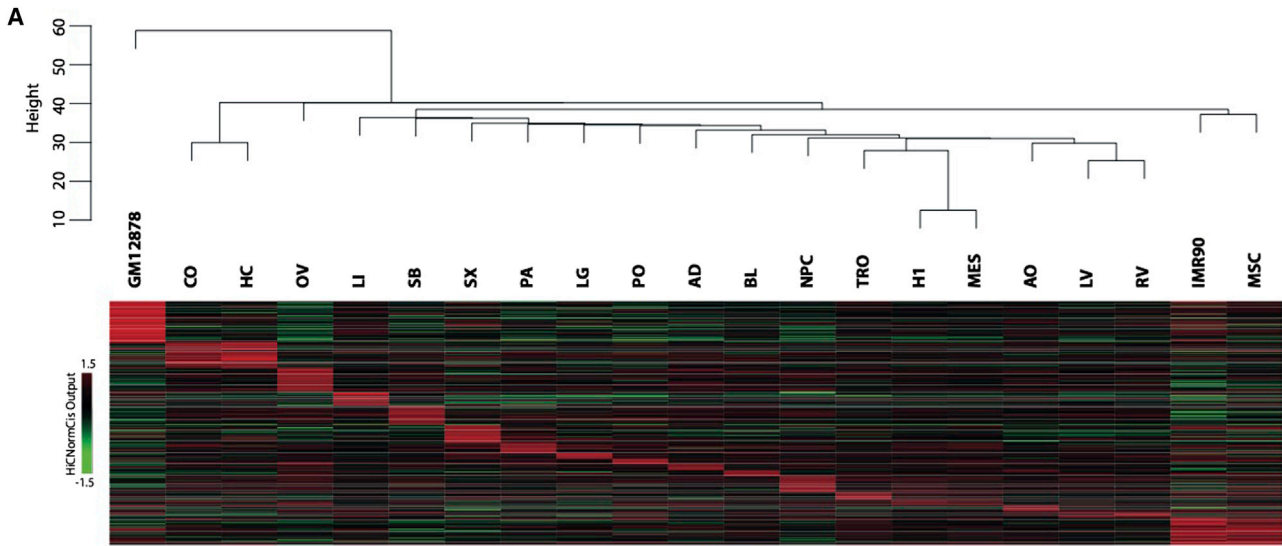
(B) Chromosome ideograms showing the genome-wide positional distribution of FIRE bins in GM12878 (blue, n = 4,769) and IMR90 (maroon, n = 4,729). Genome-wide visualization captures both conserved and specific FIRE bins. Only autosomes are depicted.

(C) Genome browser snapshot of compartment A/B patterns in 21 samples across chromosome 6 (top), and a genome browser snapshot of a 90-Mb subset of chromosome 6 (chr6:25,000,000–115,000,000) showing compartment A/B patterns for 21 samples (top set, blue/yellow) and FIRE calls (bottom set, maroon).

(D) Bar plots showing an enrichment analysis of FIRE positioning within either compartment A or B, illustrating FIREs are enriched in compartment A and depleted in compartment B compared to random permutation of the FIRE bin location within each sample (*p < 5.0e−7; **p < 7.0e−13; #p < 2.2e−16; chi-square test). Statistical tests correspond to the significance of FIRE enrichment in compartment A.

(E) Line plot showing an example of IMR90 FIRE bin positioning relative to TADs (see Supplemental Experimental Procedures). The red line depicts the observed counts (y axis) of actual IMR90 FIRE bins, whereas the gray dashed line shows the counts of permuted FIRE bin locations. The x axis ranges from 0 to 0.5, where 0 represents TAD boundaries and 0.5 represents TAD center points.

(F) Heat map showing the TAD position enrichment analysis across all 21 samples. Shown are the log2(observed/expected) values for each distance increment, as computed in (E).

**A**

(n = 8731 bins)

**B**

**C**

**D** GM12878-specific FIREs
(n = 1464 bins)

**E** Brain-specific FIREs
(n = 912 bins)

*(legend on next page)*

comparative analysis (Figure 3A; Table S6). Approximately 38.8% (8,142/20,974 bins) of FIREs were identified in only one tissue or cell type, and approximately 57.7% (12,094/20,974 bins) of FIREs were identified in two or fewer, revealing the highly tissue-specific nature of FIREs (Figure S2K). Further, a hierarchical clustering analysis of genome-wide FIRE scores revealed similarities among certain cell types, such as H1 and MES, as well as MSC and IMR90 (Dixon et al., 2015) (Figure 3A). As expected, tissues from the same organ (brain: cortex and hippocampus; heart: left ventricle and right ventricle) clustered together (Figure 3A). Tissue-specific FIREs tend to be positioned in close proximity to genes related to the cellular identity (Figures 3B and 3C). For example, within a GM12878-specific FIRE is the promoter for *CD70*, a gene well known for its role in immune cell activation and maturation (Arens et al., 2004) (Figure 3B). Moreover, ~110 kb from a FIRE region present only in brain tissues is an alternative *ROBO1* promoter, a gene involved in axon guidance during development (Leyva-Díaz et al., 2014) (Figure 3C). To extend these observations to all tissue-specific FIREs and to interpret the functional roles and disease relatedness of these FIREs, we performed GREAT analysis (McLean et al., 2010) (Tables S8 and S9). The results showed that genes in close proximity to tissue-specific FIREs are related to the functionality of that tissue/cell type (Figures 3D and 3E; Tables S8 and S9). Moreover, using only our 5-kb resolution FIRE calls in GM12878 and IMR90, we also found abundant sample-specific FIREs (~57% of FIREs are sample specific), and confirmed that sample-specific FIREs are positioned near cell identity genes (Tables S8 and S9) at a higher resolution. Collectively, these results suggest that FIREs are closely associated with cell identity and tissue function.

### FIREs Are Enriched for Active Enhancers and Super-Enhancers

Because FIREs tend to be positioned near genes related to cell identity and tissue function, we posited that FIREs may be enriched for active enhancers. To test this hypothesis, we analyzed previously generated ChIP-seq data for six histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3, and H3K9me3) for these tissues and cell types (Roadmap Epigenomics Consortium et al., 2015). We observed that FIREs display a high density of active chromatin features (e.g., H3K27ac and H3K4me1), and overlap with super-enhancers found in the same tissues (Hnisz et al., 2013) (Figure 4A). We then characterized the histone modification signatures across 1-Mb regions centered at FIREs. FIREs are ubiquitously enriched for two active enhancer marks, H3K4me1 and H3K27ac, and depleted for the repressive chromatin mark H3K27me3 (Figure 4B), whereas enrichment of other marks did not show clear patterns (Figure S3A). FIREs also overlap with typical enhancers and super-enhancers (Hnisz et al., 2013) annotated in the cell lines and tissues where such data are available (Figures 4C and 4D). For example, 35.0% of typical enhancers and 77.8% of super-enhancers annotated in GM12878 cells overlap FIREs (Fisher's exact test p value < 2.2e−16) (Figures 4C and 4D). Importantly, we also found significant enrichment for FIREs at typical enhancers and super-enhancers (chi-square test p value < 2.2e−16) when analyzing FIREs at 5-kb bin resolution (Table S6) using previously published high-resolution Hi-C data in GM12878 and IMR90 (Rao et al., 2014) (Figure S3B). Also, with respect to previously annotated chromatin loops (Rao et al., 2014), we find that the aforementioned 90% of FIREs that do not overlap loop anchors are also significantly enriched for typical and super-enhancers (chi-square test p value < 2.2e−16). For example, we observed GM12878-specific FIREs corresponding to a GM12878-specific super-enhancer, whereas the same locus in IMR90 lacks any enhancer or FIRE, despite sharing a conserved chromatin loop (Figure S3C). These FIRE analyses at 5-kb resolution corroborate our findings at 40-kb resolution, and indicate that FIREs represent distinct structural entities with differing biochemical properties compared to chromatin loops. As anticipated, we also find a significant overlap between FIREs and super-enhancer domains in mouse embryonic stem cells (mESCs) at 40-kb resolution (chi-square test p value = 0.0052), but not polycomb domains (Dowen et al., 2014; Ji et al., 2016), further underscoring the role of FIREs in active gene regulation (Figure S3D).

