Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex

Paola Giusti-Rodríguez[1†], Leina Lu[2†], Yuchen Yang[3†], Cheynna A Crowley[3], Xiaoxiao Liu[2], Julien Bryois[4], Ivan Juric[5], Joshua S Martin[3], S. Colby Allred[1], NaEshia Ancalade[1], James J Crowley[1], Jerry Guintivano[1,6], Philip R Jansen[11,12], George J Jurjus[13,14], Gouri Mahajan[15], Grazyna Rajkowska[15], James C Overholser[16], Sirisha Pochareddy[17], Gabriel Santpere[17], Jeanne E Savage[7], Yurae Shin[18], Antonio F Pardiñas[19], Michael C O'Donovan[19], Michael J Owen[19], Danielle Posthuma[7], Nenad Sestan[17], Craig A Stockmeier[15], James T R Walters[19], Gregory E Crawford[21,22], Yan Li[2], Fulai Jin[2,23,24 *], Ming Hu[5 *], Yun Li[1,3 *], Patrick F Sullivan[1,4,6 *]

† *Equal contributions.* * *Co-last authors.*

Correspond with: PF Sullivan (patrick.sullivan@ki.se), Department of Medical Epidemiology and Biostatistics, Karolinska Institutet (Stockholm, Sweden) and the Departments of Genetics and Psychiatry, University of North Carolina (Chapel Hill, NC, USA).

Affiliations

1 Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
2 Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, US
3 Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
4 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SWE
5 Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA
6 Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA
7 Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, NL
8 Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, NL
9 Department of Psychiatry, Louis Stokes VA Medical Center, Cleveland, OH, USA
10 Department of Psychiatry, Case Western Reserve University, Cleveland, OH, USA
11 Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA
12 Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, USA
13 Department of Neuroscience & Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA
14 National Research Foundation of Korea, South Korea
15 MRC Centre for Neuropsychiatric Genetics & Genomics, Division of Psychological Medicine & Clinical Neurosciences, Cardiff University, Cardiff, Wales, UK
16 Center for Genomic and Computational Biology, Duke University, Durham, NC, USA
17 Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA
18 Department of Population & Quantitative Health Science, Case Western Reserve University, Cleveland, OH, US
19 Department of Engineering & Computer Science, Case Western Reserve University, Cleveland, OH, US

Abstract

Most human genetic studies assume that connections between genomic regions are adequately captured by linkage disequilibrium (LD). LD reflects historical events that shaped the individuals in a sample, and LD patterns are consistent in DNA sampled from essentially any cell at any point in development from embryonic stem cells to highly specialized post-mitotic neurons. LD-based studies have been highly useful in identifying genetic associations for idiopathic maladies like schizophrenia; however, even as LD is fundamental to these successes, LD confounds attempts to derive confident biological insights from genetic results as it imperfectly reflects the functional state of DNA in cell nuclei. Interpretation of most genetic findings for schizophrenia is complicated by the presence of many significant and highly correlated associations in non-coding regions. Here, we created a high-resolution map of three-dimensional genome architecture by applying Hi-C to adult and fetal brain cortex with concomitant RNA-seq, ATAC-seq, and ChIP-seq data (CTCF, H3K27ac, and H3K4me3). An extensive set of analyses established the quality, information content, and salience of these new Hi-C data. We used these 3D data to functionally connect genetic results for schizophrenia to specific genes. We show that LD-based approaches provide a limited view of the complexity of schizophrenia GWAS findings. Gene set analyses based on functional genomic data provide an expanded view of the biological processes involved in the etiology of schizophrenia.

Introduction

In the last decade, detailed genome-wide searches for genetic variation fundamental to human maladies of exceptional public health importance became feasible [1]. Genomic studies are particularly important for idiopathic disorders like schizophrenia that have few proven biological risk factors despite twin-family studies conclusively establishing a role for inheritance [2]. Hopes that exon variation would provide a key to schizophrenia [3,4] have not eventuated as exome sequencing studies have identified only two genes to date [5,6] (for comparison, exome sequencing studies of autism identified over 60 genes with about half the number of cases as for schizophrenia [7,8]). The lack of exonic findings for schizophrenia is unfortunate given the available tools for experimental modeling of single genes.

In contrast, genome-wide association studies (GWAS) of common genetic variation in schizophrenia have yielded more findings. A landmark GWAS in 2014 identified 108 significant loci for schizophrenia [9], and a subsequent study found 145 loci [10]. Most of the "genetic architecture" of schizophrenia lies in common variants of relatively subtle effects identifiable by GWAS [11], and other complex human diseases have analogous conclusions (e.g., type 2 diabetes mellitus) [12]. As of 08/2018, there have been 3,029 GWAS paper that identified 31,976 genome-wide significant associations for 2,520 diseases, disorders, traits, or lab measures (***URLs***), and these are almost always common variants of subtle effects (median odds ratio, OR, 1.22) and ORs infrequently exceed 2 [13].

Although GWAS findings are surprisingly informative in aggregate [9,10,14,15], delivering strong hypotheses about their connections to specific genes has proven difficult. For schizophrenia, fewer than 10% of the GWAS loci have been connected to specific genes via gene expression quantitative trait loci (eQTL), SNP prioritization algorithms, chromatin interactions, or other approaches [10,16-19]. Investigators often rely on genomic location to connect significant SNPs to genes but this is problematic because GWAS loci usually contain many correlated and highly significant SNP associations over hundreds of Kb and multiple genes expressed in tissues of interest. The lack of direct connections to genes constrains subsequent experimental modeling and efforts to develop improved therapeutics.

The three-dimensional (3D) arrangement of chromosomes enables physical and regulatory interactions between genomic regions located far apart [20]. Chromosome conformation capture (3C) methods enable identification of 3D interactions *in vivo* [21,22] and can clarify GWAS findings. For example, an intergenic region associated with

multiple cancers was shown to be an enhancer for *MYC* via a long-range chromatin loop [23,24], and intronic *FTO* variants are robustly associated with body mass but influence expression of distal genes via long-range interactions [25]. Roussos et al. [17] used 3C methods to identify an intergenic chromatin loop for *CACNA1C* (from intronic GWAS associations for schizophrenia and bipolar disorder to its promoter). Won et al. used Hi-C to assess the 3D chromatin interactome in fetal brain, and asserted connections of some schizophrenia associations to specific genes [18].

These examples suggest that knowledge of the 3D chromatin interactome in human brain could help clarify the meaning of the many GWAS findings for schizophrenia [10]. Because 3D interactome data from human brain are limited, we sequenced adult and fetal cortical samples to greater depth than in any prior study of a human tissue to enable a detailed portrait of the brain 3D chromatin interactome. After establishing the quality and informativeness of these new brain Hi-C data, we contrasted LD-based gene identification with those from high-confidence regulatory chromatin interactions. We show what LD-based approaches provide a limited view of the functional complexity of schizophrenia GWAS findings. Gene set analyses based on functional genomic data provide an expanded view of the biological processes involved in the etiology of schizophrenia.

Results

Description and general properties of chromatin organization in brain

To gain further understanding of 3D chromatin organization of the brain and to evaluate relevance to schizophrenia, we applied "easy Hi-C" (eHi-C) [26] to six postmortem samples (*N*=3 adult temporal cortex and *N*=3 fetal cortex). We generated 10.4 billion reads and, following quality control, there were 1.323 billion high-confidence *cis*-contacts to enable a kilobase resolution map of the chromatin interactome (i.e., uniquely mapped, PCR-duplicates removed, intra-chromosomal reads >20 Kb and ≤2 Mb apart). To our knowledge, these are the deepest Hi-C data on any human tissue (***Figure 1a***). We also processed all human Hi-C tissue data publicly available in 01/2018 for a total of 25 Hi-C datasets from human tissues and cell lines (***Table S1***) [18,27]. We generated 22.5X as many *cis*-contacts for adult cortex and 1.56X as many for fetal cortex compared to the next largest datasets [18,27]. We generated four Hi-C readouts. Compartments (100 kb resolution) are "A" (active) or "B" (inactive) [28]. Topologically associating domains (TADs, Mb-scale) [29,30] segment the genome into block-like regions within which chromatin interactions have a strong tendency to occur [29,30]. Frequently interacting regions (FIREs, 40 Kb) are regions with significantly greater *cis*-connectivity than expected under a null hypothesis of random collisions. [27,31] FIREs can exist in isolation or in contiguous blocks called super FIREs [32]. Chromatin interactions identify genomic regions that are physically proximal in the nuclear 3D space despite being far apart in genomic location [20,31-37]. Chromatin interactions in this study consisted of two 10 Kb anchors that were from 20 Kb to 2 Mb apart. Some chromatin interactions occur transiently, randomly, or in subsets of cells, but many others are stable structural or regulatory features of cells in a tissue. To aid interpretation of the eHi-C readouts, we generated or assembled RNA-seq, ATAC-seq, and ChIP-seq data (CTCF, H3K27ac, and H3K4me3) from adult and fetal cortex (***Table S2***) [32,38-40]. ***Figure S1*** shows a circular genome plot of all eHi-C readouts.

Before evaluating genetic findings for schizophrenia, we analyzed the properties of these brain eHi-C data. Because these analyses are foundational to our goal of furthering understanding of schizophrenia, we summarize the analyses here with fuller details in the ***Supplemental Note***. First, we compared our eHi-C readouts to external Hi-C datasets (including brain samples) for A/B compartments (***Figures S2a-b***), TAD boundaries (***Figures S3a-b***), FIREs (***Figures S4a-b***), and chromatin interactions (***Figure S5***). We found good agreement with external Hi-C data. Second, we evaluated whether these Hi-C readouts captured biologically relevant information (summary in ***Table 1***, details in ***Table S3***). We found that FIREs and super FIREs recapitulated key functions of the source tissues: fetal results pointed at differentiation and neurogenesis and adult at core neuronal functions. As a non-brain control, analysis of FIREs and super FIREs from heart ventricle

Hi-C data were consistent with basic myocardial functions. GREAT analyses of fetal chromatin interactions enriched for transcriptional regulation and core functions of the major cell types (glia, oligodendrocytes, and neurons), and the adult results pointing at postsynaptic density and excitatory synapse. TAD boundaries showed no enrichment consistent with their cell type-independent insulative and structural roles. Third, we assessed enrichment of functional genomic features (***Tables 1*** and ***S4***). As expected [27], adult FIREs were enriched in adult cortical H3K27ac histone marks, the presence of an enhancer, and open chromatin while depleted in H3K4me3 histone marks. Fetal brain FIREs were enriched for fetal H3K27ac histone marks and CTCF binding sites. Chromatin interactions were enriched for open chromatin, typical enhancer, CTCF, and H3K27ac in adult cortex along with protein-coding TSS while depleted in H3K4me3 in adult cortex. In fetal cortex, chromatin interactions were enriched for H3K27ac and CTCF with depletion of H3K4me3 and gene expression. TAD boundaries in adult and fetal brain were enriched for CTCF and TSS of protein coding genes, as expected [29]. Fourth, we evaluated the relation between TADs and LD given that TADs and LD [41] are both on the scale of 10^5-10^6 bases and segment the genome into blocks. TADs are defined via observations of 3D chromatin interactions in samples of cell nuclei whereas LD reflects historical events for a given human sample. Consistent with prior studies [18,42], we found that TADs and LD blocks identify different regions. This conclusion is consistent with observations of inbred laboratory mice that have chromosome scale LD blocks but far smaller TADs that are conserved between mouse and human [29]. Fifth, given the importance of brain FIREs, we conducted a set of evolutionary analyses. The findings were coherent: brain FIREs have stronger evidence for ancient and recent positive selection, less population differentiation, and fewer singleton/doubleton single nucleotide variants. These observations suggest that brain FIREs are important genomic regions under stronger population genetic constraints. Sixth, given the role of the chromatin interactome on transcriptional regulation, we evaluated the importance of chromatin interactome features on gene expression in fetal and adult brain and found: (a) developmentally specific adult vs fetal FIREs had a strong relation to gene expression; (b) in a multivariable model, we found that variables assessing A/B compartment, FIREs, and chromatin interaction were significant predictors of gene expression (while controlling for H3K27ac and H3K4me3 marks) with model R^2 0.0475 (***Figure S6***) – thus, chromatin interactome data provide orthogonal information compared to epigenetic data predicting dynamic gene expression; and (c) we found a strong overlap of adult cortex chromatin interactions with adult cortex eQTLs (from CommonMind Consortium [16]), particularly for chromatin interactions whose anchors coincided with eQTL SNP-gene pairs that also had H3K27ac or H3K4me3 marks (***Figure 1b***).

In conclusion, the brain eHi-C data we generated were consistent with prior Hi-C datasets. Although Hi-C data directly incorporate few genome annotations, Hi-C readouts appear to be a particularly informative functional genomic datatype. These readouts are often orthogonal to other datatypes (e.g., histone marks and LD-based location) but have a strong relation to brain gene expression. Evaluation of the relation of eHi-C readouts to genetic risk for schizophrenia is thus logical and empirically-grounded.

Hi-C readouts and the genetic architecture/SNP-heritability of schizophrenia

GWAS data can capture the common variant genetic architecture of a trait via a direct, SNP-based assessment of heritability from genome-wide data [43]. A more refined question can also be posed [44]: to what extent is a specific genome annotation enriched for SNP-heritability? This addresses the heterogeneous impact of different annotations on disease risk. We determined SNP-heritability enrichment using partitioned LD score regression [44]. As its name indicates, this technique leverages LD; we believe that it is a valid way to understand the importance of 3D Hi-C readouts given its genome-wide, gigabase focus. As we show below, LD and 3D approaches can yield different when used to dissect individual loci at a scale of 100s of Kb. The largest published study of the genetic architecture of schizophrenia is the CLOZUK GWAS that identified 145 genome-wide significance loci (40,675 cases and 64,643 controls) [10]. We included large GWAS from multiple brain-related and somatic traits, and contrasted Hi-C readouts from six datasets along with multiple other functional annotations.

The results are summarized in ***Figure 2*** (***Table S5***). The A compartment annotation comprised ~40% of the genome and showed considerable SNP-heritability enrichment that was not specific to the Hi-C tissue source or to any GWAS. FIREs (~5% of the genome) and super FIREs (~2% of the genome) showed more limited enrichments but were only moderately specific to a tissue. TAD boundaries (1.6% of the genome) showed no SNP-heritability enrichments. There were far more chromatin interactions in fetal than adult brain (52% vs 40% of the genome) but similar numbers of regulatory chromatin interactions (~10% in both fetal and adult brain), and these showed SNP-heritability enrichments that were relatively specific to brain traits (recalling that body mass has functional genomic brain connections [45]). Gene expression in liver was enriched for SNP-heritability for body mass and height, and gene expression in blood was particularly enriched for SNP-heritability for height, inflammatory bowel disease, and total cholesterol. For CLOZUK, 21 of 44 comparisons showed significant enrichment (Bonferroni correction for $P<0.001$). The functional annotations that showed the most specific connections to schizophrenia and brain traits were open chromatin in adult cortex and regulatory chromatin interactions. Thus, across the genome, annotations that capture dynamic genome processes central to gene regulation in brain have particular salience to genetic risk for schizophrenia.

Using functional genomic data to connect schizophrenia GWAS loci to genes

The CLOZUK schizophrenia GWAS identified 145 genome-wide significant loci. [10] These loci were defined using the standard PGC algorithm of LD-based "clumping" followed by merging overlapping loci. The CLOZUK paper demonstrated that these loci defined in this way were statistically independent of each other [10].