Because many FIRE bins were found in clusters, we stitched together adjacent FIRE bins and ranked them by cumulative Z score, revealing that a small proportion of FIRE clusters (termed "super-FIREs") contain the majority of bins with the most significant local interaction frequency (Figure S3E). Strikingly, compared to all FIREs (Figure S3F), we observed some tissues, in which nearly 100% of super-FIREs contain either a super-enhancer or typical enhancer (Figure S3G), suggesting that the bins with the highest local interaction frequency almost always mark active enhancer(s). Analysis of super-FIREs not containing an enhancer revealed a moderate enrichment for H3K27me3 across most testable samples, but no other clear trends (Figures S3H–S3M). Given this striking relationship, we wondered to what

---

**Figure 3. FIREs Are Tissue-type Specific and Enriched Near Genes Involved in Tissue Function**

(A) At the top is a dendrogram resulting from a hierarchical clustering analysis using genome-wide FIRE scores for each sample. The y axis is the Euclidean distance between FIRE scores from any two samples. The heat map below shows a subset of FIRE bins (n = 8,371), corresponding to FIRE bins that are called as FIRE in only one or two samples. For ventricle tissues, brain tissues, IMR90/MSC, and H1/MES, FIREs specific to two samples are allowed in the definition of sample specific.

(B) Genome browser snapshot showing a GM12878-specific FIRE region (chr19:6,560,000–6,640,000) (top, maroon) in an 800-kb region around *CD70* (chr19:6,583,193–6,604,114). Below is a line plot of FIRE scores for each sample, showing the GM12878-specific FIRE peak (blue).

(C) Genome browser snapshot showing a brain-specific FIRE region (chr3:78,920,000–78,960,000), shared by CO and HC, in a 760-kb region within *ROBO1* (chr3:78,646,338–79,068,609). Below is a line plot of FIRE scores for each tissue showing CO (yellow) and HC (pea green) FIRE peaks.

(D) GREAT biological process analysis of genes surrounding GM12878-specific FIRE bins (n = 1,464 bins), showing biological processes highly related to immune functions. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.

(E) Same as (D), except using genes surrounding brain (CO and HC) specific FIRE bins (n = 912 FIRE bins) showing several significant processes highly related to brain functionality. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.
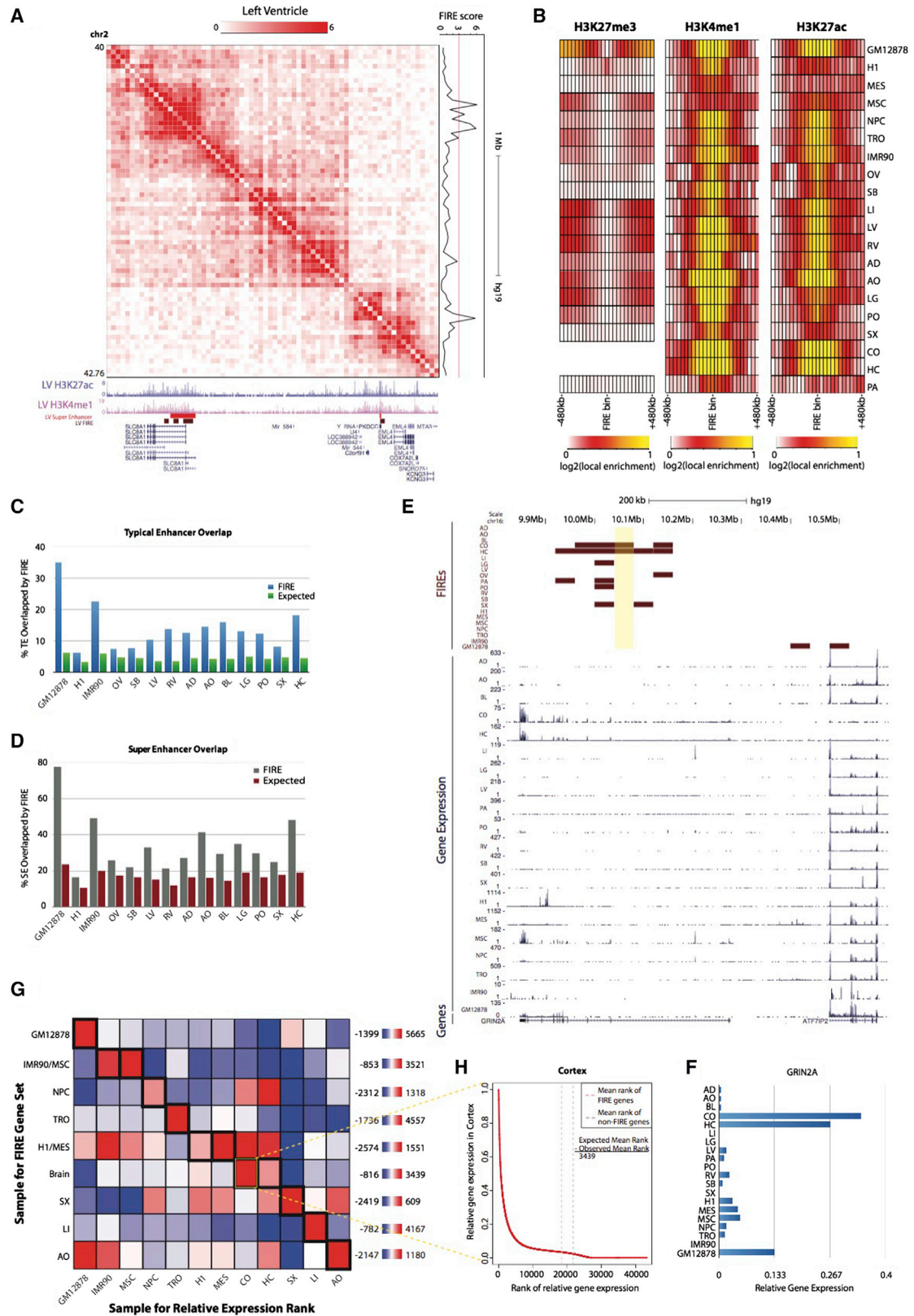
(legend on next page)

extent FIRE analysis could be used to predict the locations of typical and super-enhancers in GM12878. By varying the significance thresholds for FIRE calling and performing a receiver operating characteristic (ROC) area under curve (AUC) analysis, we find an impressive predictive power of FIRE analysis to identify typical enhancers and super-enhancers using Hi-C data alone (AUC = 0.813 and AUC = 0.906, respectively) (Figures S3N and S3O). Taken together, the high overlap between super-enhancers and FIREs, as well as the enrichment of tissue identity genes near tissue-specific FIREs, implicates a potential *cis*-regulatory role for FIREs in facilitating tissue-specific gene expression.

### FIREs Are Near Tissue-Specifically Expressed Genes

Because super-enhancers are known to be tissue-specific and positioned near cell identity genes, we asked if FIREs are nearby genes that are more transcriptionally active in the corresponding tissue/cell types. By re-analyzing publicly available RNA-seq data (Roadmap Epigenomics Consortium et al., 2015), we indeed found a strong correlation between cell/tissue-specific FIREs and cell/tissue-specific expression of nearby genes. For example, the *GRIN2A* gene, which encodes an important ligand- and voltage-gated N-methyl-D-asparate (NMDA) receptor subunit implicated in epilepsy (Kingwell, 2013) and schizophrenia (Ohi et al., 2016), is predominantly expressed in brain tissues, and the transcription start site (TSS) is ~197 kb from a brain-specific FIRE (Figure 4E). In *GRIN2A*, the relative gene expression in cortex (CO) is the highest among all tissues (Figure 4F; see Supplemental Experimental Procedures). We also calculated the relative gene expression for each gene within 200 kb of a tissue-specific FIRE across all tissues and found significant correlation between tissue-specific FIREs and tissue-specifically expressed genes (Figure S3P). For example, we found that the GM12878-specific FIRE gene set contained genes with significantly higher relative expression in GM12878 compared to any other FIRE gene set (two-sample t test p value < 9.26e−6) (Figure S3P).

Intrigued by these observations in brain tissue and lymphoblast cells, we applied a more systematic mean-rank gene set enrichment test (see Supplemental Experimental Procedures) to further understand the relationship between FIREs and gene expression patterns. For example, in cortex tissue, there is a clear difference between the mean ranks of genes neighboring brain-specific FIREs compared to random FIRE positioning (Figures 4G and 4H). Importantly, this type of analysis can be used to study the extent to which tissue-specific FIRE genes are expressed by testing all combinations of relative expression rank lists and tissue-specific FIRE gene sets (Figure 4G). In other words, if tissue-specific FIRE genes are primarily expressed in that same sample, the enrichment signal should track the diagonal of an all by all comparison (Figure 4G) and generally lower enrichment off the diagonal where the sample for the rank list and FIRE gene set are different. Indeed, we observed this trend, although the neural progenitor cell (NPC)-specific FIRE gene set is ranked higher in the cortex and hippocampus, which may be expected, given that they prominently consist of neural cells or neural progenitors. Taken together, our results suggest that tissue-specific FIREs are likely involved in tissue-specific gene expression.

### FIREs Are Conserved in Humans and Mice

If FIREs play a role in gene regulation and developmental programs, one would expect that such chromatin features would be conserved evolutionarily (Dixon et al., 2012, 2015; Vietri Rudan et al., 2015). To test this hypothesis, we compared FIREs between humans and mice in three different sample types (embryonic stem cells, neural progenitor cells, and cortex tissue) (Dixon et al., 2012, 2015; Fraser et al., 2015; Shen et al., 2012). We found that FIREs are significantly conserved in these comparisons (Figure 5A). Specifically, 33.0% of human cortex FIREs

**Figure 4. FIREs Are Enriched for Active Enhancers and Positioned Near Tissue-Type-Specific Genes**

(A) Normalized Hi-C contact matrix in left ventricle tissue showing a 2.76-Mb locus (chr2:40,000,000–42,760,000). Below are genome browser tracks for previously published (Hnisz et al., 2013) LV super-enhancers (red), LV FIRE bins (brown), and UCSC genes, including isoforms (blue). To the right is the continuous LV FIRE score along this locus.

(B) Heat maps showing the local enrichment (see Supplemental Experimental Procedures) of H3K27me3 (left), H3K4me1 (middle), and H3K27ac (right), centered on FIRE bins for each cell line or adult tissue. H3K27me3 data were not available for CO or HC.

(C) Bar plot showing the observed overlap between actual FIRE bins and previously characterized typical enhancers (blue) (Hnisz et al., 2013) for each available cell line or tissue that has both Hi-C data and typical enhancer calls. Expected values are also shown (green), which are calculated by permuting the location of FIRE bins within each tissue and calculating the overlap with typical enhancers. The y axis shows the percentage of typical enhancers overlapped by FIREs.

(D) Same as (C), except showing the percentage of super-enhancers overlapped by FIRE bins for each testable cell line or tissue.

(E) Genome browser snapshot showing an example of sample-specific gene expression near sample-specific FIREs. Shown here is a 780-kb locus (chr16:9,820,000–10,600,000) around *GRIN2A* (chr16:9,852,375–10,276,611). At the top, FIRE tracks (maroon) for each sample, showing the brain-specific FIRE (chr16:10,040,000–10,080,000, highlighted in yellow) ~197 kb away from *GRIN2A* TSS. Below, RNA-seq data (Roadmap Epigenomics Consortium et al., 2015) for all samples except OV (blue), showing *GRIN2A* is mainly expressed in brain tissues.

(F) Bar plot indicating the relative gene expression (see Supplemental Experimental Procedures) of *GRIN2A* across 20 samples.

(G) All-by-all mean-rank enrichment analysis result showing gene expression specificity of genes within 200 kb of sample-specific FIRE bins (see Supplemental Experimental Procedures). Each row is a different sample type for which the sample-specific FIRE gene set is collected, and columns are the sample type used to calculate the relative expression rank of each gene. IMR90/MSC, M1/MES, and brain tissues were previously shown to have highly overlapped FIRE bins (Figure 3A) and are therefore grouped. The color for each row of the heat map indicates the enrichment. Outlined in thick black boxes along the diagonal are the matrix entries for which the sample for the sample-specific FIRE gene set and expression rank list are the same. Highlighted in a thin yellow box is the analysis portrayed in (H).

(H) Line plot illustrating a single mean-rank enrichment analysis. The plot shows the relative gene expression values (y axis) in the cortex as a function of their numeric ranking (x axis) in the cortex. Vertical dashed lines show the position of the observed mean rank of cortex-specific FIRE genes (red dash), and the expected mean rank based on size-matched randomly selected non-FIRE bins in the cortex (gray dash). The inset is the calculation of the enrichment score.
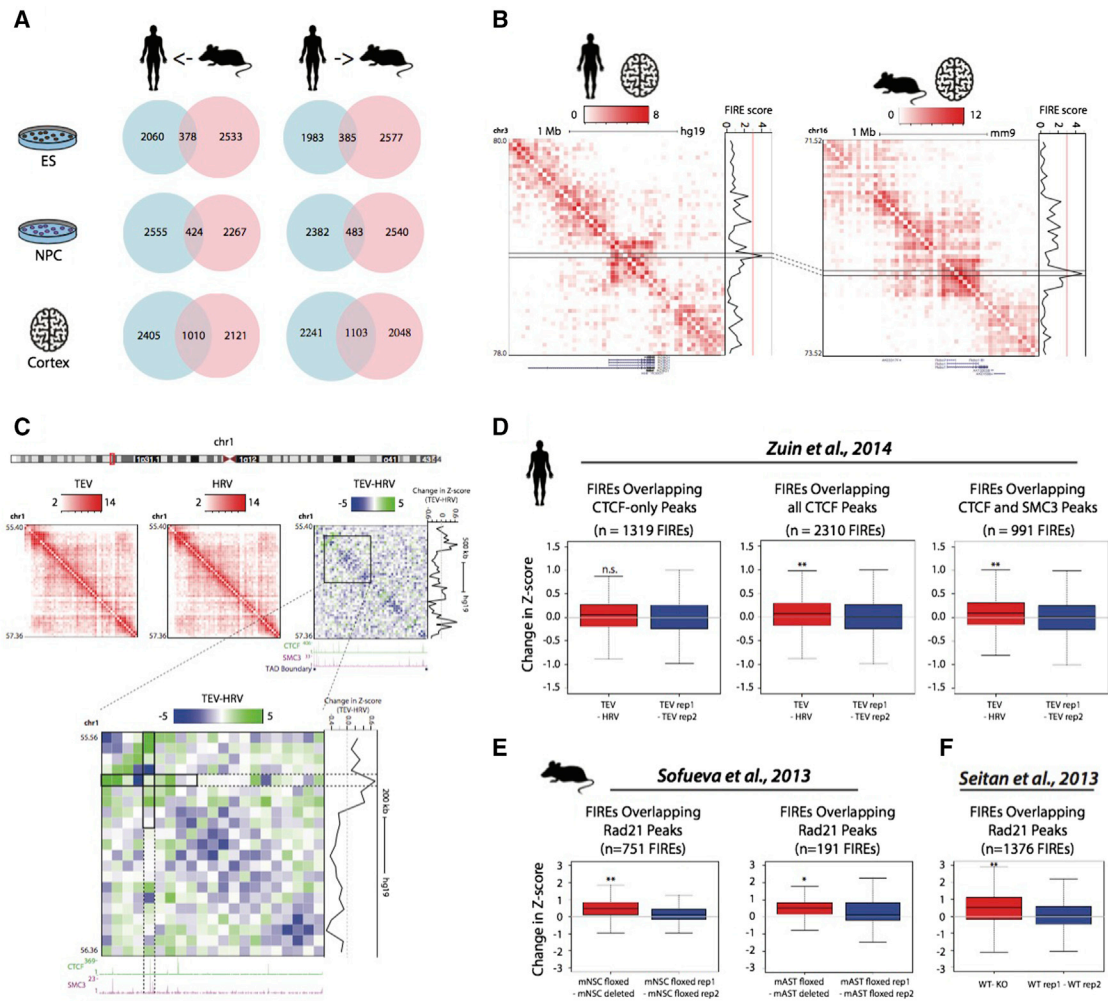
**Figure 5. FIREs Are Conserved across Evolution and Mediated by Cohesin**

(A) Venn diagrams showing the significant number of conserved FIRE bins when lifting over mouse FIREs onto the human genome (left column) or lifting over human FIREs onto the mouse genome (right column) in either embryonic stem cells (top row, p value < 5.0e−16), neural progenitor cells (middle row, p value < 2.2e−16), and cortex tissue (bottom row, p value < 2.2e−16). Significance evaluated using a Fisher's exact test (see Supplemental Experimental Procedures).