These LD-defined loci were a median of 100 Kb in size (interquartile range, IQR 40-333 Kb), and 37 (25.5%) were intergenic, 54 (37.2%) intersected a single protein-coding gene, and 54 (37.2%) intersected multiple protein-coding genes. The intergenic loci were smallest (median 110 Kb) and were a median of 181 Kb from the nearest gene (IQR 60-366 Kb). Single-gene loci were intermediate in size (median 182 Kb) and 53/54 intersected a brain-expressed gene. Multigenic loci were largest (median 501 Kb) and contained a median of 4 brain-expressed genes (IQR 2-10). Most loci were within a single adult cortex TAD but 9 loci spanned 2 or 3 TADs.

We compared three methods of assigning SNPs to genes: (a) by gene location within one of the 145 GWAS loci, (b) using adult cortex eQTL-gene data from CommonMind and GTEx ($q<0.05$ where the eQTL-SNP was within a locus) [16,46], and (c) using eHi-C chromatin interactions from adult cortex (Bonferroni $P<0.001$, where at least one anchor was within a locus and the other in a gene). To increase specificity of the eHi-C data, we selected 103K eHi-C chromatin interactions that were regulatory enhancer-promoter or promoter-promoter loops. A promoter anchor was defined as the intersection of: an eHi-C *HindIII* fragment, the start site of a transcript expressed in adult cortex (±2 Kb), and a region of open chromatin in adult cortex. An enhancer anchor was defined as the intersection of: an eHi-C *HindIII* fragment, a region of open chromatin in adult cortex, and either a H3K27ac histone mark in adult cortex or the combination of a H3K4me3 histone mark in adult cortex and a brain expressed transcript start site. We focused on adult cortex given empirical data connecting this tissue to schizophrenia using orthogonal functional genomics data (mRNA-seq, single cell RNA-seq, enhancer marks, and open chromatin) [38,40,47,48]

First, we identified 35 genes that were implicated by two different schizophrenia loci. We reasoned that that a such a "pincer" (different loci pointing at the same gene) increases the likelihood that a gene is implicated by the genomic findings [49]. The large gene *DPYD* (917 Kb) was implicated by two statistically independent loci that were 320 Kb apart (*DPYD* also has both eQTL and regulatory loop evidence). Ten genes (*CSDC2, DESI1, DNAH1, EMB, NT5DC2, PBRM1, POLR3H, SATB2, TOB2,* and *WBP2NL*) were the eQTL-gene for eQTL-SNPs located in different loci. For example, two statistically independent loci chr22:41.027-41.753 Mb (P=5.54e-13) and chr22:42.226-42.689 Mb (P=2.15e-14) are 473 Kb apart but each contains eQTL-SNPs to the same five genes. Finally, 27 genes contain an anchor for regulatory chromatin interaction with the other anchor in different significant loci. However, 17 of these genes are in chr22:39.840-40.091 Mb (P=1.76e-12), chr22:41.027-41.753 Mb (P=5.54e-13),

and chr22:42.226-42.689 Mb (*P*=2.15e-14). This suggests that these statistically independent loci may nonetheless be functionally related particularly given their location within the same adult cortex TAD (chr22:38.220-43.020 Mb).

Second, 17/37 intergenic loci could be connected to 32 different brain-expressed genes (13 genes via -SNP-gene eQTLs, 18 genes via Hi-C, and 1 gene by both). Five of the intergenic loci implicated 20 genes. The typical locational assumption that an intergenic GWAS association influenced only the nearest brain-expressed gene was supported for 7 loci.

Third, 51/54 of the single-gene loci could be connected to a brain-expressed gene. The typical assumption that the GWAS association influenced only the brain-expressed gene which it intersected was supported by eQTL or Hi-C evidence for only 11/51 loci (for 11/51 additional genes, there was the absence of evidence or no eQTL or Hi-C evidence connecting the locus to some other gene). However, this understates the complexity of single-gene loci as 29/51 loci could be connected to 2-12 genes via eQTL or Hi-C evidence. These genes were a median of 243 Kb (IQR 1.1-672 Kb) from the locus. For example, the *CACNA1C* intron 3 association (*P*=5.63e-20) supports a connection to *CACNA1C* via eQTL and Hi-C data but also has Hi-C regulatory interactions to *CACNA2D4*, *DCP1B*, *FKBP4*, *LRTM2*, and *TSPAN9* (172-672 Kb from the GWAS locus). The *DRD2* association has a similar pattern of regulatory chromatin interaction to *DRD2* as well as 6 other brain-expressed genes.

Fourth, as anticipated, the 54 multigenic loci were complex with eQTL or Hi-C evidence supporting a connection to a median of 11 genes per locus (IQR 6-23). Of these, 56.5% of the genes were outside the locus and often at a considerable distance (median 291 Kb, IQR 126-601 Kb).

In summary, of 14,890 brain-expressed protein-coding genes, 1,197 (8.1%) had evidence of a connection to a schizophrenia GWAS locus (***Figure 3***). The supporting evidence was by location (42.6%), eQTL (41.5%), and/or regulatory chromatin interactions (86.7%). Most of the genes with eQTL or chromatin interaction evidence (57.8%) were not in the locus and generally at a considerable distance outside the locus (median 305 Kb, IQR 128-620 Kb).

Gene set analyses, location versus functional genomic connections

Most gene set analyses incorporate LD [62], but functional genomic readouts (eQTL and Hi-C) are often unrelated to LD (as shown above, chromatin interactions can extend far beyond LD-based loci). Thus, we conducted a series of gene set analyses using "gene2function" hypergeometric tests in FUMA (***URLs***) [63] We contrasted two gene sets: (a) the standard approach of analyzing brain-expressed, protein-coding genes within LD-defined GWAS loci CLOZUK (339 genes); and (b) genes implicated by eQTL and/or regulatory chromatin interactions as described above (999 genes, including 314 genes implicated by location). We excluded genes in the extended MHC region (due to its complexity and gene density), and selected gene sets with ≥10 gene overlap and adjusted *P* values < 0.001.

The results are ***Tables 3*** and ***S6***. The functional genomic approach yielded 24 significant GO gene sets whereas location-based gene identification yielded only 12 gene sets. Six gene sets were in common between the two methods (e.g., cell and neuron projection). However, the functional genomic approach yielded a rather different portrait. Specifically, it implicated a variety of regulatory processes, particularly chromosome organization, synaptic transmission, and synaptic plasticity.

Rare genetic variation

Exon variation. We evaluated the salience of Hi-C readouts for genes implicated in rare variant studies of intellectual disability (i.e., the lower tail of the cognitive ability distribution). We compiled a gene set from literature reviews, OMIM, and exome sequencing studies [50-52]. We could not analyze schizophrenia because too few genes have been identified. Full details of this analysis are in ***Table S7***. There were significant univariate

associations between the implication of a gene in intellectual disability via exonic variation and its intersection with FIREs, TAD boundaries, or the number of chromatin interactions in both adult and fetal eHi-C datasets. In a multivariable model for the adult eHi-C data, intersection with a TAD boundary increased the odds of being an intellectual disability gene by 54% (OR=1.54, *P*=2.56e-4) as did the number of chromatin interactions (OR=1.26 for each doubling of the number of interactions, *P*=6.61e-11). Intersection with an adult FIRE was not significant in this model. In a multivariable model for the fetal eHi-C data, the odds of being an intellectual disability gene increased markedly with the number of chromatin interactions (OR=4.11 for each doubling of the number of interactions, *P*=8.0e-13) but fetal FIRE and TAD boundary intersections were not significant in this model.

Copy number variation (CNVs). Cases with schizophrenia have more rare CNVs than controls [53]. Excess CNV "burden" in cases can be attributed to genic CNVs with greater effects for deletion than duplication CNVs [54,55]. Greater CNV burden in cases persists after removing ~10 large CNVs individually associated with schizophrenia (e.g., 22q11 or 16p11) [54]. It is plausible that a CNV that disrupts a FIRE or a TAD [56] could explain some of the excess CNV burden. We evaluated this hypothesis using carefully curated CNVs in 4,719 schizophrenia cases and 5,917 controls [55,57]. We excluded large CNVs known to be associated with schizophrenia [54], and controlled for genotyping batch and ancestry. The presence of a CNV deletion intersecting one or more adult brain FIREs was significantly associated with schizophrenia (OR=1.72, P=3.27e-6) whereas CNV duplications were not associated (OR=1.10, *P*=0.17). The presence of a CNV deletion intersecting one or more fetal brain FIREs was modestly associated with schizophrenia (OR=1.31, *P*=4.7e-3) as were CNV duplications (OR=1.22, *P*=7.9e-3). CNVs that intersected TADs were not notably associated with schizophrenia: adult TAD-CNV deletion OR=1.08 (P=0.60); adult-TAD CNV duplication OR=1.16 (*P*=0.047); fetal TAD-CNV deletion OR=1.07 (*P*=0.64); and fetal-TAD CNV duplication OR=1.16 (*P*=0.042).

Discussion

In many respects, our understanding of the human genome is best at the extremes. At a chromosomal to megabase scale, there is considerable knowledge of the structure of the genome and the prevalence and medical relevance of variation (i.e., large structural variants). At the base pair scale, we have increasingly detailed surveys of the nature and frequency of genetic variation from studies like TOPMed and gnomAD (***URLs***). Between these extremes, some annotations appear to be increasingly complete for a few crucial topics (e.g., gene models, variation causal for Mendelian disorders, or expression patterns in human tissues) and also via the work of ENCODE and Roadmap consortia. However, particularly for complex diseases of profound societal importance, there is an unsolved problem at intermediate scales: given the typical paucity of exonic findings, precisely how do the thousands of significant, subtle, and common associations that account for most of inherited liability act mechanistically to increase risk for disease? In effect, human genetic studies of complex disease attempt to provide a parts list, and more complete parts lists and even wiring diagrams are an important next steps.

In this paper, we used the deepest Hi-C dataset of human brain to evaluate its relevance for schizophrenia and cognitive ability. Instead of considering the genome as a 1D object, we used a 3D functional snapshot of genome organization in brain cells. In doing so, we sought to identify a more complete parts list.

Most human genetic studies leverage LD. LD is a fundamental feature of the genome with a large body of supporting statistical genetic theory and analytical methods [66,67]. Indeed, LD has been essential to genetic discovery for decades. LD was crucial in linkage mapping for Mendelian diseases in large affected pedigrees. LD is a fundamental reason why GWAS "works" at all: there are millions of polymorphic SNPs in the genome but widespread LD in humans means that genotyping a relatively small number of SNPs can capture a large fraction of the information contained in common variation. LD-based imputation, LD score regression, and gene set analyses are essential parts of the GWAS toolkit [43,62,68]. However, LD is a double-edged sword: following identification of a significant locus for a complex disease, LD almost always confounds attempts to identify

specific genes. There usually are many loci with approximately the same significance values and, even with dense additional genotyping, rarely can a single variant with markedly greater significance be identified. 1D epigenetic information certainly can help in prioritization [69] although many epigenetic marks are common (e.g., brain open chromatin regions are about as large as the exome) [38]. LD remains a problem with large number of significant SNPs in high LD that overlap promoters, open chromatin, and certain histone marks.

We suggest that difficulties of LD-based approaches following locus identification in complex disease warrants incorporation of more informative data types. LD arises from historical population genetic processes whereas the chromatin interactome captures the functional organization of cells in a disease-relevant tissue. These two sets of features do not overlap well. Our results provide support for the idea that, following genetic identification of a locus for a complex disease like schizophrenia, it is essential to incorporate knowledge of the 3D interactome in the appropriate tissue. Connecting GWAS findings to genes using 1D-based methods like LD is often misleading given that the genes implicated by 3D interactome are not in LD or not the nearest gene.

Figure legends

Figure 1 shows comparisons of high-level Hi-C metrics and features. Figure 1a: Metrics for 25 Hi-C datasets. The X-axis is the total number of reads passing quality control (uniquely mapped, PCR duplicates removed), and the Y-axis is the number of informative *cis*-reads (uniquely mapped, PCR duplicates removed, intra-chromosomal, >15kb apart). Point sizes are proportional to the numbers of informative *cis*-reads. Red diamonds show data we generated in human brain using eHi-C: filled red diamond is adult temporal cortex (Adult, N=3), and open red diamond is fetal cortex (Fetal, N=3). Yellow open circles show fetal germinal zone (GZ, N=3) and cortical plate (CP, N=3) [18]. From Schmitt et al. [27], filled grey circles show 14 human tissues and 7 cell lines. Tissues: AD=adrenal gland; AO=aorta; BL=bladder; DLPFC=brain dorsolateral prefrontal cortex; Hippo=brain hippocampus; LG=lung; LI=liver; LV=heart left ventricle; OV=ovary; PA=pancreas; PO=psoas skeletal muscle; RV=heart right ventricle; SB=small intestine; SX=spleen. Cell lines: GM12878=lymphoblast; H1=human embryonic stem cell (hESC); IMR90=lung fibroblast; MES=mesoderm; MSC=mesenchymal stem cell; NPC=neural progenitor cell; TRO=trophoblast-like cell. Figure 1b: Brain eQTLs and chromatin interactions. Scatter plot of genomic distance between chromatin interaction anchors (X-axis) and mean Hi-C contact frequency (Y-axis). Using the CommonMind DLPFC eQTL dataset, we stratified by these data when the chromatin interaction anchors did not overlap an eQTL (red), when they did overlap an eQTL (blue). The subset of the chromatin interaction-eQTL overlapping which have H3K27ac or H3H4me3 mark are in green.

Figure 2 depicts a heat map for results of partitioned LD score regression for seven large GWAS for functional genomic readouts (Hi-C, ChIP-seq, gene expression, and open chromatin) for multiple tissues. See Figure 1 for tissue identifiers.

Figure 3 shows a Venn diagram implicating brain-expressed genes in schizophrenia. A gene can be implicated by location in an schizophrenia LD-based GWAS locus (Location), by adult cortex eQTL data, or via adult cortex Hi-C regulatory chromatin interactions. The insert shows percentages in each group. Hi-C dominates.

Supplemental figure legends

Figure S1. Circos plot summarizing the adult and fetal brain Hi-C readouts in the autosomal genome. From the outside inward, the tracks are: (a) ideogram (chr1ptel to 22qtel in clockwise direction); (b) gene density per Mb; 1 Mb A (green) and B (yellow) compartment in (c) fetal brain and (d) adult brain; locations of super FIREs in (e) fetal brain (red) and (f) adult brain (green); TAD boundary locations in (g) fetal brain and (h) adult brain; and chromatin interaction density per Mb in (i) fetal brain and (j) adult brain.

Figure S2. Figure S2a: Results of PCA on Hi-C compartment scores (100 Kb bins) from 25 Hi-C datasets. X-axis is PC1 (75% of variance) and Y-axis is PC2 (8.0% of variance). The fetal samples clustered together as did the adult brain samples. Point sizes are proportional to the numbers of informative *cis*-reads. Filled red diamond is our adult anterior temporal cortex dataset (Adult) and the open red diamond is our fetal cortex dataset (Fetal). Yellow open circles are fetal brain germinal zone (GZ) and cortical plate datasets (CP). Filled grey circles show 14 human tissues and 7 cell lines (see Figure 1a legend for abbreviations). Figure S2b: Clustered heat map of Jaccard index values describing the degree of overlap of compartments across Hi-C datasets.

Figure S3. Figure S3a: Results of PCA on Hi-C TAD boundary scores (40 Kb bins) from 25 Hi-C datasets. X-axis is PC1 (77% of variance) and Y-axis is PC2 (5.0% of variance). All brain samples clustered on the dominant PC1. Point sizes are proportional to the numbers of informative *cis*-reads. Filled red diamond is our adult anterior temporal cortex dataset (Adult) and the open red diamond is our fetal cortex dataset (Fetal). Yellow open circles are fetal brain germinal zone (GZ) and cortical plate datasets (CP). Filled grey circles show 14 human tissues and 7 cell lines (see Figure 1a legend for abbreviations). Figure S3b: Clustered heat map of Jaccard index values describing the degree of overlap of TAD boundaries across Hi-C datasets.