(B) Normalized Hi-C contact matrix in human cortex (left) and mouse cortex (right) for a 2-Mb syntenic region (human chr3:78,000,000–80,000,000; mouse chr16:71,520,000–73,520,000) showing a conserved FIRE (connected black lines) within the same tissue type but across species. Below is a UCSC gene track, and to the right of the contact matrix is the continuous FIRE score across the locus. For the human data, the Hi-C contact matrix, gene track, and FIRE score plot have been inverted to show synteny with the mouse data.

(C) Normalized Hi-C contact matrices (red and white) or delta matrix (green and blue) for the 1.96-Mb locus (chr1:55,400,000–57,360,000) illustrating the change of interaction frequency between TEV and HRV. Directly below the delta matrix are binding profiles of CTCF and the Cohesin subunit SMC3 in wild-type HEK cells (Zuin et al., 2014) as well as TAD boundary annotations. To the right of the Hi-C delta matrices is the continuous FIRE Z score difference between TEV and HRV. Below is a delta matrix at a zoomed-in 800-kb region (chr1:55,560,000–56,360,000) for TEV-HRV, showing the greatest reduction of FIRE score occurs at the bin with co-binding of CTCF and SMC3. The FIRE Z score difference is plotted to the right of the subtraction matrices.

(D) Box plots showing the change in Z score at FIREs overlapping bins bound by CTCF but not SMC3 "CTCF-only" (left plot), all CTCF peaks (middle plot), and CTCF and SMC3 co-binding (right plot) for the comparison of TEV and HRV. The red boxes show distributions of FIRE score change at FIRE bins called in wild-type cells minus the mutant cells, whereas the blue boxes are distributions for FIRE score change at FIRE bins called in wild-type cells but between biological replicates of wild-type cells. These comparisons show the significant reduction of FIRE score at all CTCF peaks, and especially at CTCF SMC3 co-bound peaks overlapping FIRE bins (*p = 1.0e−4; **p = 4.04e−5; two-sample t test).

(E) Similar to (D), except analysis of Z score change was done considering FIREs overlapping the Cohesin subunit Rad21 peaks using previously published Hi-C data and Rad21 ChIP-seq data in mouse neural stem cells (left plot) and mouse post-mitotic astrocytes (middle plot) (Sofueva et al., 2013). Comparison of Z score change upon deletion of Rad21 shows a significant decrease compared to changes observed between biological replicates (*p < 0.01; **p < 2.2e−16; two-sample t test).

(F) Similar to (E), except analysis of Z score change was conducted on previously published Hi-C data and Rad21 ChIP-seq data in mouse thymocytes (Seitan et al., 2013). Comparing the distributions of Z score changes at FIRE bins bound by Rad21 shows a significant reduction in Z score between the wild-type and Rad21 knockout cells compared to changes between wild-type biological replicates (**p < 2.2e−16; two-sample t test).

are also FIREs in the mouse cortex, whereas only 8.7% is expected by chance (Fisher's exact test p value < 2.2e−16). For example, returning to the *ROBO1* locus, we found that both the mouse and human cortex have only one FIRE bin in the 2-Mb region around *ROBO1*, and the single FIRE position is conserved across species (Figure 5B). Interestingly, the degree of FIRE conservation between a human and mouse is the highest in cortex tissue and less, although statistically significant, in embryonic stem cells and neural progenitor cells (ESC p value < 5.0e−16; NPC p value < 2.2e−16, Fisher's exact test) (Figure 5A). More generally, by randomly sampling syntenic bins across a range of FIRE scores, we find a modest yet significant correlation of FIRE score between a human and a mouse in each cell type (Pearson correlation coefficient = 0.20–0.42; p value < 2.2e−16) (Figures S4A–S4F). These data indicate a tendency for the local contact frequency to be conserved in syntenic regions throughout the human and mouse genome as well as conservation of the strongest locally interacting hotspots.

### CTCF and Cohesin Complex Contribute to Establishment of FIREs

We posited that FIREs might be mediated by the Cohesin complex, which has been previously shown to modulate enhancer/promoter interactions in mammalian cells (Kagey et al., 2010). To test this hypothesis, we re-analyzed three previously published Hi-C datasets, in which a Cohesin subunit was experimentally depleted in human or mouse cells (Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014), and investigated FIRE scores upon loss of a Cohesin subunit. We began by systematically examining the Hi-C datasets generated in HEK293 cells before and after depletion of the Cohesin subunit SMC3 (Figure 5C). Because the Cohesin complex is frequently bound together with CTCF throughout the genome, we focused our analysis to CTCF-only binding sites and CTCF/SMC3 co-bound peaks. SMC3-only peaks were ignored because only ∼0.7% of SMC3 peaks overlapping FIREs were not co-occupied with CTCF (Figure S4G). We then compared FIRE score changes at FIRE bins upon loss of SMC3. We observed a significant decrease of the FIRE score at CTCF/SMC3 co-bound sites (two-sample t test p value = 6.78e−6 for TEV-HRV) (Figures 5C and 5D). By contrast, there is no statistically significant FIRE score decrease at FIRE bins that had CTCF binding *without* binding of SMC3 (Figure 5D). Quantitatively similar results were seen in mouse neural stem cells, post-mitotic astrocytes, and thymocytes in the case of Rad21 deletion (two-sample t test p value = 0.0011 for post-mitotic astrocytes; two-sample t test p value < 2.2e−16 for both neural stem cells and thymocytes) (Figures 5E and 5F) (Seitan et al., 2013; Sofueva et al., 2013). Importantly, the significant decrease of the FIRE score was only observed at FIRE bins. Cohesin loss did not systemically affect FIRE scores at randomly selected and size-matched (5% of the genome) control regions (Figures S4H and S4I). We also re-analyzed Hi-C data in HEK293 cells, in which CTCF had been experimental knocked down (Zuin et al., 2014), and again observed that FIRE score is most significantly reduced at FIRE bins occupied by CTCF/SMC co-binding in wild-type cells (Figure S4J). Collectively, these results, as well as the significant enrichment of Cohesin at FIRE bins (Figure S4K), suggest that both CTCF and the Cohesin complex contribute to the formation of FIREs, and such a mechanism is likely conserved across the human and mouse.
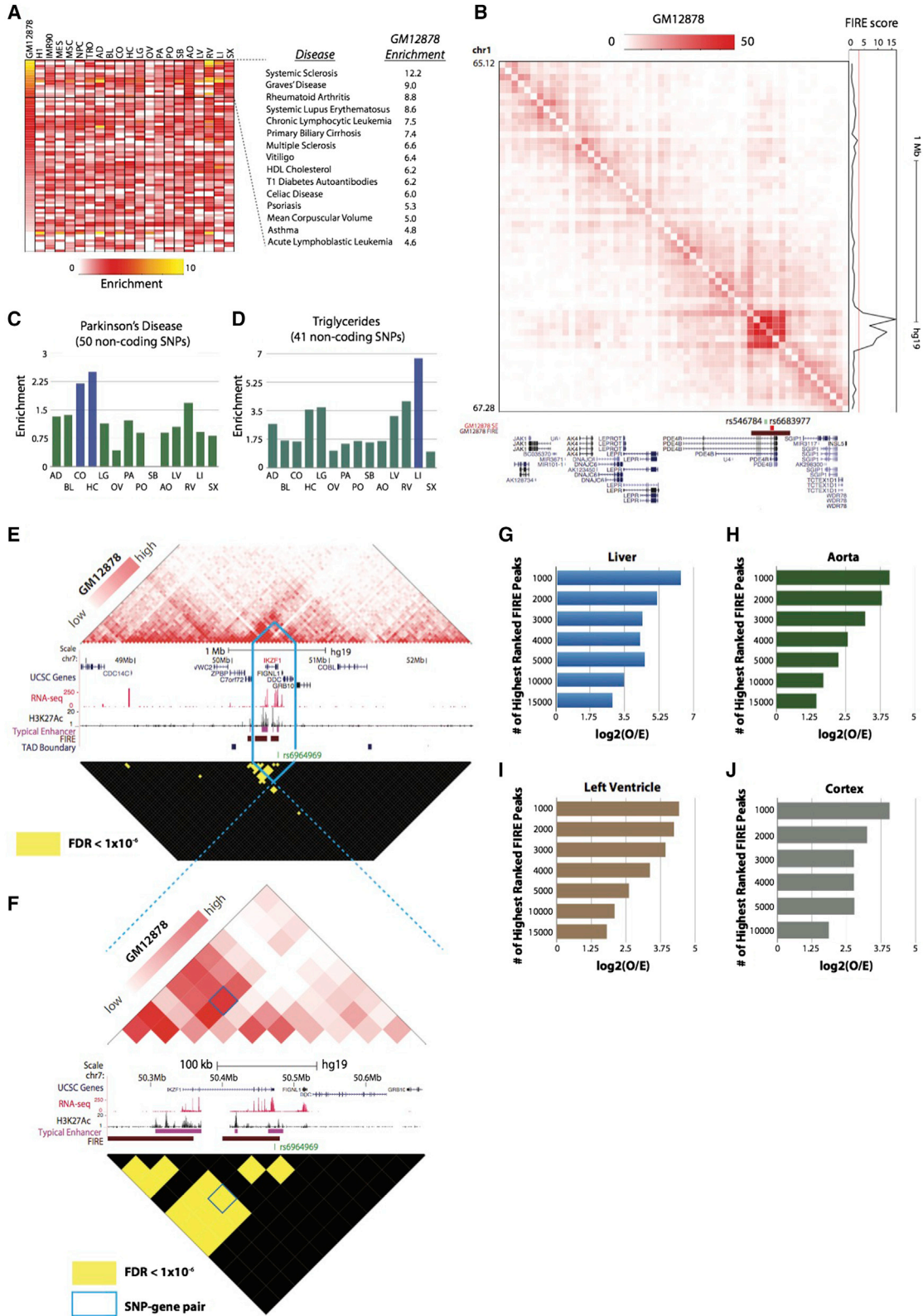
### FIREs Are Enriched for Disease-Associated SNPs

Our analyses have indicated that FIREs are enriched for active enhancers and super-enhancers (Figures 4A–4D; Figures S3B, S3C, S3F, S3G, S3N, and S3O). Because typical and super-enhancers contain a significant proportion of disease-associated SNPs (Hnisz et al., 2013), we further investigated the overlap between FIREs and disease-associated SNPs. First, we mapped 4,327 previously annotated disease-associated non-coding SNPs to FIREs defined in each cell line and tissue (see Supplemental Experimental Procedures) (Hnisz et al., 2013). Consistent with previous results (Hnisz et al., 2013), we observed 7.06 and 3.76 SNPs per megabase, and among 354 GM12878 FIREs overlapped with super-enhancers and 2,800 GM12878 FIREs overlapped with typical enhancers, respectively (Figure S5A). Surprisingly, among 1,615 GM12878 FIREs that do *not* overlap an annotated enhancer, we also observed 3.33 SNPs per megabase, which is ∼2.3-fold higher than the genome-wide SNP density (1.42 SNPs per megabase) (Figure S5A). Importantly, these SNPs would not be captured by directly overlapping super-enhancers or typical enhancers with disease-associated SNPs (Hnisz et al., 2013).

Next, we examined the overlap between disease-associated SNPs and FIREs for 456 diseases and quantitative traits (Hnisz et al., 2013). We defined the enrichment score for each disease as the ratio between the proportion of SNPs overlapped with FIREs and the proportion of FIRE bins in the genome. Strikingly, numerous immune-related diseases exhibit strong SNP enrichment in GM12878, but mild or weak enrichment in the other cell lines or tissues (Figure 6A). In fact, the vast majority of the top enrichment scores come from diseases previously implicated with immune pathology (Jostins et al., 2012) (Figure 6A). Motivated by these observations, we closely examined genes near FIREs harboring disease-associated SNPs, and found many genes associated with that type of disease. For example, two SNPs associated with acute lymphoblastic leukemia (ALL), rs6683977 and rs546784, are within a GM12878-specific super-FIRE (Figure 6B) and within *PDE4B*, a gene associated with ALL (Yang et al., 2011).

We then conducted an SNP enrichment analysis for the tissue datasets and observed similar results for some diseases and quantitative traits, with the most striking findings in the brain and liver (Figures 6C and 6D; Figures S5C and S5D). A careful examination of SNP and FIRE overlap also revealed disease candidate genes. For example, two Alzheimer's disease-associated SNPs, rs3851179 and rs536841, are within a brain FIRE (Figure S5B). Here, rs3851179 is within a brain-specific super-enhancer, whereas rs536841 is outside the super-enhancer. Interestingly, this brain-specific FIRE overlaps with *PICALM*, which contains the SNP (rs3851179) previously related to the incidence of late-onset Alzheimer's disease (Liu et al., 2016).

The presence of deleterious variants has been shown to mediate the expression of distal genes and confer pathology through DNA looping (Smemo et al., 2014). Therefore, we posited that significantly interacting bin pairs (i.e., "peaks")

A



| Disease | GM12878 Enrichment |
|---|---|
| Systemic Sclerosis | 12.2 |
| Graves' Disease | 9.0 |
| Rheumatoid Arthritis | 8.8 |
| Systemic Lupus Erythematosus | 8.6 |
| Chronic Lymphocytic Leukemia | 7.5 |
| Primary Biliary Cirrhosis | 7.4 |
| Multiple Sclerosis | 6.6 |
| Vitiligo | 6.4 |
| HDL Cholesterol | 6.2 |
| T1 Diabetes Autoantibodies | 6.2 |
| Celiac Disease | 6.0 |
| Psoriasis | 5.3 |
| Mean Corpuscular Volume | 5.0 |
| Asthma | 4.8 |
| Acute Lymphoblastic Leukemia | 4.6 |

B



C — Parkinson's Disease (50 non-coding SNPs)

D — Triglycerides (41 non-coding SNPs)

E



F



G — Liver

H — Aorta

I — Left Ventricle

J — Cortex

(legend on next page)

anchored at SNP-bearing FIREs (termed "FIRE peaks") may be enriched for SNP-gene pairs, relative to peaks anchored at non-FIRE bins (termed "non-FIRE peaks"). To explore this, we first used Fit-Hi-C (Ay et al., 2014) (see Supplemental Experimental Procedures) and a stringent statistical significance (FDR < 1e−6) cutoff to obtain the most confident peak calls within a 2-Mb genomic distance for all samples in our primary cohort (Supplemental Information). We found that this significance cut-off corresponds well to previously published total peak counts (Jin et al., 2013) and can also be used to link disease-associated SNPs to genes previously implicated in a particular disease. For example, Fit-Hi-C peak-calling analysis in GM12878 lympho-blasts reveals a highly significant (FDR = 6.29e−83) pairwise Hi-C contact between a bin containing a SNP associated with ALL (rs6964969) and a distal (~130 kb) TSS of *IKZF1*, a gene pre-viously implicated in ALL (Mullighan et al., 2009) (Figures 6E and 6F). To further explore SNP-gene-pair linkages in our tissue da-tasets, we collected statistically associated SNP-gene pairs from the GTEx eQTL database in tissues matching our Hi-C da-tasets (GTEx Consortium, 2015; Lonsdale et al., 2013). We then selected six of our higher resolution tissue Hi-C datasets that were also present in GTEx for further analysis and found that FIRE peaks were indeed significantly enriched for SNP-gene pairs compared to non-FIRE peaks (Table S4). However, this may be expected because FIREs are enriched for disease-asso-ciated SNPs, and FIREs are likely to have more local peaks than non-FIREs based on the definition of FIRE. Therefore, we analyzed the enrichment of GTEx SNP-gene pairs in subsets of the most significant FIRE peaks (i.e., the lowest FDR bin pairs). We found that the most statistically significant FIRE peaks ex-hibited the strongest enrichment of SNP-gene pairs, and relaxing the FDR for peak calling results in statistically significant, but less enriched, SNP-gene pairs (Figures 6G–6J; Table S4).