Figure S4. Figure S4a: Results of principal components analysis (PCA) on Hi-C FIRE scores (40 Kb bins) from 25 Hi-C datasets. X-axis is PC1 (69% of variance) and Y-axis PC2 which captured far less variance (5.4%). All brain samples had high scores on PC1. Point sizes are proportional to the numbers of informative *cis*-reads. Filled red diamond is our adult anterior temporal cortex dataset (Adult) and the open red diamond is our fetal cortex dataset (Fetal). Yellow open circles are fetal brain germinal zone (GZ) and cortical plate datasets (CP). Filled grey circles show 14 human tissues and 7 cell lines (see Figure 1a legend for abbreviations). Figure S4b: Clustered heat map of Jaccard index values describing the degree of overlap of FIREs across Hi-C datasets.

Figure S5. Venn diagram of chromatin interactions in our Adult and Fetal eHi-C datasets and fetal cortical plate (CP) and germinal zone (GZ). Chromatin interactions are between 10 Kb bins that were ≥20 Kb apart and ≤2 Mb apart (i.e., *in cis*). We evaluated these four brain Hi-C datasets because of relatively high read depths (***Figure 1a***). All chromatin interactions exceeded a stringent *P*-value (Bonferroni correction α=0.001). We identified 2,195,401 chromatin interactions in any of these four brain Hi-C datasets. Fetal data had greatest number of chromatin interactions (1.25 million), somewhat more than CP and GZ (0.998 million each), and over twice as many as adult cortex (0.509 million).

Online Methods

General & data availability

All procedures on data from human research subjects were approved by the appropriate ethical committees. All genomic coordinates are given in NCBI Build 37/UCSC hg19. Upon acceptance of this paper, full eHi-C readouts will be posted on the PGC website, the psychENCODE portal, and GEO (***URLs***).

Samples

Anterior temporal cortex was dissected from postmortem samples from three adults of European ancestry with no known psychiatric or neurological disorder (Dr Craig Stockmeier, University of Mississippi Medical Center). Cortical samples from three fetal brains were obtained from the NIH NeuroBiobank (gestational age 17-19 weeks), and none were known to have anatomical or genomic disease. Samples were dry homogenized to a fine powder using a liquid nitrogen-cooled mortar and pestle. All samples were free from large structural variants (>100 Kb) detectable using Illumina OmniExpress arrays. Genotypic sex matched phenotypic sex for all samples.

Easy Hi-C (eHi-C) methods

We used eHi-C to assess chromatin interactome [26]. The eHi-C protocol is biotin-free and uses sequential enzymatic reactions to maximize the recovery of DNA products from proximity ligation. The main advantage of this Hi-C adaptation is that it can generate Hi-C libraries that are comparable to traditional Hi-C but with lower sample input (as little as 10 mg brain tissue) and increased yield. All of these features of eHi-C are crucial for relatively uncommon human postmortem brain samples.

We followed the protocol described in Lu et al. [26]. Pulverized tissue (~100 mg) was crosslinked with formaldehyde (1% final concentration) and the reaction was quenched using glycine (150 mM). We lysed samples on ice with brain tissue-specific lysis buffer (10 mM HEPES; pH 7.5, 10 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, 0.5% Nonidet-40 and protease inhibitor cocktail), Dounce homogenized, and digested using the six base pair restriction enzyme *HindIII*. This was followed by *in situ* ligation. Samples were reverse cross-linked with proteinase K and purified using phenol-chloroform. DNA was then digested with four base pair restriction enzyme *DpnII* followed by size selection using PCRClean DX beads (Aline Biosciences) (choosing fragments between 100-1000 bp). The DNA products were self-ligated overnight at 16°C using T4 DNA ligase. Self-ligated DNA was purified with phenol-chloroform, digested with lambda exonuclease, and purified using PCRClean DX beads. For re-linearization of circular DNA, bead-bound DNA was eluted and digested with *HindIII* and purified using PCRClean. Bead-bound DNA was eluted in 50 µl nuclease-free water. Re-linearized DNA (~50 ng) was used for library generation (Illumina TruSeq protocol). DNA was end-repaired using End-it kit (Epicentre), "A-tailed" with Klenow fragment (3'–5' exo–; NEB), and purified with PCRClean DX beads. The 4 µl DNA product was mixed with 5 µl of 2X quick ligase buffer, 1 µl of 1:10 diluted annealed adapter and 0.5 µl of Quick DNA T4 ligase (NEB). Ligation was done by incubating at room temperature for 15 minutes. DNA was purified using DX beads, and eluted using 14 µl nuclease-free water. To sequence eHi-C libraries, we used custom TruSeq adapters in which the index is replaced by 6 base random sequences. Libraries were then PCR-amplified and deeply sequenced (2-5 independent libraries/sample) using Illumina HiSeq4000 (50 bp paired-end).

Because nearly all mappable reads start with the *HindIII* sequence AGCTT, we trimmed the first 5 bases from every read and added the 6-base sequence AAGCTT to the 5' of all reads. These reads were aligned to the human reference genome (hg19) using Bowtie [70]. After mapping, we kept reads where both ends were exactly at *HindIII* cutting sites. PCR duplicates with the same positions and UMI were removed [26]. We also removed read pairs with the two ends within the same *HindIII* fragment. To further filter valid ligation products from *cis- HindIII* pairs, we split reads into three classes based on strand orientation: "same-strand" had both ends on the same strand; "inward" had the upstream end on forward strand; and "outward" where the upstream end was on the reverse strand [26]. "Outward" read pairs with gap distance <1 Kb between the two corresponding fragments were removed because they might originate from undigested *HindIII* sites. "Inward" read pairs with gap distance <25 Kb between the two corresponding fragments were removed because they might come from self-circled DNA with undigested *HindIII* sites. All other reads are valid ligation products and were processed as described below (FIREs, chromatin interactions, TADs, and compartment A/B designation).

Hi-C readouts

We adapted in-house pipelines to process eHi-C and conventional Hi-C data from external datasets as described previously [27,71] with slight modifications. We used *bwa mem* to map each read to the hg19 reference genome, retaining only uniquely mapped reads. For chimeric reads overlapping *HindIII* cut sites, we only used the 5' end. We kept reads within 500 bp of the nearest *HindIII* cut site, and removed any intra-chromosomal read pairs within 15 Kb [71]. Processed data from our eHi-C and external Hi-C were binned into 100 Kb, 40 Kb, and 10 Kb resolution contact matrices for downstream analysis. As shown in ***Figure S1a***, we evaluated Hi-C read summary statistics including: total number of uniquely mapped reads per sample (PCR duplicates removed), total number of intra-chromosomal reads, and total number of informative intra-chromosomal reads which are >15 Kb. For

comparison, we include in ***Table S1*** all Hi-C human tissue data available as of 1/2018. Hi-C comparison data included Schmitt et al. [27] (14 adult tissues and 7 cell lines) and Won et al. [18] (3 paired fetal samples from brain germinal zone and cortical plate). Data quality from this study was comparable to prior studies although our read depth is the best of any currently available Hi-C dataset from human brain tissue.

We evaluated the reproducibility of the 40 Kb bin resolution eHi-C contact matrices. The biological replicates of each sample showed high reproducibility (Pearson correlation coefficients > 0.96) enabling pooling of all biological replicates for downstream analyses. eHi-C analyses are based on N=3 adult and N=3 fetal cortex samples for chr1-chr22 and N=2 male adult and N=2 male fetal samples for chrX (given that chromatin interactions are distinctive in females).

Frequently interacting regions (FIREs)

Following our prior study [27], we applied an in-house pipeline to identify FIRE bins. Using 40 Kb resolution Hi-C contact matrix for each chromosome. For each 40 Kb bin, we calculated the total number of *cis* (intra-chromosomal 15-200 Kb) interactions. We removed "bad" 40 Kb bins with low effective fragment size, GC content, or mappability score [72]. We also removed 40 Kb bins in the MHC region (chr6:28,477,797-33,448,354). After filtering, 64,222 40 Kb bins remained for analysis. Next, we applied HiCNormCis [27] to remove systematic biases from local genomic features, including effective fragment size, GC content and mappability [72]. The normalized total number of *cis* intra-chromosomal reads is defined as the FIRE score [27]. We then performed quantile normalization of FIRE scores across all samples using the R package "preprocessCore" to remove potential batch effect among different samples. We transformed FIRE score using $\log_2$(FIRE score+1), and transformed FIRE scores to *Z*-scores (i.e., mean=0 and standard deviation=1). We designated FIRE bins as 40 Kb regions with FIRE score one-sided *P*-value < 0.05.

Super FIREs

FIREs often cluster into contiguous runs of bins termed super FIREs. We used an in-house pipeline to call super FIREs, motivated by super enhancer calling algorithms [73,74]. The method is also described in our previous study [27]. For each Hi-C sample, we began with 40 Kb FIRE bins as described above. We then merged consecutive 40 Kb FIRE bins into contiguous FIRE regions, allowing for up to one 40 Kb bin gap. We ranked these contiguous FIRE regions by their cumulative Z-scores, and plotted the ranked FIRE regions as a function of their cumulative Z-score. Finally, we identified the inflection point of such plot, and designated all FIRE regions to the right of the inflection point as super FIREs.

Chromatin interactions

We applied a combination of Fit-Hi-C [75] with default parameters and our FastHiC [76,77] to detect long-range chromatin interactions. Starting from 10 Kb bin resolution raw Hi-C contact matrices, we removed any 10 Kb bin overlapping the ENCODE "blacklist" (uniquely mappable regions that nonetheless yield artificially high signal, ***URLs***) or the MHC region (chr6:28,477,797-33,448,354). We then ran the Fit-Hi-C+FastHiC combination caller on all 10 Kb bin pairs ≥20 Kb and ≤2 Mb apart, resulting in a total of 43,222,677 bin pairs. Specifically, we first applied Fit-Hi-C to generate informative initial values, from which we ran our FastHiC. We applied a stringent Bonferroni correction, and only considered 10 Kb bin pairs with P-value < $0.001/43{,}222{,}677 = 2.31\times10^{-11}$ as statistically significant chromatin interactions.

Since we used the six base pair restriction enzyme *HindIII* in eHi-C, it was not possible to directly apply HiCCUPS, another method to detect chromatin interactions (predominantly CTCF-mediated chromatin loops) as it was designed for *in situ* Hi-C data with the four base pair restriction enzyme *MboI*. Our approach is consistent with the long-range chromatin interaction calling algorithm adopted by Won et al. [18]. Because Fit-Hi-C and FastHiC model the global background, the resulting chromatin interactions identified (i.e., 3D peaks called) are enriched

with long-range enhancer/promoter interactions. In contrast, HiCCUPS adopts a local background, and thus the resulting peaks are enriched with CTCF mediated chromatin loops.

Topographically associated domains (TADs)

We identified TAD boundary regions using an in-house pipeline to implement the insulation score method, as described in Crane et al. [78]. Starting from 40 Kb bin resolution raw Hi-C contact matrix we applied HiCNorm [72] to obtain the normalized chromatin interaction frequency matrix. Next, each 40 Kb bin, one at a time treated as the anchor bin, obtained its "insulation score" by calculating the sum of normalized interaction frequency between the anchor bin and all bins ±1 Mb of the anchor bin. We further performed quantile normalization on the insulation score across all samples using the R package "preprocessCore". Finally, we called a 40 Kb bin a TAD boundary region if its insulation score is the minimal in its local neighboring ±1 Mb region.

A/B compartments

An output of Hi-C is determination of "A" and "B" compartments corresponding to contiguous regions of active (A) and inactive chromatin (B) which tend to self-associate. A/B compartment analysis was accomplished with an in-house pipeline following initial paper by Lieberman-Aiden et al. [28]. Hi-C data from our adult cortex, our fetal cortex, fetal germinal zone and cortical plate samples from Won et al. [18], and the 21 Hi-C datasets from Schmitt et al. [27] were all processed identically. We identified A/B compartments at both 100 Kb bin resolution. We applied quantile normalization using the R package "preprocessCore" for batch effect removal across samples. We then applied PCA to the quantile normalized matrix of compartment scores, and graphed samples in PC1 vs. PC2 plots. The PCA analysis shows clear distinctions between human brain and non-brain tissues and cell lines, and developmental stage differences for the brain samples.

Additional functional genomic data

To aid in the interpretation of our eHi-C data, we generated additional data from human brain samples. ***Table S2*** summarizes the data types and sample developmental stage/brain region. Methods are provided below.

RNA-sequencing

We generated bulk-tissue RNA-sequencing data from nine fetal cortex and nine adult DLPFC (dorsolateral prefrontal cortex) samples. The collection of dorsolateral prefrontal cortex and the psychiatric characterization are detailed in Zhu et al. [89]. All controls had no neurological disease or severe mental illness. All samples were free from large structural variants (>100 Kb) detectable using Illumina OmniExpress arrays. Genotypic sex matched phenotypic sex for all samples. We extracted total RNA from 25 mg of pulverized tissue per sample using Norgen's Fatty Tissue RNA Purification Kit (Norgen Biotek, Thorold, ON Canada). Extracted RNA from the cell pellets (each with ~3-6 million cells) using Norgen's Total RNA extraction kit. RNA concentration was measured using fluorometry (Qubit 2.0 Fluorometer), and RNA quality verified using a microfluidics platform (Bioanalyzer, Agilent Technologies). Barcoded libraries were created (Illumina TruSeq RNA Sample Preparation Kit v4) using 1 µg of total RNA as input. Samples were randomly assigned to bar codes and lanes. Libraries were quantified using fluorometry and equal amounts of all barcoded samples were pooled and sequenced using Illumina HiSeq 2000 (100 bp paired-end). We used the quasi-mapping-based mode of Salmon (version 0.8.2) [90] to generate transcript-level counts from RNAseq reads using the GENCODE gene models (v26). We used tximport to generate gene-level counts [91] and DESeq2 [92] for differential expression analysis. We used the WGCNA R package [93] for co-expression network analysis of our RNA-seq data from fetal and adult cortex.

Total-stranded RNA-seq

We also generated total-stranded RNA-seq from one control DLPFC sample and one fetal cortical sample. The purpose was to enable detection of RNA sequences that were being generated in brain but not present in GENCODE gene models.