## FIREs Display Promiscuous Local Chromatin Interactions

Although FIREs are identified on the basis of their cumulative local contact frequency, this could result from FIREs either hav-ing a single local target with exquisitely high contact frequency or numerous local targets with moderate to high contact frequency. Because FIREs and super-FIREs are highly enriched for active enhancers, exploring the interaction patterns of FIRE regions may provide further insight into the interaction behavior of active *cis*-regulatory loci. First, as expected, we find that FIREs are highly enriched for local interactions compared to non-FIREs, but, unexpectedly, this contact enrichment extends in many cases to an ~500-kb genomic distance (Figure 7A). Because FIREs tend to be positioned near the TAD center, it's likely that FIREs are highly interactive with all loci within the confines of their respective TADs. Next, using the most statistically con-fident (FDR < 1e−6) Hi-C contacts determined by Fit-Hi-C, we find that FIREs have significantly more local (≤200 kb) peaks compared to non-FIREs (Figures 7B and 7C; Table S4) (two-sample t test p value < 0.01 for ovary (OV) and small bowel (SB); < 2.2e−16 for remaining samples), with an average of three to seven local peaks per FIRE bin, depending on the sample and sequencing depth (Figures 7B and 7C; Table S4). One example is the *BCL11A* locus in GM12878 lymphoblast cells, where numerous enhancer-bearing FIRE bins significantly interact with each other and with the bin containing the promoter for *BCL11A* (Figure 7D). Interesting, *BCL11A* is also known to be involved in numerous lymphoid pathologies (Satterwhite et al., 2001).

To further quantify the contacts between FIREs, we examined the contact frequencies of FIREs and non-FIRE bins across a spectrum of genomic distances within 2 Mb. We find a signifi-cantly high contact frequency between FIREs beyond 200 kb

---

**Figure 6. FIREs Are Enriched with Disease-Associated GWAS SNPs**

(A) Heat map showing the enrichment of disease-associated GWAS SNPs (see Supplemental Experimental Procedures) in FIRE bins for each cell line or tissue (columns). Rows represent the enrichment of disease-associated SNPs for one disease, and all rows in the presented heat map are sorted from high to low based on enrichment score in GM12878 (lymphoblast cell line). Only diseases with >15 SNPs are shown. Noted to the right are the top 15 diseases for which disease-associated SNPs are most enriched in GM12878 FIREs, showing the high enrichment of several diseases (all except mean corpuscular volume) with previously noted immune-mediated pathology (Jostins et al., 2012).

(B) Normalized Hi-C contact matrix of a 2.16-Mb locus (chr1:65,120,000–67,280,000) in GM12878 cells. The tracks below depict the presence of two SNPs associated with acute lymphoblastic leukemia (rs546784 and rs6683977) located within a FIRE bin (brown, chr1:66,760,000–66,800,000), ~30 kb outside of a GM12878-specific super-enhancer (red) and also within the *PDE4B* gene sequence. To the right of the Hi-C contact matrix is the FIRE score.

(C) Bar plots showing the enrichment of Parkinson's disease-associated SNPs across 14 primary adult tissue FIRE annotations, also highlighting the highest enrichment in FIREs from both brain tissues (CO and HC).

(D) Bar plots showing the enrichment of SNPs associated with the quantitative triglycerides trait across 14 primary adult tissue FIRE annotations, also highlighting the highest enrichment in liver FIREs.

(E) Normalized Hi-C contact matrix (top) in GM12878 for a 4.04-Mb locus (chr7:48,440,000–52,480,000) centered on *IKZF1* (red text). The Hi-C color scale ranges from the 15th to 99th percentile normalized contact frequencies within this locus. The reflected matrix shows the statistically significant (FDR < 1e−6) bin-pairs within 2-Mb genomic distance across the locus. Only bin pairs with FDR < 1e−6 are yellow; the rest are black. Between the matrices are a UCSC gene annotations (blue, top), RNA-seq data (red), H3K27Ac data (black), typical enhancer annotations (Hnisz et al., 2013) (purple), FIRE annotations (brown), TAD boundary calls (blue), and an SNP that is statistically linked to the IKZF1 TSS (green). The blue lines outline the 440-kb locus (chr7:50,240,000–50,680,000) that is shown in (F).

(F) Same as (E), except a zoomed-in snapshot of a 440-kb locus (chr7:50,240,000–50,680,000) centered on a SNP-bearing FIRE bin (chr7:50,440,000–50,480,000) containing the 3′ UTR of *IKZF1* and the SNP rs6964969. The blue box outlines the bin pair that is the significant interaction between previously known SNP-gene pairs.

(G) Bar plots showing the enrichment of liver GTEx eQTLs in FIRE peak bin pairs as a function of the subset of top liver FIRE peaks (based on the lowest false discovery rate) determined by Fit-Hi-C.

(H) Same as (G), except using aorta GTEx eQTLs, FIREs, and FIRE peaks.

(I) Same as (G), except using left ventricle GTEx eQTLs, FIREs, and FIRE peaks.

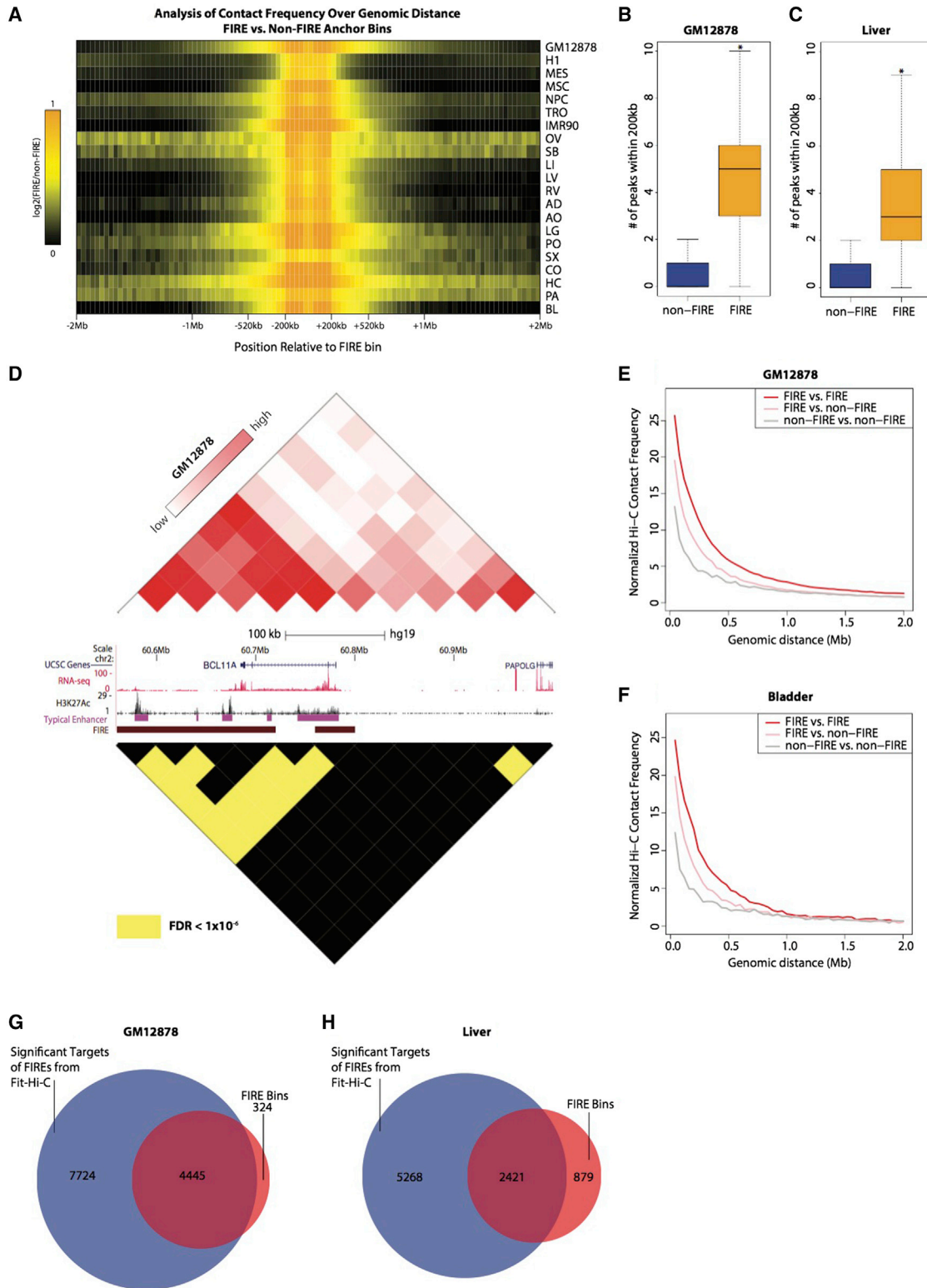(J) Same as (G), except using cortex GTEx eQTLs, FIREs, and FIRE peaks.

**Figure 7. FIREs Have Several Targets and Are Self-Interactive**

(A) Heat map showing the relationship between the mean observed contact frequencies at FIREs compared to the mean observed contact frequency at non-FIREs. Enrichment is shown as the ratio between the two contact observed mean contact frequencies (FIRE:non-FIRE) per unit genomic distance, from ± 40 kb to ± 2 Mb, centered on FIRE bins. Each row represents the analysis of a different sample, and the color intensity corresponds to the enrichment value.

*(legend continued on next page)*

(Figures 7E and 7F), often up to ∼500 kb and even up to 2 Mb in some cell lines and tissues (Figure S5E; Table S4). Furthermore, we find a significant proportion of FIREs are targets of other FIREs (chi-square test p value < 1e−5 for OV and < 2.2e−16 for the rest of the samples) (Figures 6, 7E, 7G, 7H, and S5E; Table S4). Taken together, these data support the notion that FIREs represent spatially active regions in the genome.

## DISCUSSION

3C and related technologies have been instrumental for understanding the hierarchical organization of mammalian genomes. Comparative analyses across cell types or species have thus far revealed a number of organizational features, including dynamic chromosomal compartments (Dixon et al., 2015; Lieberman-Aiden et al., 2009), TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), sub-TADs (Phillips-Cremins et al., 2013), insulated neighborhoods (Dowen et al., 2014), and chromatin loops (Rao et al., 2014). Here, through a comprehensive survey of chromatin organization in 21 human tissues and cell types, we report the finding of a previously under-appreciated feature of chromatin organization, FIRE, defined as regions that show substantial levels of local chromatin interactions. FIREs are distinct structural features compared to the previously described 3D genome features, such as TADs, chromatin loops, and compartments. FIREs are enriched in compartment A and display strong tissue-type specificity, with nearly 60% of the FIREs found in two or fewer tissues and cell types out of 21 surveyed. Perhaps most surprisingly, FIREs appear to engage in promiscuous chromatin interactions within their local chromatin neighborhood. The majority of the FIREs identified interact with multiple partners, while the reported chromatin loops typically connect two genomic regions together. Thus, FIREs are hotspots of local chromatin interactions. Finally, FIREs likely represent genomic regions actively engaged in gene regulation. Indeed, they reside near cell-identity genes, harbor significant levels of active chromatin marks, and are enriched for active enhancers, especially super-enhancers.

Further analysis reveals FIREs are closely related to previously reported super-enhancers (Hnisz et al., 2013). In GM12878 cells, in which deeply sequenced Hi-C data were available, nearly 100% of the super-enhancers are FIREs. Such an observation sheds light on the spatial architecture of super-enhancers and other active enhancers. Specifically, our results suggest that in addition to the high density of transcription factor binding and active chromatin modification, these long-range control elements also share a unique spatial feature: a high level of local chromatin interactions. Three additional properties about FIREs carry implications for the understanding of chromatin organization of enhancers. First, FIREs are not only highly interactive within 200 kb, but also highly interactive beyond 200 kb. Because FIREs are often positioned toward the TAD center, this likely means these FIREs are free to explore and interact with a substantial fraction of the TAD structure. Second, we find that FIREs often have numerous significant local interaction partners. Coupled with the observation that FIREs and superFIREs are highly enriched for enhancers, this uncovers the promiscuously interactive behavior of active enhancer sequences. This could mean that enhancers are likely to explore and physically engage with several loci in their local neighborhood in search for compatible targets. Lastly, we find that FIREs are highly self-interactive, even beyond the local (±200 kb) neighborhood. This underscores the significant degree of active cis-regulatory element spatial clustering occurring within the topological framework of larger domains. These observations, in conjunction with the notion that FIREs exhibit a high degree of tissue-specificity, reveal the degree to which tissues contain unique chromatin folding signatures at their active cis-regulatory elements. Through their heightened local contact frequency, FIREs are likely to engage with several cis-regulatory elements in their TADs and cooperatively regulate gene expression.

By analyzing the effects of Cohesin depletion in three independent studies involving both mouse and human cells, we found that the Cohesin complex is a key mediator of FIREs, and this mechanism is conserved across species. Previous analyses of chromatin architecture in mammalian cells indicated that loss of Cohesin results in a reduction of interaction frequency within TADs ("intra-TAD"), whereas knockdown of CTCF results in both loss of intra-TAD contact frequency and an increase in inter-TAD contact frequency (Zuin et al., 2014). Our re-analysis of these data in the context of very local chromatin interaction frequency indicates that upon loss of Cohesin or CTCF, the most dramatic reduction in FIRE score at FIRE bins was observed at loci containing CTCF/Cohesin co-bound peaks but not CTCF-only sites. We further demonstrate the Cohesin dependence of FIREs in murine neural progenitor cells, astrocytes, and thymocytes, supporting a conserved mechanism of FIRE establishment.

In sum, by generating a rich resource of chromatin contact maps across 21 human tissues and cell types and exploring

(B) Box plot for GM12878 showing the distributions of a number of statistically significant (FDR < 1e−6) Hi-C contacts within 200 kb emanating from non-FIRE (blue box) or FIRE (yellow box) bins (two-sample t test p value < 2.2e−16).
(C) Same as (B), except analysis of liver data.
(D) Comparison of the normalized contact matrix (top triangle) to statistically confident (FDR < 1e−6) pairwise contacts (bottom triangle) in GM12878 across a 440-kb locus centered on BLC11A. Between the matrices are the UCSC gene annotations (blue), RNA-seq (red), H3K27Ac (black), typical enhancer annotations (purple) (Hnisz et al., 2013), and FIRE annotations (brown). Color bar values of the Hi-C contact matrix correspond to the 15th and 99th percentiles, respectively, across this locus. In the lower triangle matrix, only the most confident bin pairs (FDR < 1e−6) are colored yellow.
(E) Line plots in GM12878 showing the normalized Hi-C contact frequency (y axis) as a function of genomic distance (x axis) for three categories of pairwise interactions: FIRE-FIRE interactions (red line), FIRE-non-FIRE interactions (pink line), and non-FIRE-non-FIRE interactions (gray line).
(F) Same as (E), except analysis is in bladder tissue.
(G) Venn diagram showing the overlap between all annotated FIRE bins (red circle) in GM12878 and all bins that are involved in statistically significant (FDR < 1e−6) pairwise contacts (blue circle).
(H) Same as (G), except analysis is in liver tissue.

with integrative analytic methods, we have cataloged 3D genome interactions at various hierarchical levels and uncovered the highly dynamic nature of local interaction hotspots. These results provide insights into the chromatin organization in mammalian cells.