ChIP-seq

To assist in interpreting the eHi-C data, we generated epigenetic marks using postmortem brain tissue from fetal and adult samples. The marks evaluated were H3K27ac (N=4 fetal cortex samples), H3K4me3 (N=3 fetal cortex; N=2 adult anterior temporal cortex), and CTCF (N=2 fetal cortex). All assays were done using the ChIPmentation protocol [94]. Brain tissues were crosslinked with 1% formaldehyde at room temperature followed by glycine quenching. To isolate nuclei, tissues were lysed with Lysis buffer I (10 mM HEPES; pH 7.5, 10 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, 0.5% Nonidet-40, and protease inhibitor cocktail) for 10 minutes at 4°C. The collected nuclei were then washed with a lysis buffer II (200mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH8.0, 10 mM Tris-Cl pH 8.0 and protease inhibitor cocktail) for 20 minutes at room temperature. The nuclei pellets were resuspended in lysis buffer III (10mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, and protease inhibitor cocktail) for sonication. The chromatin was sheared for 10 cycles (15 seconds on and 45 seconds off at constant power 3) on Branson 450 sonifier. For the pulldowns, 20-50 µl of chromatin was used for H3K4me3 (Abcam, ab8580) and H3K27ac (Abcam, ab4729), and 100-150 µl for CTCF (Abcam, ab70303). First, 11 µl of Dynabeads M-280 (Life Technologies, Sheep Anti-Rabbit IgG, Cat# 11204D) was washed three times with 0.5 mg/ml of BSA/PBS on ice and then incubated with each designated antibody for at least 2 hours at 4°C. The bead-antibody complexes were then washed with BSA/PBS. The pulldown was down in binding buffer (1% Trixon-X 100, 0.1% sodium deoxycholate, and protease inhibitor cocktail in 1X TE) by mixing the bead-antibody complexes and chromatin. After pulling down overnight, the bead-antibody-chromatin complexes were washed with RIPA buffer (50mM HEPES pH 8.0, 1% NP-40, 0.7% sodium deoxycholate, 0.5M LiCl, 1mM EDTA, and protease inhibitor cocktail). The bead complexes were then subjected to ChIPmentation by incubating with homemade Tn5 transposase in tagmentation reaction buffer (10mM Tris-Cl pH 8.0 and 5mM $MgCl_2$) for 10 minutes at 37°C. To remove free DNA, beads were washed twice with 1x TE on ice. The pulldown DNA was recovered by reversing crosslinks overnight followed by PCRClean DX beads purification. To generate ChIP-seq libraries, PCR was applied to amplify the pulldown DNA with illumina primers. Size selection was then done with PCRClean DX beads to choose fragments ranging from 100-1000 bp. All ChIP-seq libraries were sequenced on Illumina HiSeq2500 platform (50 bp single-end).

All ChIPmentation reads were mapped to hg19 of the human genome using Bowtie [70]. The first 36 bases of each reads were applied for alignment with up to 2 mismatches allowed. To remove duplication, only uniquely mapped reads were kept for further analysis. Peak calling was performed using MACS2 [95].

Open chromatin using ATAC-seq

Assay for transposase-accessible chromatin sequencing (ATAC-seq) was used to map chromatin accessibility genome-wide [96]. This method probes DNA accessibility with hyperactive Tn5 transposase, which inserts sequencing adapters into accessible regions of chromatin. We used tissue from adult DLPFC with no history of psychiatric or neurological disorders (N=137) [38]. Approximately 20 mg of pulverized brain tissue was used for ATAC-seq. Frozen samples were thawed in 1 ml of nuclear isolation buffer (20 mM Tris-HCL, 50 mM EDTA, 5mM Spermidine, 0.15 mM Spermine, 0.1% mercaptoethanol, 40% glycerol, pH 7.5), inverted for 5 minutes to mix, and samples were filtered through Miracloth to remove larger pieces of tissue. Samples were centrifuged at 1100 x *g* for 10 min at 4°C. The resulting pellet was washed with 50 µl RSB buffer, centrifuged again, and supernatant was removed. The final crude nuclear pellet was re-suspended in transposition reaction mix and libraries prepared for sequencing as described in Buenrostro et al. [96]. All samples were barcoded, and combined into pools. Each pool contained 8 randomly selected samples (selection balanced by case/control status and sex). Each pool was sequenced on two lanes of an Illumina 2500 or 4000 sequencer (San Diego, CA, USA). Raw fastq files were processed through cutadapt (version 1.2.0) [97] to remove adaptors and low-quality reads. cutadapt-filtered reads were aligned to hg19 using Bowtie2 [70] using default parameters. In alignment, all reads are treated as single-read sequences, regardless of whether ATAC-seq libraries were sequenced as single-end or paired-end. The aligned bam files were sorted using samtools (version 0.1.18) [98], duplicates removed using

Picard MarkDuplicates, and then converted to bed format using BedTools (version: v2.17.0) [99]. ENCODE "blacklist" regions (***URLs***) were removed (i.e., empirically identified genomic regions that produce artefactual high signal in functional genomic experiments). Narrow open chromatin peaks were called from the final bed files using MACS2.

Comparing Hi-C readouts to human genetic results

We compared the eHi-C readouts to multiple sets of human genetic results. We selected these phenotypes due to the availability of large empirical studies for common and rare genetic variation. For common variation, the CLOZUK study [10] identified 145 loci that met genome-wide significance in 40,675 schizophrenia cases and 64,643 controls. A genome-wide association study of cognitive ability [48] identified 205 genome-wide significant loci in 269,867 individuals.

We evaluated the connection between Hi-C readouts with schizophrenia and cognitive ability using partitioned LD score regression [100,101]. In effect, we estimate the degree to which the SNP-heritability of schizophrenia or cognitive ability is enriched in a set of genomic features (e.g., adult FIREs). Partitioned LD score regression is an extension of LD score regression allowing to estimate whether one or more sets of pre-specified genomic regions are enriched for the SNP-heritability of a trait based on GWAS summary statistics. Briefly, LD score regression [43] estimates common-variant SNP heritability by regressing the χ^2 association statistics from GWAS against the LD scores (the sum of r^2 for each SNP in a reference population). For multigenic traits, SNPs in high LD should on average be more associated with the trait than SNPs in low LD as they are expected to capture more of the genetic effects on the trait. The relationship between χ^2 statistics and LD scores is directly dependent on the proportion of genetic variance of the trait, which allows estimation of SNP-heritability [43].

Partitioned LD score regression [100] uses the same principle except that SNPs are partitioned into diverse functional categories. If some categories are enriched in causal variants, the relationship between χ^2 statistics and LD scores should be stronger than for categories with few causal variants. This allows estimation of the degree of enrichment of SNP-heritability in one or more functional categories. We used the partitioned LD score regression baseline model which consists of 53 functional categories. LD scores are computed for each annotation based on the presence of the SNP in the annotation (1 if a SNP is located in the annotation, 0 otherwise). For each annotation of interest (e.g. FIREs), we added SNPs from the "baseline model" located within the genomic coordinates of the regions as an extra annotation (1 if a SNP is in the region, 0 otherwise). For heritability enrichment, we added an extra annotation surrounding the annotation of interest (e.g. FIREs) by 500 bp on each side as recommended in order to prevent upward bias in the heritability enrichment estimates [100]. Significance was assessed using the enrichment P-value, which is not corrected for other genomic annotations (e.g. conserved regions). For comparison of annotations across tissues, we only added the annotation of interest and used the coefficient Z-score P-value, which is corrected for other genomic annotations, to compare the same annotation across different tissues (as recommended) [100]. For continuous annotations across the genome (e.g., FIRE scores), we added the continuous value as an extra annotation to the baseline model and the relevance of the continuous annotation was estimated using the coefficient Z-score P-value [101].

Hi-C features and human evolutionary history

We assessed two main ideas. First, we wished to evaluate whether TADs had any relation to LD blocks. TADs and LD blocks are Kb to Mb sized genomic regions that define key regions of interest in functional genomics and human genetics. We did this by evaluating recombination rates and LD decay centered on TAD boundaries. Second, given the importance of brain FIREs, we assessed whether these regions had evidence of evolutionary selection.

TAD boundaries and recombination rates

To examine the relationship between TAD boundaries and recombination rates, we downloaded the HapMap recombination map (***URLs***) [102]. We divided the genome in 40 Kb bins and excluded regions with <10 SNPs as well as those with poor mappability, GC content, centromeric location, or poor performance in functional genomic assays via the ENCODE blacklist. Each remaining 40kb bin was dichotomized as a TAD boundary or not, based on our TAD calling results. For each bin, we calculated minimal, maximal and median recombination rate. Previous studies have suggested that GC content and SNP density both influence recombination rate [103,104]. We therefore calculated summary statistics for these two factors and tested their differences in TAD boundaries and non-TAD boundary bins.

To test whether recombination rates differ in TAD boundaries, we took two approaches, one based on multiple regression and that other on resampling to account for GC content and SNP density [103,104]. The regression approach simultaneously adjusted for GC content and SNP density [103,104]. The resampling approach also controlled for GC content or SNP density. Specifically, let n denote the number of TAD boundary bins. We repeated the following procedure 10,000 times: from the set of TAD non-boundary bins, we created a subset s_i, $i = 1,2, \ldots 1,000$ of length n bins, by sampling bins, with replacement, such that the distribution of GC content (or SNP density) in s_i matched that in the set of TAD non-boundary bins. For each s_i, we calculated the median (across the n bins) of relevant summary statistics (median, mean and max recombination rates), t_i. We thus generated an empirical null distribution based on which empirical p-values were derived for significance testing.

TAD boundaries and linkage disequilibrium

To examine the relationship between TAD boundaries and LD, we computed LD r^2 values for variants in the 1000 Genomes Project [67]. After removing structural variants [105] and SNPs overlapping with small insertions and deletions, we grouped SNPs by whether they reside in a TAD boundary bin, similarly as in the previous recombination rate analysis. We then evaluated LD decay pattern for SNPs in or not in TAD boundary bins.

Brain FIREs and ancient positive selective sweep

We first determined whether brain FIREs have evidence of positive selection since human divergence from Neanderthals, as indicated by the top 5% Neanderthal selective sweep scores (NSS, ***URLs***) [106]. We first compared brain FIRE bins to non-FIRE bins to assess enrichment for NSS scores. Since FIRE and non-FIRE bins differ in a number of other aspects, which may also relate to positive selection since divergence from Neanderthals, we performed Fisher's exact test as well as performed logistic regression analysis adjusting for GC content.

Since FIREs are dichotomized by thresholding on continuous FIRE scores (***Online Methods***), signals may be diluted by lumping the remaining ~95% all as the "other" non-FIRE category. We therefore also (a) contrasted the extremes (i.e., top versus bottom 5% of FIRE score bins, or FIREs versus DIREs, depleted interacting regions) using logistic regression and (b) regressed on continuous FIRE scores. For covariates in regression models, we assessed GENCODE gene density, CTCF intensity, the histone marks brain H3K27ac and H3K4me3), open chromatin in DLPFC, and enhancer status in hippocampus (***Table S2***). We included only those that were nominally significant at 0.05 level as covariates in testing for FIRE effects.

Brain FIREs and extent of population differentiation

Global population differentiation (measured by the fixation index, F_{st}) is informative for soft selective sweeps where adaptive mutations increase in frequency but do not reach fixation [107,108]. We obtained global F_{st} scores from the 1000 Genomes Selection Browser (***URLs***) [109]. For each bin, we calculated summary statistics (including mean and median) for F_{st} scores for SNPs within the bin.

Brain FIREs and signals of recent positive selection

Signatures of classical and recent selective sweeps can be gauged by a number of metrics [110-114]. We focused on the integrated haplotype score (iHS) statistic that accommodates variants that have not reached fixation for signals of recent selection, integrating signals over 1,000 generations [111,114]. We obtained iHS statistics from the 1000 Genomes Selection Browser (***URLs***) [109]. For each bin, we calculated summary statistics (including mean and median) for iHS scores for SNPs within the bin.

Brain FIREs and extremely rare variant frequency

We assessed mutation tolerance for extremely rare variant frequency by comparing singleton and doubleton density based on TOPMed freeze 5b (***URLs***, N=62,784 whole genome sequences).

URLs

1000 Genomes Selection Browser, http://hsb.upf.edu

ENCODE "blacklists", http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human

FUMA, http://fuma.ctglab.nl/gene2func

gnomAD, http://gnomad.broadinstitute.org

GREAT, http://bejerano.stanford.edu/great

HapMap recombination map, ftp://ftp.ncbi.nlm.nih.gov/hapmap/recombination/2011-01_phaseII_B37

HUGIn, http://yunliweb.its.unc.edu/hugin

Juicer, https://github.com/theaidenlab/juicer

Neanderthal selection, ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/ntSssTop5p.txt.gz

NHGRI/EBI GWAS catalog, https://www.ebi.ac.uk/gwas

PsychENCODE portal, https://www.synapse.org//#!Synapse:syn4921369/wiki/235539

Psychiatric Genomics Consortium, http://www.med.unc.edu/pgc/results-and-downloads

SALMON, https://combine-lab.github.io/salmon

TOPMed, https://www.nhlbiwgs.org

TOPMed Bravo server, https://bravo.sph.umich.edu/freeze5

Competing Financial Interests

PFS reports the following potentially competing financial interests in the past 3 years: Lundbeck (advisory committee, grant recipient), Pfizer (Scientific Advisory Board), and Element Genomics (consultation fee).

Acknowledgements & Funding

PFS gratefully acknowledges support from the Swedish Research Council (Vetenskapsrådet, award D0886501) and the UNC Departments of Genetics and Psychiatry. MH and IJ were supported by NIH grant U54 DK107977. CC, YY, JSM, and YL are supported by NIH grant R01 HL129132 (awarded to YL). PGR was supported by NIMH grant K01 MH109772. LL, XL, YanL, and FJ are funded by R01 HG009658 (awarded to FJ). P.R.J. was funded by the Sophia Foundation for scientific research (grant nr: S14-27). The primary schizophrenia GWAS data were generated with support from Medical Research Council (MRC) Centre (MR/L010305/1), Program Grant (G0800509) and Project Grant (MR/L011794/1), and funding from the European Union's Seventh Framework

Programme for research, technological development and demonstration under grant agreement n° 279227 (CRESTAR Consortium). The authors acknowledge contributions by Drs. Herbert Meltzer and Bryan Roth, the assistance of Lesa Dieter, and the support of the families of the deceased and the staff of the Cuyahoga County Medical Examiner's Office, Cleveland, Ohio. The tissue collection and psychiatric assessment was supported by NIGMS COBRE Center for Psychiatric Neuroscience (P30 GM103328, awarded to CAS). We thank the NIH NeuroBioBank for providing the fetal brain samples used in this study. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The PsychENCODE Consortium was supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, and P50MH106934 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University).