## EXPERIMENTAL PROCEDURES

### Hi-C

Hi-C experiments on all human tissues were performed as previously described using the HindIII restriction enzyme (Lieberman-Aiden et al., 2009), with minor modifications pertaining to handling flash frozen primary tissues (Leung et al., 2015). All previously published Hi-C datasets analyzed in this study were generated using the original "dilution" Hi-C protocol (Lieberman-Aiden et al., 2009) and HindIII, unless otherwise noted (Table S1).

### Hi-C Data Processing

Newly generated Hi-C datasets were sequenced on either the Illumina HiSeq2000 or HiSeq2500 instrument. Published datasets were obtained from the SRA and converted to fastq files. Data were then processed using a custom pipeline, beginning with aligning each read end to the mm9 or hg19 reference genomes using BWA –mem. Chimeric read ends were filtered to keep only 5′ alignments with MAPQ > 10, and then read ends were paired and de-duplicated. Raw contact matrices were constructed using in-house scripts, and then further processed using HiCNormCis (described below) or using HiCNorm (Hu et al., 2012), Vanilla Coverage (Rao et al., 2014), or ICE (Imakaev et al., 2012), where indicated.

### Compartment A/B Identification

Compartment A/B analysis was performed at 1-Mb resolution, as previously described (Lieberman-Aiden et al., 2009), using the "prcomp" function in R on the Pearson correlation matrix.

### Identification of Topological Domains

Topological domain boundaries were identified at 40-kb bin resolution using the previously described insulation score analysis approach, with two minor modifications (Crane et al., 2015). Because mammalian TAD have been previously identified to be ∼1 Mb, a 1-Mb genomic region was used rather than 500 kb. Additionally, a 200-kb window, rather than 100 kb, was used for calculation of the delta vector.

### Identifying Frequently Interacting Regions

We developed a Poisson-regression-based normalization approach, named "HiCNormCis," to identify FIRE bins. Specifically, we first partitioned the entire genome into bins, and calculated the total number of intra-chromosomal ("cis") interactions in the contact distance range of 15–200 kb for each bin. Bins with low mappability (<0.9) around HindIII cut sites were removed. HiCNormCis then takes into account biases from three known factors known to bias observed Hi-C contact counts, including effective fragment length, GC content, and mappability (Yaffe and Tanay, 2011) (related to Figures 2 and S2). Let $Y_i$ represent the total cis interactions (15–200 kb) for the $i$th bin. Additionally, let $F_i$, $GC_i$, and $M_i$ represent the effective fragment length and GC represent content and mappability in the $i$th bin, respectively. The detailed calculation of $F_i$, $GC_i$, and $M_i$ is described in our previous work (Hu et al., 2012). Assume $Y_i$ follows a Poisson distribution, with a mean of $\theta_i$. We fitted a Poisson regression model as follows: $\log \theta_i = \beta_0 + \beta_F F_i + \beta_{GC} GC_i + \beta_M M_i$, and defined the residual $R_i = Y_i / \exp(\widehat{\beta}_0 + \widehat{\beta}_F F_i + \widehat{\beta}_{GC} GC_i + \widehat{\beta}_M M_i)$ as the normalized total cis interaction. Noticeably, $\exp(\widehat{\beta}_0)$ is proportional to the overall sequencing depth, and the residual $R_i$ has a mean of 1. Therefore, the normalized total cis interactions are robust to different sequencing depths, and are directly comparable among different samples. Visual inspection revealed that $R_i$ follows a Gaussian distribution (related to Figure S2). Therefore, we converted $R_i$ to the corresponding $Z$ score and $-\ln(p$ value). The same approach can theoretically be applied to any Hi-C dataset generated using a restriction enzyme and at any bin size.

### Identification of Significant Hi-C Contacts

Statistically significant contacts in Hi-C data were identified at 40-kb resolution using Fit-Hi-C, as previously described (Ay et al., 2014) (see Supplemental Experimental Procedures). We used the default Fit-Hi-C code to calculate a p value and q value for each bin pair within a 2-Mb genomic distance. For all analyses in this study, we used a conservative peak-calling threshold of FDR < 1e−6.

### REFERENCES

Arens, R., Nolte, M.A., Tesselaar, K., Heemskerk, B., Reedquist, K.A., van Lier, R.A.W., and van Oers, M.H.J. (2004). Signaling through CD70 regulates B cell activation and IgG production. J. Immunol. 173, 3901–3908.

Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res. 24, 999–1011.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523, 240–244.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science 295, 1306–1311.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin domains: the unit of chromosome organization. Mol. Cell *62*, 668–680.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Dryden, N.H., Broome, L.R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I., et al. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome Res. *24*, 1854–1868.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C., Aitken, S., et al.; FANTOM Consortium (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol. Syst. Biol. *11*, 852.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B., and Liu, J.S. (2012). HiCNorm: removing biases in Hi-C data via Poisson regression. Bioinformatics *28*, 3131–3133.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I., et al. (2016). 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell *18*, 262–275.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IIBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kingwell, K. (2013). Epilepsy: GRIN2A mutations identified as key genetic drivers of epilepsy-aphasia spectrum disorders. Nat. Rev. Neurol. *9*, 541.

Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.A., Lin, S., Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature *518*, 350–354.

Leyva-Díaz, E., del Toro, D., Menal, M.J., Cambray, S., Susín, R., Tessier-Lavigne, M., Klein, R., Egea, J., and López-Bendito, G. (2014). FLRT3 is a Robo1-interacting protein that determines Netrin-1 attraction in developing axons. Curr. Biol. *24*, 494–508.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al.

(2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Liu, G., Xu, Y., Jiang, Y., Zhang, L., Feng, R., and Jiang, Q. (2016). PICALM rs3851179 variant confers susceptibility to Alzheimer's disease in Chinese population. Mol. Neurobiol, Published online April 5, 2016. http://dx.doi.org/10.1007/s12035-016-9886-2.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al.; GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Meaburn, K.J., and Misteli, T. (2007). Cell biology: chromosome territories. Nature *445*, 379–781.

Montavon, T., and Duboule, D. (2013). Chromatin organization and global regulation of Hox gene clusters. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120367.

Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., Schulman, B.A., et al. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N. Engl. J. Med. *360*, 470–480.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Ohi, K., Shimada, T., Nitta, Y., Kihara, H., Okubo, H., Uehara, T., and Kawasaki, Y. (2016). Specific gene expression patterns of 108 schizophrenia-associated loci in cortex. Schizophr. Res. *174*, 35–38.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Satterwhite, E., Sonoki, T., Willis, T.G., Harder, L., Nowak, R., Arriola, E.L., Liu, H., Price, H.P., Gesk, S., Steinemann, D., et al. (2001). The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. Blood *98*, 3413–3420.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. *23*, 2066–2077.

Selvaraj, S., R Dixon, J., Bansal, V., and Ren, B. (2013). Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nat. Biotechnol. *31*, 1111–1118.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013).

Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 3119–3129.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell *163*, 1611–1627.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. *10*, 1297–1309.

Xu, Z., Zhang, G., Jin, F., Chen, M., Furey, T.S., Sullivan, P.F., Qin, Z., Hu, M., and Li, Y. (2015). A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in Hi-C data. Bioinformatics *32*, 650–656.

Xu, Z., Zhang, G., Wu, C., Li, Y., and Hu, M. (2016). FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. Bioinformatics *32*, 2692–2695.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. *43*, 1059–1065.

Yang, J.J., Cheng, C., Devidas, M., Cao, X., Fan, Y., Campana, D., Yang, W., Neale, G., Cox, N.J., Scheet, P., et al. (2011). Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat. Genet. *43*, 237–241.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van IJcken, W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U S A *111*, 996–1001.