Table 1. Analysis of Hi-C features

Hi-C feature	*Dataset*	*Analysis*	*Summary of significant findings*
FIREs	Fetal	GREAT	CNS axonogenesis, stem cell differentiation, neural nucleus development
		Model	Enriched: H3K27ac and CTCF in fetal cortex Depleted: H3H4me3 in fetal cortex and TSS
	Adult	GREAT	Regulation of autophagy, phosphoprotein phosphatase, neural nucleus development, detection of calcium ion
		Model	Enriched: typical enhancers, open chromatin, & H3K27ac in adult cortex Depleted: H3K4me3 in adult cortex
	Ventricle	GREAT	Actin cytoskeleton, myofibril, contractile fiber, sarcomere
Super FIREs	Fetal	GREAT	Stem cell/neuron/CNS differentiation, neurogenesis, cortex radial cell migration
	Adult	GREAT	Cytoskeletal protein binding, ion channel binding, membrane organization, neuron differentiation/migration, etc.
	Ventricle	GREAT	Actin binding, cytoskeletal protein binding
Chromatin Interactions	Fetal	GREAT	Transcription regulation, glial proliferation, oligodendrocyte differentiation, growth cone
		Model	Enriched: H3K27ac and CTCF in fetal cortex Depleted: H3K4me3 & gene expression in fetal cortex
	Adult	GREAT	Postsynaptic density, excitatory synapse
		Model	Enriched: open chromatin, typical enhancer, CTCF, & H3K27ac in adult cortex; TSS Depleted: H3K4me3 in fetal cortex
TAD Boundaries	Fetal	GREAT	None
		Model	Enriched: CTCF in fetal cortex; TSS Depleted: H3K27ac in fetal cortex
	Adult	GREAT	None
		Model	Enriched: CTCF in adult cortex; TSS Depleted: typical enhancer in adult cortex
	Ventricle	GREAT	None

*GREAT results for biological enrichments are summarized above with full results in **Table S3** (binomial FDR q-value < 0.01). Super FIREs are relatively few, and the results represent all fetal brain samples (our fetal cortex plus fetal germinal zone and cortical plate) and all adult brain samples (our adult cortex plus DLPFC and hippocampus). Most of the genome had ≥1 chromatin interactions, so we evaluated genomic regions with ≥50 interactions for our adult brain data and ≥100 interactions for our fetal data. The statistical modeling is also summarized above (there were no available data for heart ventricles), and full results are in **Table S4**.*

Table 2. Gene set analyses of genes implicated by location or by functional genomic evidence

GO category	*Gene set*	*eQTL and/or Hi-C*	*Location*
BP	Cellular response to UV	yes	
	Chromosome organization	yes	
	Establishment of localization in cell	yes	
	Modulation of synaptic transmission	yes	
	Positive regulation of mitochondrion organization	yes	
	Regulation of cellular component biogenesis	yes	
	Regulation of dephosphorylation	yes	
	Regulation of mitochondrion organization	yes	
	Regulation of organelle organization	yes	
	Regulation of synaptic plasticity	yes	
CC	Cell body	yes	yes
	Cell projection	yes	yes
	Chromatin	yes	
	Dendrite		yes
	Excitatory synapse		yes
	Mitochondrial part	yes	
	Mitochondrion	yes	yes
	Neuron part	yes	yes
	Neuron projection	yes	yes
	Neuron spine		yes
	Postsynapse		yes
	Ribonucleoprotein granule	yes	
	Somatodendritic compartment	yes	yes
	Synapse		yes
	Synapse part		yes
MF	Chaperone binding	yes	
	Enzyme binding	yes	
	Identical protein binding	yes	
	Poly A RNA binding	yes	
	RNA binding	yes	

*Gene set analyses for schizophrenia, contrasting LD/location-based gene implication and implication by results for biological enrichments. Significant gene sets (corrected P < 0.001) are summarized above with full results in **Table S6**.*

References

1 Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5-22, doi:10.1016/j.ajhg.2017.06.005 (2017).

2 Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187-1192 (2003).

3 McClellan, J. M., Susser, E. & King, M. C. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* **190**, 194-199 (2007).

4 Goldstein, D. B. Common genetic variation and human traits. *N Engl J Med* **360**, 1696-1698 (2009).

5 Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* **19**, 571-577, doi:10.1038/nn.4267 (2016).

6 Steinberg, S. *et al.* Truncating mutations in RBM12 are associated with psychosis. *Nat Genet* **49**, 1251-1254, doi:10.1038/ng.3894 (2017).

7 Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233, doi:10.1016/j.neuron.2015.09.016 (2015).

8 Genovese, G. *et al.* Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nature Neuroscience* **19**, 1433-1441, doi:10.1038/nn.4402 (2016).

9 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

10 Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).

11 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190 (2014).

12 Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47, doi:10.1038/nature18642 (2016).

13 Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry* **175**, 15-27, doi:10.1176/appi.ajp.2017.17030283 (2018).

14 Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984-994, doi:10.1038/ng.2711 (2013).

15 Bulik-Sullivan, B. K. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241, doi:1546-1718 (Electronic) 1061-4036 (Linking) (2015).

16 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).

17 Roussos, P. *et al.* A role for noncoding variation in schizophrenia. *Cell reports* **9**, 1417-1429, doi:10.1016/j.celrep.2014.10.015 (2014).

18 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).

19 Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-183, doi:10.1038/nature16549 (2016).

20 Dekker, J. Mapping the 3D genome: Aiming for consilience. *Nat Rev Mol Cell Biol* **17**, 741-742, doi:10.1038/nrm.2016.151 (2016).

21 Dekker, J. Gene regulation in the third dimension. *Science* **319**, 1793-1794, doi:10.1126/science.1152850 (2008).

22 Ethier, S. D., Miura, H. & Dostie, J. Discovering genome regulation with 3C and 3C-related technologies. *Biochim Biophys Acta* **1819**, 401-410, doi:10.1016/j.bbagrm.2011.12.004 (2012).

23 Pomerantz, M. M. *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* **41**, 882-884, doi:10.1038/ng.403 (2009).

24 Wright, J. B., Brown, S. J. & Cole, M. D. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol* **30**, 1411-1420, doi:10.1128/MCB.01384-09 (2010).

25 Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-375, doi:10.1038/nature13138 (2014).

26 Lu, L., Liu, X., Peng, J., Li, Y. & Jin, F. Easy Hi-C: A simple efficient protocol for 3D genome mapping in small cell populations. doi:10.1101/245688 (Submitted).

27 Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell reports* **17**, 2042-2059, doi:10.1016/j.celrep.2016.10.061 (2016).

28 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

29 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

30 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).

31 Burgess, D. J. Epigenomics: Deciphering non-coding variation with 3D epigenomics. *Nat Rev Genet* **18**, 4, doi:10.1038/nrg.2016.161 (2016).

32 Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol* **17**, 743-755, doi:10.1038/nrm.2016.104 (2016).

33 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531 (2017).

34 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 772, doi:10.1038/nrg.2016.147 (2016).

35 Krijger, P. H. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* **17**, 771-782, doi:10.1038/nrm.2016.138 (2016).

36 Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* **17**, 756-770, doi:10.1038/nrm.2016.126 (2016).

37 Rajarajan, P., Gil, S. E., Brennand, K. J. & Akbarian, S. Spatial genome organization and cognition. *Nat Rev Neurosci* **17**, 681-691, doi:10.1038/nrn.2016.124 (2016).

38 Bryois, J. *et al.* Evaluation of Chromatin Accessibility in Prefrontal Cortex of Schizophrenia Cases and Controls. *Nature communications* (In press).

39 Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res*, doi:10.1101/gr.232488.117 (2018).

40 Baró, G. S. & Sestan, N. Integrative Functional Genomic Analysis of the Developing Human Brain. (Submitted).

41 International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299-1320, doi:10.1038/nature04226 (2005).

42 Whalen, S. & Pollard, K. S. Most regulatory interactons are not in linkage disequilibrium. doi:http://dx.doi.org/10.1101/272245 (Submitted).

43 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).

44 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).

45 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. doi:10.1101/274654 (Submitted).

46 GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

47 Skene, N. G. *et al.* Brain cell types and the genetic basis of schizophrenia. *Nature Genetics* (2018).

48 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* **50**, 912-919, doi:10.1038/s41588-018-0152-6 (2018).

49 Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668-681, doi:10.1038/s41588-018-0090-3 (2018).

50 Inlow, J. K. & Restifo, L. L. Molecular and comparative genetics of mental retardation. *Genetics* **166**, 835-881 (2004).

51 McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* **80**, 588-604 (2007).

52 Nguyen, H. T. *et al.* Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Med* **9**, 114, doi:10.1186/s13073-017-0497-y (2017).

53 International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-241 (2008).

54 CNV Working Group of the Psychiatric Genomics Consortium & Schizophrenia Working Group of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* **49**, 27-35, doi:10.1038/ng.3725 (2017).

55 Szatkiewicz, J. *et al.* Copy number variation in schizophrenia in Sweden. *Molecular Psychiatry* **19**, 762-773 (2014).

56 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237, doi:10.1016/j.tig.2016.01.003 (2016).

57 Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* **45**, 1150-1159, doi:10.1038/ng.2742 (2013).

58 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).

59 Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nature communications* **6**, 10069, doi:10.1038/ncomms10069 (2015).

60 Duggal, G., Wang, H. & Kingsford, C. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res* **42**, 87-96, doi:10.1093/nar/gkt857 (2014).

61 GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

62 de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat Rev Genet*, doi:10.1038/nrg.2016.29 (2016).

63 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).

64 Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome are associated with schizophrenia risk. (Submitted).

65 Rhie, S. K. *et al.* A 3D Epigenomic Map of Olfactory Neuronal Cells Reveals Schizophrenia-Associated Genes. (Submitted).

66 Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58, doi:nature09298 [pii] 10.1038/nature09298 (2010).

67 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

68 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).

69 Encode Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046, doi:10.1371/journal.pbio.1001046 (2011).

70 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

71 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).

72 Hu, M. *et al.* HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics* **28**, 3131-3133, doi:10.1093/bioinformatics/bts570 (2012).

73 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).

74 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

75 Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res* **24**, 999-1011, doi:10.1101/gr.160374.113 (2014).

76 Xu, Z., Zhang, G., Wu, C., Li, Y. & Hu, M. FastHiC: a fast and accurate algorithm to detect long-range chromosomal interactions from Hi-C data. *Bioinformatics* **32**, 2692-2695, doi:10.1093/bioinformatics/btw240 (2016).

77 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* (2015).

78 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).

79 Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nat Methods* **6**, 690-691, doi:10.1038/nmeth1009-690 (2009).

80 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).

81 Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58**, 192-203, doi:10.1016/j.ymeth.2012.07.022 (2012).

82 Dostie, J., Zhan, Y. & Dekker, J. Chromosome conformation capture carbon copy technology. *Curr Protoc Mol Biol* **Chapter 21**, Unit 21 14, doi:10.1002/0471142727.mb2114s80 (2007).

83 Ferraiuolo, M. A., Sanyal, A., Naumova, N., Dekker, J. & Dostie, J. From cells to chromatin: capturing snapshots of genome organization with 5C technology. *Methods* **58**, 255-267, doi:10.1016/j.ymeth.2012.10.011 (2012).

84 Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* **8**, 509-524, doi:10.1038/nprot.2013.018 (2013).

85 Mitchell, A. C. *et al.* The genome in three dimensions: a new frontier in human brain research. *Biol Psychiatry* **75**, 961-969, doi:10.1016/j.biopsych.2013.07.015 (2014).

86 van Berkum, N. L. & Dekker, J. Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol Biol* **567**, 189-213, doi:10.1007/978-1-60327-414-2_13 (2009).

87 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).

88 Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72**, 65-75, doi:10.1016/j.ymeth.2014.10.031 (2015).

89 Zhu, H. *et al.* Quantitative analysis of focused A-to-I RNA editing sites by ultra-high-throughput sequencing in psychiatric disorders. *PLoS ONE* **7**, e43227, doi:10.1371/journal.pone.0043227 (2012).

90 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

91 Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521, doi:10.12688/f1000research.7563.2 (2015).

92 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

93 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

94 Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* **12**, 963-965, doi:10.1038/nmeth.3542 (2015).

95 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

96 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

97 Martin, M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal* **17**, doi:10.14806/ej.17.1.200.

98 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:btp352 [pii] 10.1093/bioinformatics/btp352 (2009).

99 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).

100 Finucane, H. K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. *Nature Genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

101 Finucane, H. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics*, doi:10.1101/103069 (2017).

102 Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321-324, doi:10.1126/science.1117196 (2005).

103 Jeffreys, A. J. & Neumann, R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum Mol Genet* **14**, 2277-2287, doi:10.1093/hmg/ddi232 (2005).

104 Jorgenson, E. *et al.* Ethnicity and human genetic linkage maps. *Am J Hum Genet* **76**, 276-290, doi:10.1086/427926 (2005).

105 Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81, doi:10.1038/nature15394 (2015).

106 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

107 Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet* **10**, 639-650, doi:10.1038/nrg2611 (2009).

108 Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370, doi:10.1111/j.1558-5646.1984.tb05657.x (1984).

109 Pybus, M. *et al.* 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res* **42**, D903-909, doi:10.1093/nar/gkt1188 (2014).

110 Schrider, D. R. & Kern, A. D. Soft Sweeps Are the Dominant Mode of Adaptation in the Human Genome. *Mol Biol Evol* **34**, 1863-1877, doi:10.1093/molbev/msx154 (2017).

111 Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764, doi:10.1126/science.aag0776 (2016).

112 Berg, J. J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet* **10**, e1004412, doi:10.1371/journal.pgen.1004412 (2014).

113 Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu Rev Genet* **47**, 97-120, doi:10.1146/annurev-genet-111212-133526 (2013).

114 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72, doi:10.1371/journal.pbio.0040072 (2006).

115 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986, doi:10.1038/ng.3359 (2015).

116 Global Lipids Genetics Consortium *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283, doi:10.1038/ng.2797 (2013).

117 Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* **18**, 988-996, doi:ddn351 [pii] 10.1093/hmg/ddn351 (2009).

118 Stoll, G. *et al.* Deletion of TOP3beta, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. *Nature Neuroscience* **16**, 1228-1237, doi:10.1038/nn.3484 (2013).

Supplemental Note – Additional evaluation of brain Hi-C data

Giusti-Rodriguez et al., Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex.

Table of Contents

COMPARATIVE ANALYSIS OF BRAIN EHI-C DATA WITH EXTERNAL HI-C DATASETS 2

HI-C AND GENE EXPRESSION IN BRAIN 2
TOPOLOGICALLY ASSOCIATING DOMAINS (TADS) 2
FREQUENTLY INTERACTING REGIONS (FIRES) 2
CHROMATIN INTERACTIONS 3

TADS AND LD 4

RECOMBINATION RATES 4
LINKAGE DISEQUILIBRIUM DECAY 4

BRAIN FIRES AND HUMAN EVOLUTIONARY HISTORY 4

ANCIENT POSITIVE SELECTIVE SWEEPS 5
EXTENT OF POPULATION DIFFERENTIATION 5
SIGNALS OF RECENT POSITIVE SELECTION 5
FREQUENCY OF EXTREMELY RARE VARIANTS 5

HI-C AND GENE EXPRESSION IN BRAIN 5

FIRES IN ADULT AND FETAL BRAIN AND SPECIFICITY OF GENE EXPRESSION 5
DEVELOPMENTAL DIFFERENCES AND THE BRAIN CHROMATIN INTERACTOME 6
CHROMATIN INTERACTOME AND EQTLS 6

REFERENCES 8

Comparative analysis of brain eHi-C data with external Hi-C datasets

We investigated four major Hi-C readouts in our adult and fetal eHi-C data. Here we compare our data to a total of 25 Hi-C datasets from a range of human tissues and cell lines. The key readouts we evaluated were:

- At a 100 kb scale, we used these Hi-C data to determine "A" (active) and "B" (inactive) compartments [1].
- Topologically associating domains (TADs) [2,3] are Mb-scale regions within which chromatin interactions have a strong tendency to occur [2,3].
- Frequently interacting regions (FIREs, 40 kb) are genomic regions with significantly greater *cis*-connectivity than expected under a null hypothesis of random collisions. [4,5] FIREs can exist in isolation or in contiguous blocks called superFIREs [6].
- Chromatin interactions are a key Hi-C readout, and identify genomic regions that are physically proximal in the nuclear 3D space despite being far apart in 1D genomic location [5-12] (FIREs are the subset of chromatin interactions with considerably more 3D contacts). Chromatin interactions in this study consisted of two anchors (10 kb, 20 Kb to 2 Mb apart). Some chromatin interactions occur transiently, randomly, or in subsets of cells, but many others are stable structural or regulatory features of cells in a tissue.

Hi-C and gene expression in brain

We used these Hi-C data to determine "A" (active) and "B" (inactive) compartments [1] as data on chromatin spatial organization in brain is limited. Using Hi-C contact matrices in 100 Kb bins, we extracted PC1 that was combined with transcription start site (TSS) density to assign A versus B compartments. ***Figure S2a*** shows results of a principal component analysis (PCA) applied to the compartment scores in 100 Kb bins. Fetal brain samples clustered tightly on PC1 and PC2 as did adult brain samples.

The 100 Kb compartment A/B assignments had far stronger overlap within the adult brain samples (ORs >150, $P<1x10^{-300}$) and within the fetal brain samples (ORs >95, $P<1x10^{-300}$) than between our adult and fetal brain samples (OR=24.5, $P<1x10^{-300}$). A heat map of all pairwise Jaccard indices (***Figure S2b***) paralleled the PCA with tight clustering of adult and fetal brain samples. Jaccard indices for the adult brain samples were around 0.83, fetal brain samples around 0.83, and adult versus fetal 0.66 (for comparison, the Jaccard index for left versus right heart ventricle was 0.91).

Topologically associating domains (TADs)

Chromatin interactions have a strong tendency to occur within TADs (discrete megabase-scale regions with less frequent interactions outside the regions) [2,3]. ***Figure S3a*** shows results of PCA applied to insulation scores in 40 Kb bins (used to define TAD boundaries) [13]. As with FIRE scores, brain samples with high read depth clustered on the dominant PC1.

As anticipated from ***Figure S3a***, TAD boundaries in six brain Hi-C datasets strongly overlapped with pairwise ORs generally exceeding 50 ($P<1x10^{-300}$). A heat map of all pairwise Jaccard indices showed prominent clustering of TAD boundaries in our adult and fetal data along with fetal cortical plate and germinal zone (although not with the lower read depth DLPFC and hippocampal samples, ***Figure S3b***). Adult and fetal TAD boundaries clustered very strongly (OR=173, $P<1x10^{-300}$) with a Jaccard index of 0.41 (for comparison, the Jaccard index for TAD boundaries in left and right heart ventricle was 0.48).

Frequently interacting regions (FIREs)

FIREs [4,5] are perhaps the simplest Hi-C readout. FIREs are genomic regions with significantly greater *cis*-connectivity than expected under a null hypothesis of random collisions. FIREs capture key observations about the functional genome, that chromatin interactions cluster non-randomly in a tissue-specific manner. We

computed FIRE scores from Hi-C contact matrices (bin size 40 Kb, 15-200 Kb *cis* windows). ***Figure S4a*** shows PCA of FIRE scores in 25 Hi-C datasets. Brain samples with high read depth clustered on the dominant PC1 with clear separation of several cell lines.

We defined FIREs as 40 Kb genomic bins with significantly more Hi-C interactions (FIRE score $P<0.05$, an intentionally liberal threshold for sensitivity in cross-tissue comparisons given lesser read depths). We found that FIREs in our adult and fetal brain agreed well with external Hi-C datasets, and unsupervised clustering largely recapitulated cell identity. FIREs in our adult cortex data were strongly predictive of FIREs in the most similar tissue (adult DLPFC, OR=58.0, $P<1x10^{-300}$) as well as hippocampus (OR=5.67, $P<1x10^{-300}$) [4]. FIREs in fetal cortex were more strongly associated with FIREs in fetal cortical plate (OR=52.4, $P<1x10^{-300}$) than fetal germinal zone (OR=42.6, $P<1x10^{-300}$) [14]. Fetal cortex are more similar to post-mitotic fetal cortical plate neurons than to cells in the actively dividing germinal zone. Our adult and fetal FIREs had a lesser tendency to overlap (OR=8.24, $P<1x10^{-300}$). A FIRE was present in 14.17% of 40 Kb bins in at least 1 of 6 brain datasets, 23.95% of bins in 12 non-brain human tissue datasets [4], 18.18% of bins in 7 cell line datasets [4], and 31.29% of bins in any of these 25 datasets. A heat map of all pairwise Jaccard indices is in ***Figure S4b***, and clustering tended to follow cell identity.

Using GENCODE gene models, the presence of a FIRE was strongly associated with the number of protein-coding genes overlapping that bin in adult and fetal brain (Wilcoxon $P<1x10^{-300}$ for both). A FIRE was also associated with the number of protein-coding transcription start sites (TSS) in adult (Wilcoxon $P<1x10^{-300}$) but not fetal brain (Wilcoxon P=0.30) as well as with the number of protein-coding exons in adult (Wilcoxon $P<1x10^{-300}$) but not fetal brain (Wilcoxon P=0.01). There were no significant associations for other types of genes (e.g., RNA genes like miRNA or lincRNA).

FIREs tended to cluster in genomic regions. Adapting algorithms for super-enhancers [15,16], we identified regionally clustered FIREs or "super FIREs". There were 151 autosomal super FIREs in our adult cortex data (median size 320 Kb, interquartile range 280-440 Kb), and 186 super FIREs in fetal cortex data (median size 240 Kb, interquartile range 200-360 Kb). Across all 25 Hi-C datasets, a super FIRE was present in one sample in 9.03% of the genome, in two samples in 4.08%, and in ≥3 samples in 6.11%.

Chromatin interactions

Chromatin interactions are a key readout of Hi-C data, and identify genomic regions that are physically proximal in the nuclear 3D space despite being far apart in genomic location [5-12]. FIREs are a small subset of all chromatin interactions which have considerably more 3D contacts. Some chromatin interactions occur transiently, randomly, or in subsets of cells, but many others are stable structural or regulatory features of cells in a tissue.

We identified intra-chromosomal chromatin interactions between 10 Kb bins that were >20 Kb apart (i.e., not contiguous) and ≤2 Mb apart. This definition is analogous to that for *cis* gene expression eQTLs (typically defined as an eQTL-SNP within ±1 Mb of an eQTL-gene). More distal interactions may well occur; we did not evaluate these because most interactions are likely to be in *cis* and experience suggests limited resolution due to low raw read counts [4,17]. We evaluated chromatin interactions in four brain Hi-C datasets with sufficient read depth: our adult cortex and fetal cortex datasets plus with fetal cortical plate and germinal zone [14] (***Figure 1a***). Because we wished to use chromatin interactions to interpret GWAS results, we desired specificity over sensitivity and applied a conservative Bonferroni correction to each dataset ($P<2.31x10^{-11}$, or $\alpha=0.001$ corrected for 43,222,677 potential bin pair interactions).

Figure S5 shows a Venn diagram of autosomal chromatin interactions in four brain Hi-C datasets. Our fetal cortex data had the greatest number of chromatin interactions (1.25 million), somewhat more than the fetal cortical plate and germinal zone data (0.998 million in each), and over twice as many as in our adult cortex data (0.509 million). Even at a stringent threshold, one or more chromatin interactions were present in 53.2% of all 10 Kb genomic regions in our adult data (median 1, interquartile range 0-5) and 65.3% in our fetal data (median 3,

interquartile range 0-14). In our adult data, the median distance was 220 Kb (interquartile range 80-530 Kb), and 370 Kb in our fetal data (interquartile range 130-830 Kb). Greater numbers of chromatin interactions in earlier developmental stages were observed in our eHi-C results, similarly processed adult and fetal brain Hi-C datasets (D Geschwind, personal communication), differentiation of iPSCs into neurons and glia [18], and differentiation of mouse pluripotent stem cells [19]. Thus, chromatin interactions are common in brain, and greater chromatin interactions appear to characterize earlier developmental stages.

TADs and LD

We evaluated whether TADs had any relation to LD blocks. This question often arises as TADs and LD blocks [20] are both on the scale of 10^5-10^6 bases and define regions of interest in functional genomics and human genetics. TADs are defined via observations of 3D chromatin contacts in samples of cellular nuclei, and LD blocks are defined via observations of haplotype structure in human samples. We did this by evaluating recombination rates and LD decay centered on TAD boundaries.

Recombination rates

We found that GC content was slightly higher and SNP density slightly lower in 40 Kb genomic regions that contained a TAD boundary. The results were consistent in our adult and fetal eHi-C datasets (***Table S8a***). Resampling approaches taking into account differential GC content or variant density across bins resulted in largely significant *P*-values, supporting that TAD boundaries have lower recombination rates (***Table S8b***). For example, the median recombination rates in adult TAD boundaries is 0.1879 in contrast to 0.2423 in non-TAD boundary bins (resampling empirical *P*-values 0.0013 and <0.0001, matching distributions of SNP density and GC content respectively). This observation was consistent across adult and fetal eHi-C datasets, and across different summary statistics (median, maximum, or minimum). Multiple regression adjusting for GC content and SNP density confirmed the resampling findings (*P*=2.49e-4 and P=8.64e-3 for adult and fetal eHi-C datasets), with consistent directions of effects (i.e., TAD boundary bins have lower mean recombination rates).

Linkage disequilibrium decay

We evaluated the decay in LD (measured by r^2) by distance from TAD boundaries (***Figure S7***). The patterns of LD decay were not markedly greater in TAD boundaries. These findings were similar in adult and fetal eHi-C datasets.

Our results suggest that historical recombination events are depleted in TAD boundaries. Structural proteins like CTCF and cohesin are enriched at TAD boundaries in order to create and maintain TADs [7,21-23], and the presence of these proteins may interfere with recombination leading to lowered probability of crossover events. However, the depletion in recombination is modest and is not evident at the LD level. This suggests that differences in historical recombination rates are subtle or were diluted over generations of meiosis. For example, given the observed recombination rates (adult and fetal median values of 0.1879 and 0.2423) and 20 generations of recombination, the coefficient of LD ("D") will be D=0.015 for adult and D=0.0039 for fetal TAD boundaries [24], values close to linkage equilibrium. With 50 generations, the values of D would be even closer to linkage equilibrium (D=3.0e-5 for adult and D=9.4e-7 for fetal TAD boundaries).

Consistent with prior studies [14,25], these results imply that TADs and LD blocks are different and infrequently overlap. This conclusion is consistent with observations in highly inbred laboratory mice who have chromosome scale LD blocks but far smaller TADs that are conserved between mouse and human [2].

Brain FIREs and human evolutionary history

Given the importance of brain FIREs, we assessed whether these regions had evidence of evolutionary selection.

Ancient positive selective sweeps

We found that FIREs tend to be enriched in regions with signatures of selective sweeps since divergence from the Neanderthals, as indicated by top 5% of Neanderthal selective sweep scores (NSS). As expected, the difference is more pronounced when focusing on the two extremes (namely FIREs and DIREs, depleted interacting regions), resulting in significant *P*-values despite much smaller sample sizes after removing regions with FIRE scores falling in the middle 90%. For example, out of the FIREs identified in our adult samples, 39 (1.58%) overlapped top 5% NSS, while only one DIRE (0.14%) overlapped top 5% NSS regions (Fisher's exact test *P*=8.72e-4). When comparing FIREs and non-FIREs, the much diluted differences were no longer significant. We observed similar patterns in fetal and adult samples, when further adjusting for potential confounders (***Table S9a***).

Extent of population differentiation

As shown in ***Tables S9b-c***, F_{st} values tend to be slightly lower in FIREs than in DIREs, suggesting that variants in FIRE regions are less differentiated (or more conserved) across populations. These observations, consistent across fetal and adult, and regardless of the choice of summary statistics imply that FIRE regions are functionally important.

Signals of recent positive selection

Tables S9d-e show that integrated haplotype score (iHS) statistics tend to be higher in FIREs than in DIREs. This suggests that genetic variants in FIRE regions have stronger evidence for recent positive selection acting at or near SNPs in the FIRE bins, with evidence aggregated over 1,000 generations. These observations are consistent across fetal and adult and hold regardless of the choice of summary statistics (mean or median).

Frequency of extremely rare variants

We obtained sets of extremely rare variants from the Trans-Omics for Precision Medicine (TOPMed) Bravo server (***URLs***) which contains 463 million variants on 62,784 individuals with high-coverage whole genome sequence data (≥30X) [26]. We evaluated ultra-rare single nucleotide variants that were observed in only one or two TOPMed subjects (so-called singletons and doubletons with allele frequencies on the order of 8.0e-6 and 1.6e-5). As shown in ***Tables S9f-g***, we observed marked and highly significant depletion of rare single nucleotide variants in FIREs (regions with many Hi-C contacts) compared to DIREs (depleted interacting regions with few Hi-C contacts) in both adult and fetal brain (all P values <1e-8). For example, the mean number of singletons in adult FIRE bins is 2846.3, significantly less than 3074.7 in adult DIREs (Wilcoxon *P*<2.2e-16).

Taken together, these evolutionary analyses of several population genetic signatures resulted in very consistent findings. FIREs, particularly in contrast to DIREs, have stronger evidence for ancient and recent positive selection, less population differentiation, and fewer singleton and doubleton single nucleotide variants. These highly consistent observations suggest that brain FIREs are important genomic regions under stronger population genetic constraints.

Hi-C and gene expression in brain

Given the role of the chromatin interactome on transcriptional regulation, we evaluated the relative importance of chromatin interactome features and gene expression in fetal and adult brain.

FIREs in adult and fetal brain and specificity of gene expression

We used RNA-seq data on 26,871 autosomal genes with non-zero expression in one or more samples from nine fetal and nine adult cortices. After transformation, scaling, and quantile normalization, we obtained the mean gene expression in the fetal and adult samples and considered genes whose TSS were within 200 Kb of a FIRE bin.

Fetal gene expression was significantly greater than adult gene expression in the vicinity of a fetal FIRE (*P*=1.31e-4, ***Table S10***) although adult gene expression was marginally greater than fetal gene expression in the vicinity of an adult FIRE (*P*=0.016). However, these comparisons were more significant for developmentally specific FIREs. Fetal gene expression was greater than adult gene expression in the vicinity of fetal-specific FIREs (*P*=3.03e-15), and adult gene expression was greater than fetal gene expression in the vicinity of adult-specific FIREs (*P*=4.81e-13). These results suggest that FIREs play an important role in fetal/adult-specific transcription regulation.

Developmental differences and the brain chromatin interactome

We evaluated the relative importance of chromatin interactome features and gene expression in fetal and adult brain. We compared chromatin interaction anchors, FIRE scores, insulation scores, and A/B compartment scores. For example, regions with more anchors may be associated with greater gene expression; higher FIRE scores have more *cis* interactions and are usually near genes with high expression; insulation scores dip in TAD centers that are enriched for genes with tissue-specific expression; and A compartments are associated with higher gene density and gene expression.

To systematically evaluate the joint effect of four Hi-C readouts and gene expression, we mapped the TSS of all 26,871 autosomal genes into distinct 10 Kb bins (if multiple TSSs are within the same bin, the summed gene expression was assigned to that bin). For each 10Kb bin, we matched its corresponding loop anchors and FIRE scores separately for our fetal and adult datasets. Insulation and compartment scores were mapped similarly. Univariate analysis shows that the chromatin interactomic features between fetal and adult are positively correlated with the gene expression between fetal and adult. The Pearson correlation coefficients of gene expression difference (adult vs. fetal) with adult vs. fetal loop anchor difference, FIRE score difference, insulation score difference, and A/B compartment difference are: 0.0713 (*P*<2.2e-16), 0.0861 (*P*<2.2e-16), 0.0653 (*P*=1.33-e15), and 0.0444 (*P*=5.62e-9). We next evaluated active enhancer (H3K27ac) and active promoter marks (H3K4me3) in fetal and adult brain. As expected, developmental stage differences in these marks were associated with fetal- and adult-specific gene expression.

We fit a multi-variable linear regression model using gene expression differences between fetal and adult as the response variable and the dynamic loop anchor, FIRE score, insulation score, and A/B compartment score as predictors. Consistent with the univariate analyses, all four Hi-C readouts were significant predictors of dynamic gene expression (R^2=0.0124 with *P*<2.2e-16), suggesting that 1.24% of total variation of fetal-adult gene expression differences can be explained by differences in chromatin interactome between fetal and adult brain. We next evaluated H3K27ac and H3K4me3 as predictors of gene expression and found R^2=0.0387. Combining the dynamic chromatin interactomic and ChIP-Seq data, the R^2 improves to 0.0475, and three of four Hi-C readouts (loop number, FIRE score, and insulation score) were significant, suggesting that chromatin interactomic data provide orthogonal information compared to epigenetic data predicting dynamic gene expression.

Chromatin interactome and eQTLs

We evaluated the relation between chromatin interactions and eQTLs in adult cortex (***Figure 1b***). For eQTL results from the CommonMind Consortium [27], we selected 3,225,973 autosomal SNP-TSS eQTL pairs (distance 20 Kb to 2 Mb). We mapped each eQTL and TSS into 10 Kb bins for a total of 437,687 distinct bin pairs corresponding to eQTL-TSS pairs. We then selected the adult 10 Kb bin Hi-C contact matrix for all autosomal bin pairs 20 Kb to 2Mb apart. We divided the bin pairs into those that did (group A) and did not (group B) overlap with eQTL-TSS pairs. At each 1D genomic distance from 20 Kb to 1 Mb, the mean chromatin interaction frequency for bin pairs that overlap an eQTL-TSS pairs (group A) is significantly greater than for those that do not overlap eQTL-TSS pairs (group B). For example, for bin pairs at a distance of 100 Kb, the mean chromatin interaction frequency is 13.02 for group A and 11.03 for group B (two sample t-test *P*=2.02e-39). In addition, we identified 30,725 and 61,405 10 Kb bins as active promoters and active enhancers (based on adult H3K4me3 and H3K27ac ChIP-seq). Group A bin pairs that

had enhancer-enhancer, enhancer-promoter, or promoter-promoter interactions had strikingly greater interaction frequencies.

References

1 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

2 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

3 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).

4 Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell reports* **17**, 2042-2059, doi:10.1016/j.celrep.2016.10.061 (2016).

5 Burgess, D. J. Epigenomics: Deciphering non-coding variation with 3D epigenomics. *Nat Rev Genet* **18**, 4, doi:10.1038/nrg.2016.161 (2016).

6 Schmitt, A. D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol* **17**, 743-755, doi:10.1038/nrm.2016.104 (2016).

7 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531 (2017).

8 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 772, doi:10.1038/nrg.2016.147 (2016).

9 Dekker, J. Mapping the 3D genome: Aiming for consilience. *Nat Rev Mol Cell Biol* **17**, 741-742, doi:10.1038/nrm.2016.151 (2016).

10 Krijger, P. H. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* **17**, 771-782, doi:10.1038/nrm.2016.138 (2016).

11 Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* **17**, 756-770, doi:10.1038/nrm.2016.126 (2016).

12 Rajarajan, P., Gil, S. E., Brennand, K. J. & Akbarian, S. Spatial genome organization and cognition. *Nat Rev Neurosci* **17**, 681-691, doi:10.1038/nrn.2016.124 (2016).

13 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).

14 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).

15 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

16 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).

17 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).

18 Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome are associated with schizophrenia risk. (Submitted).

19 Novo, C. L. *et al.* Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell reports* **22**, 2615-2627, doi:10.1016/j.celrep.2018.02.040 (2018).

20 International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299-1320, doi:10.1038/nature04226 (2005).

21 Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56, doi:10.1038/nature24281 (2017).

22 McCord, R. P. Chromosome biology: How to build a cohesive genome in 3D. *Nature* **551**, 38-40, doi:10.1038/nature24145 (2017).

23 Pirrotta, V. Binding the boundaries of chromatin domains. *Genome Biol* **15**, 121, doi:10.1186/gb4183 (2014).

24 Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-485, doi:10.1038/nrg2361 (2008).

25 Whalen, S. & Pollard, K. S. Most regulatory interactons are not in linkage disequilibrium. doi:http://dx.doi.org/10.1101/272245 (Submitted).

26 Brody, J. A. *et al.* Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet* **49**, 1560-1563, doi:10.1038/ng.3968 (2017).

27 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).

Figure 1a

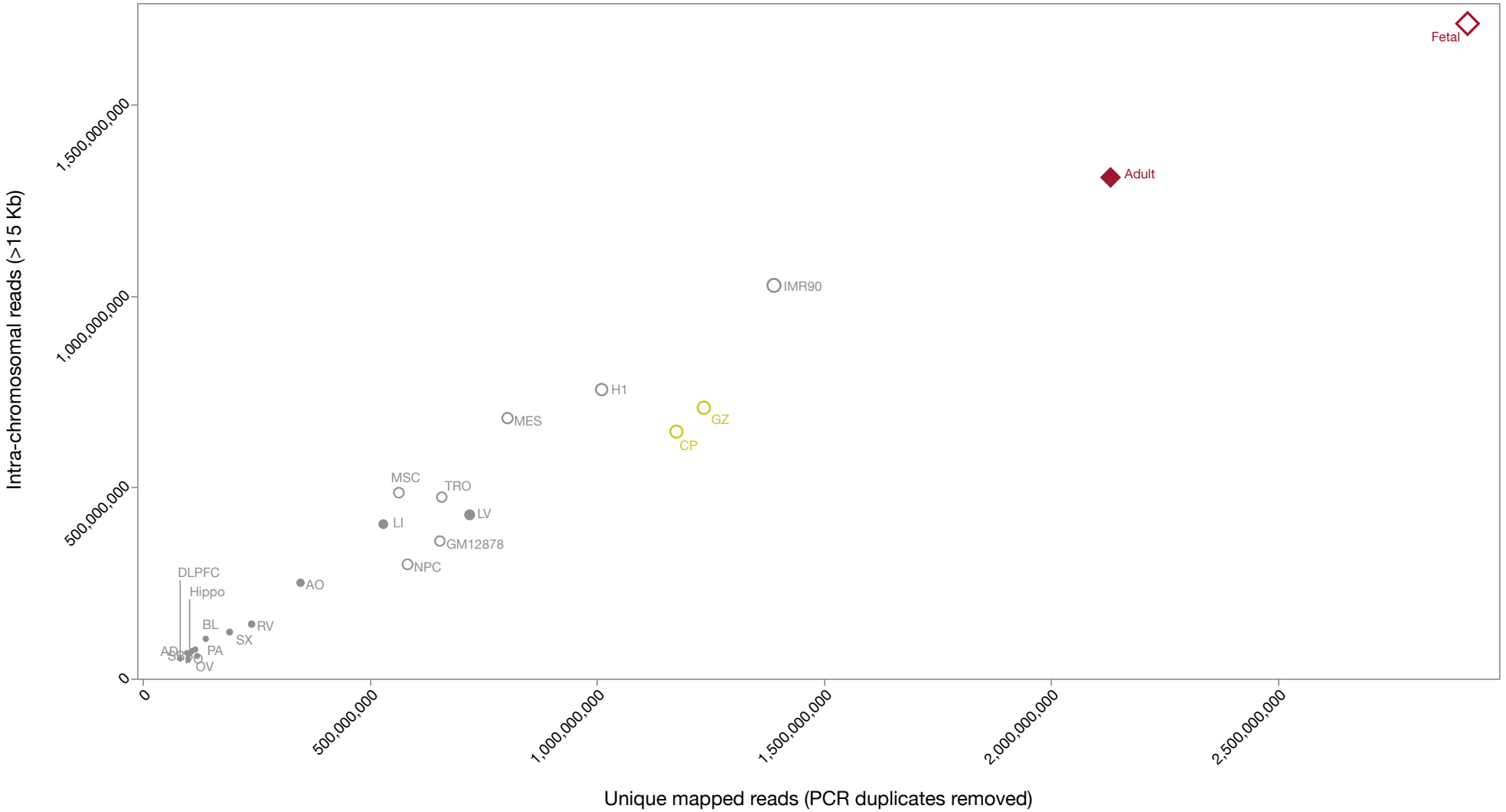

Intra-chromosomal reads (>15 Kb)
1,500,000,000
1,000,000,000
500,000,000
0
Fetal
Adult
IMR90
H1
MES
GZ
CP
MSC
TRO
LV
LI
GM12878
NPC
AO
DLPFC
Hippo
BL
RV
SX
PA
OV
0
500,000,000
1,000,000,000
1,500,000,000
2,000,000,000
2,500,000,000
Unique mapped reads (PCR duplicates removed)

Figure 1b

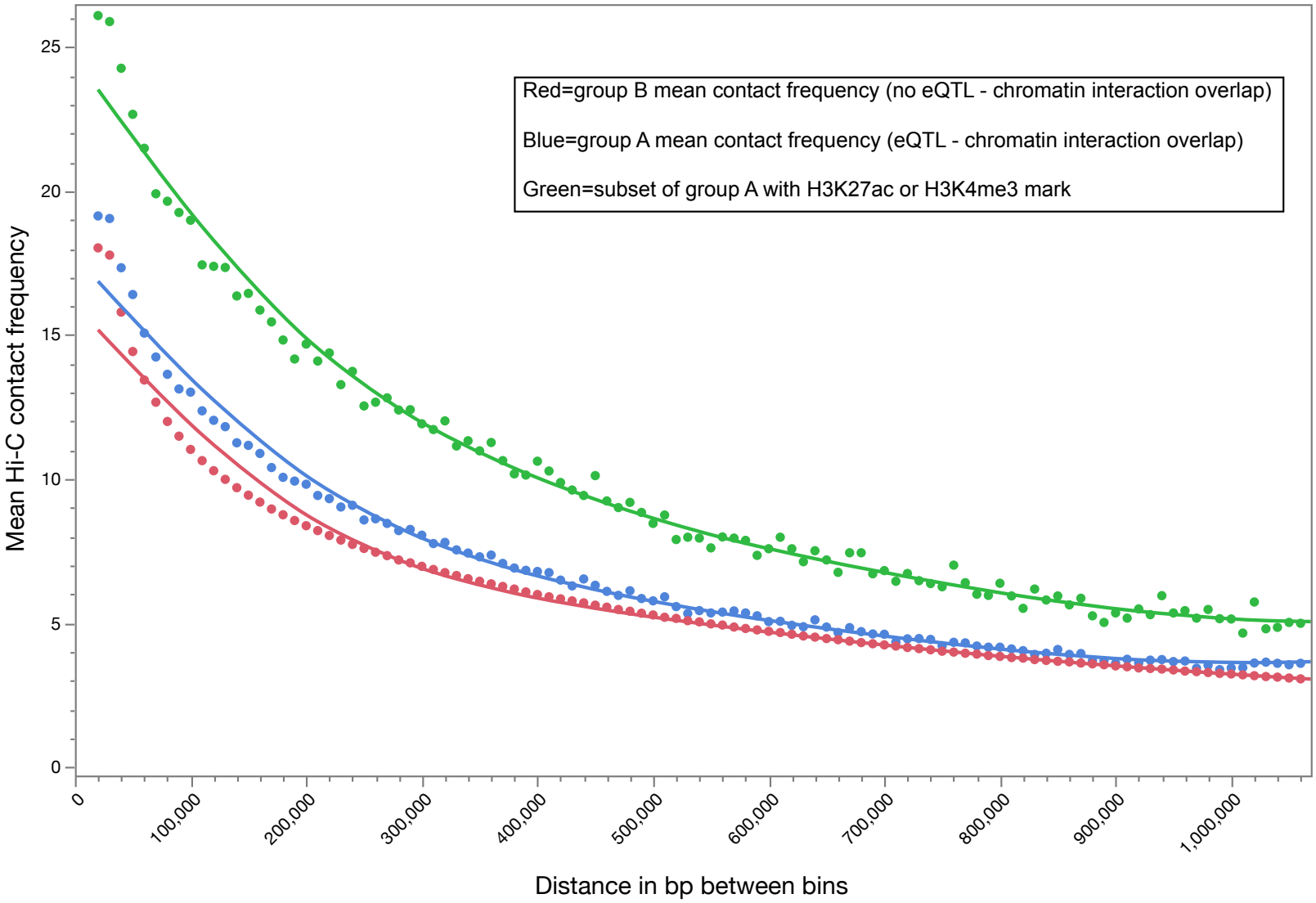

Red=group B mean contact frequency (no eQTL - chromatin interaction overlap)
Blue=group A mean contact frequency (eQTL - chromatin interaction overlap)
Green=subset of group A with H3K27ac or H3K4me3 mark
Mean Hi-C contact frequency
0
5
10
15
20
25
0
100,000
200,000
300,000
400,000
500,000
600,000
700,000
800,000
900,000
1,000,000
Distance in bp between bins

Figure 2

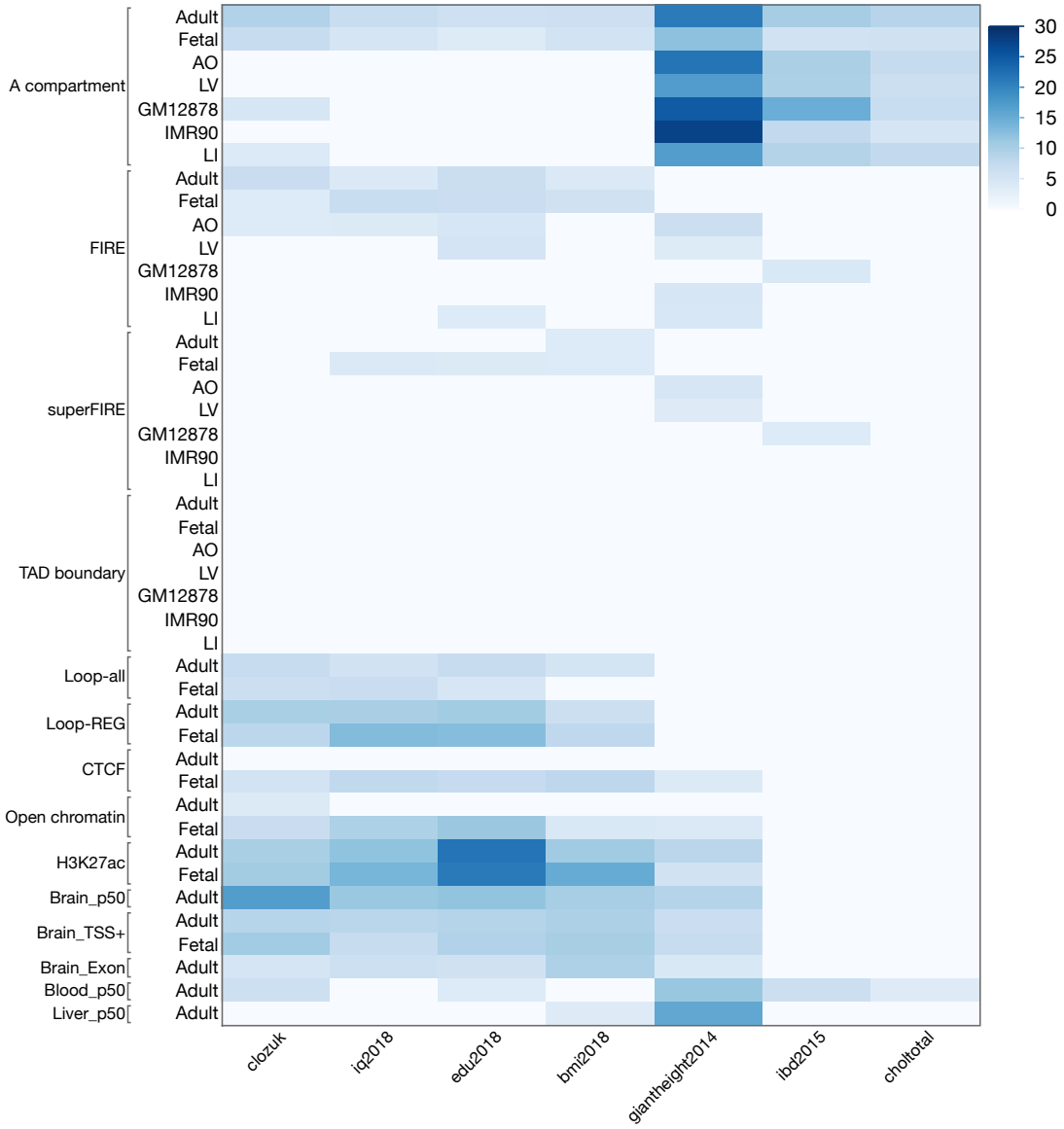
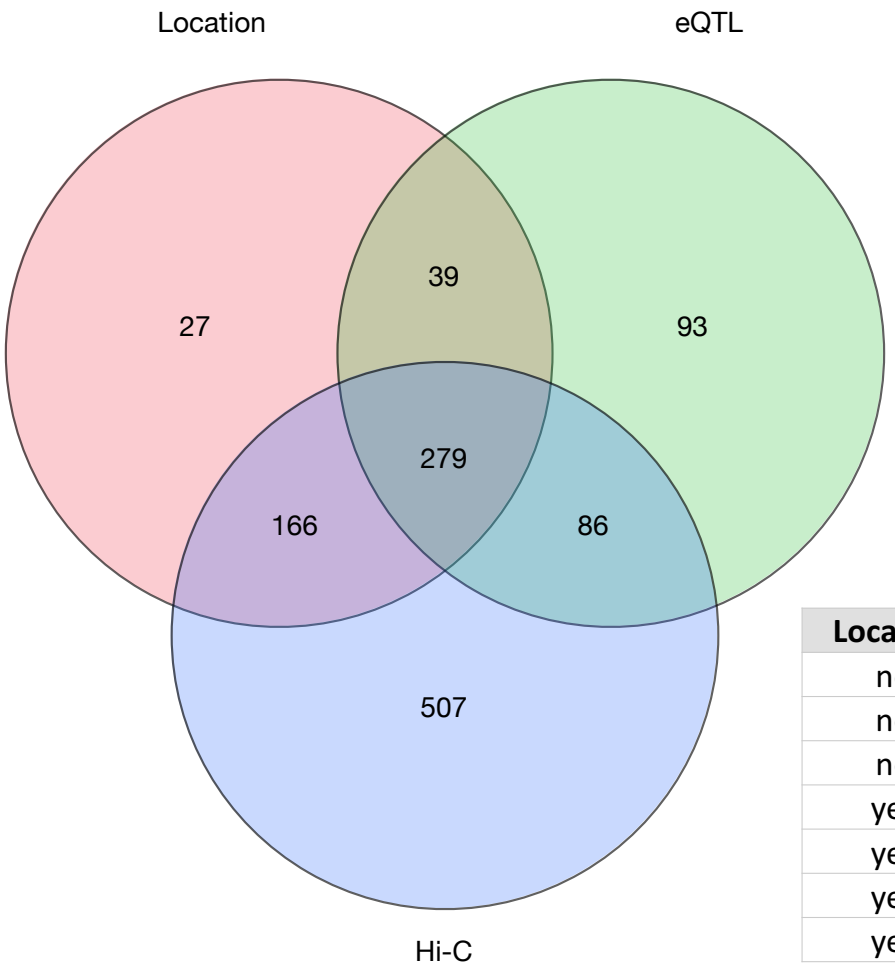

A compartment
FIRE
superFIRE
TAD boundary
Loop-all
Loop-REG
CTCF
Open chromatin
H3K27ac
Brain_p50
Brain_TSS+
Brain_Exon
Blood_p50
Liver_p50
Adult
Fetal
AO
LV
GM12878
IMR90
LI
clozuk
iq2018
edu2018
bmi2018
giantheight2014
ibd2015
choltotal
30
25
20
15
10
5
0

Figure 3



Location	eQTL	Hi-C	Genes	Percent
no	no	yes	507	42.4%
no	yes	no	93	7.8%
no	yes	yes	86	7.2%
yes	no	no	27	2.3%
yes	no	yes	166	13.9%
yes	yes	no	39	3.3%
yes	yes	yes	279	23.3%

Figure S1

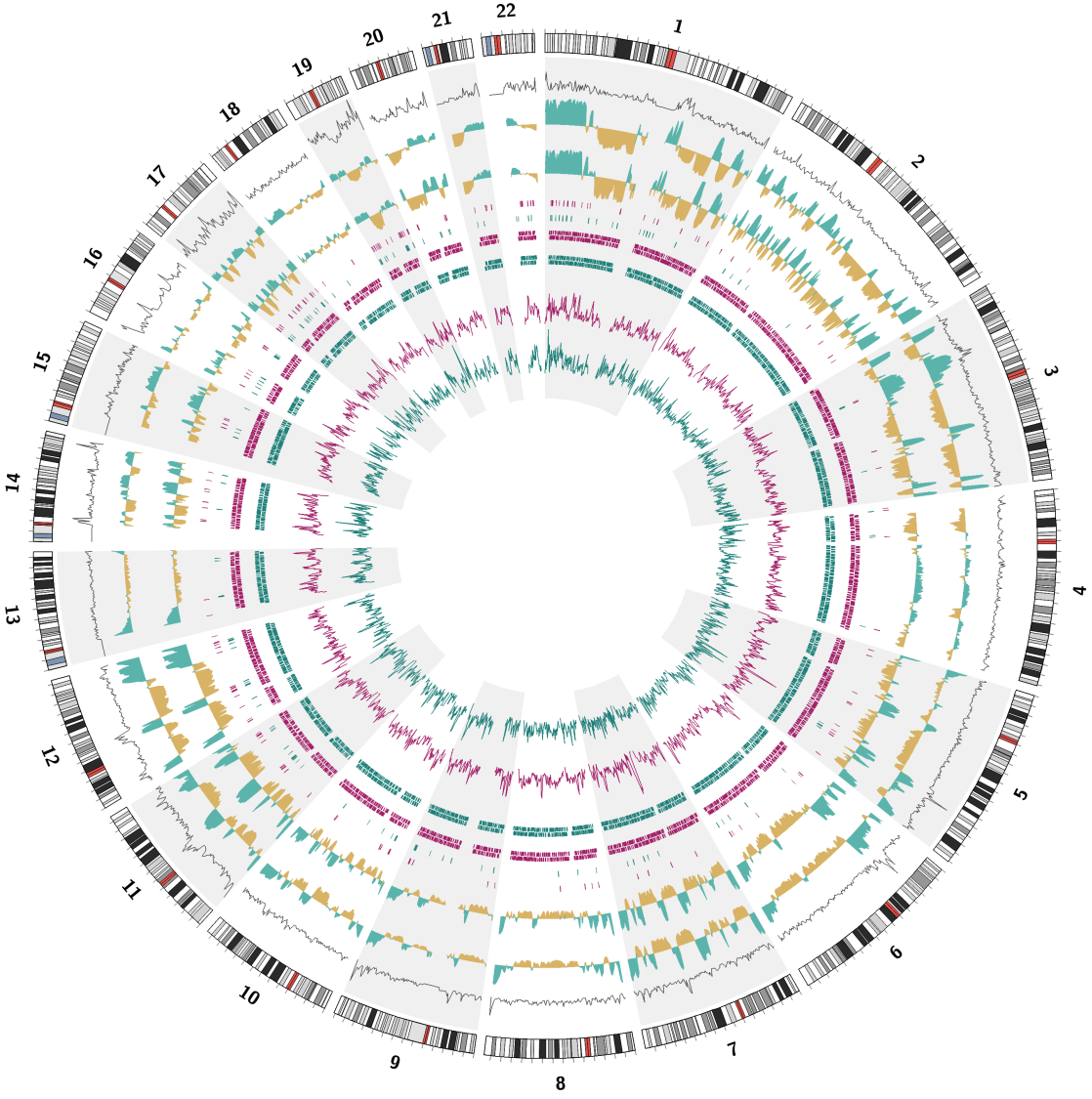


Figure S2a

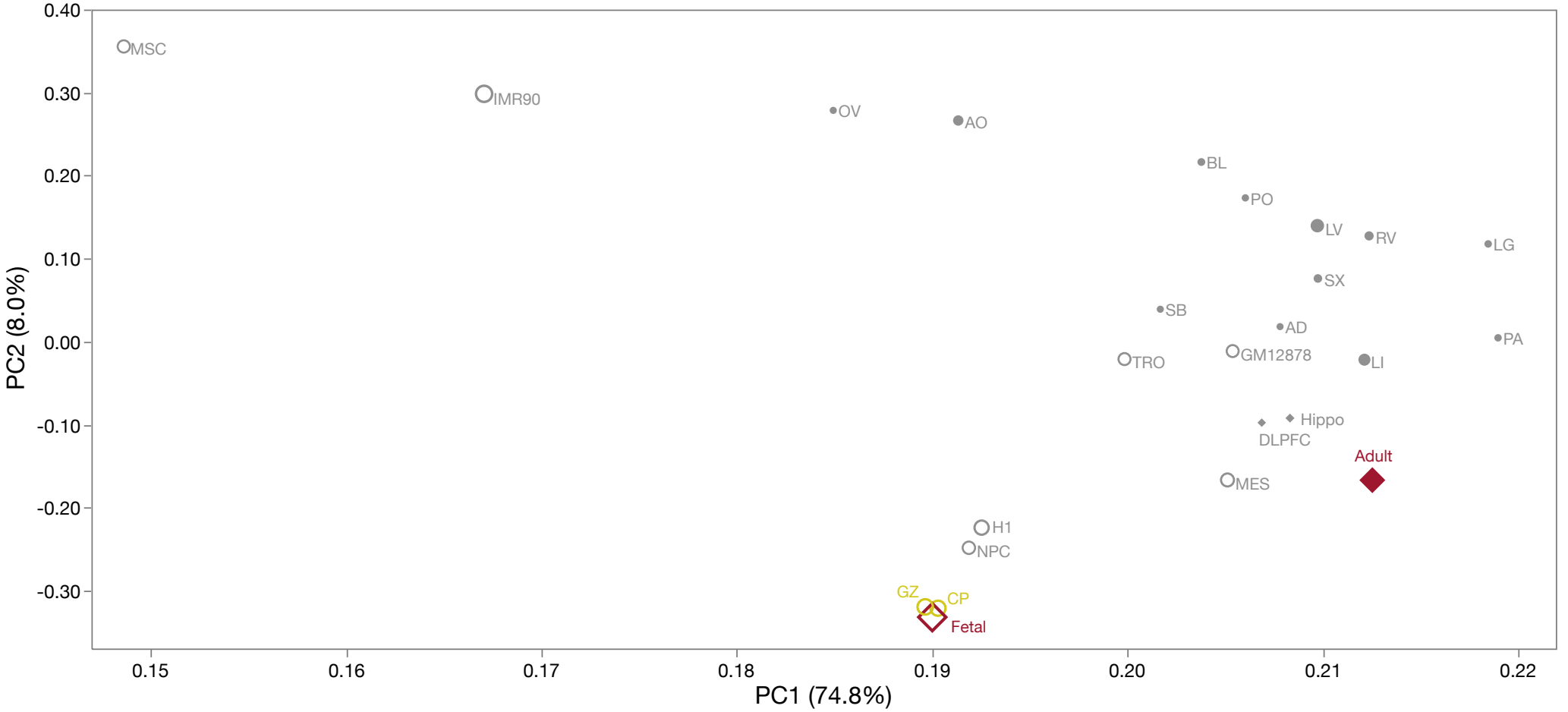

0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
PC2 (8.0%)
0.15
0.16
0.17
0.18
0.19
0.20
0.21
0.22
PC1 (74.8%)
MSC
IMR90
OV
AO
BL
PO
LV
RV
LG
SX
SB
AD
PA
TRO
GM12878
LI
Hippo
DLPFC
Adult
MES
H1
NPC
GZ
CP
Fetal

Figure S2b

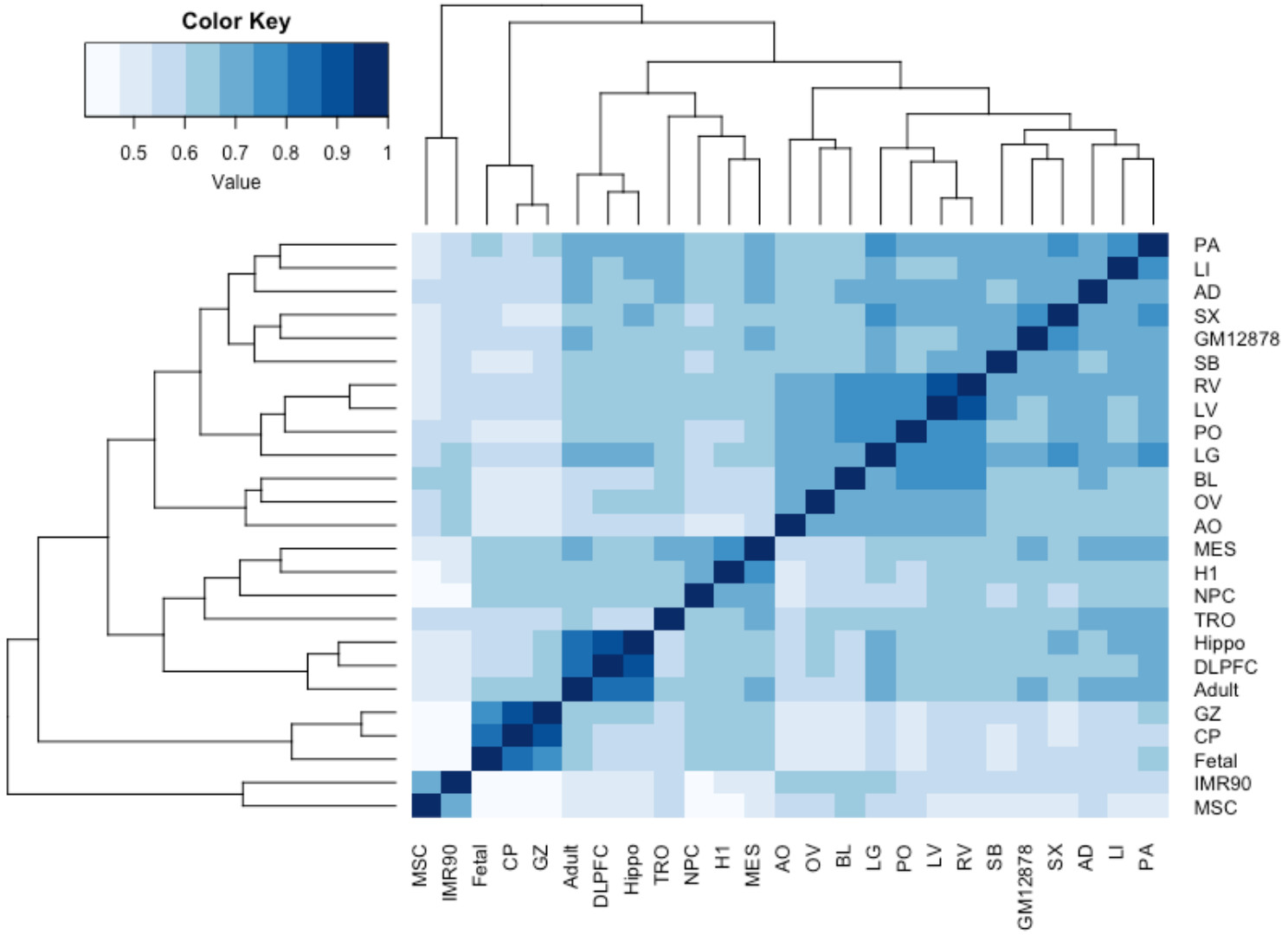

Color Key
0.5 0.6 0.7 0.8 0.9 1
Value
PA
LI
AD
SX
GM12878
SB
RV
LV
PO
LG
BL
OV
AO
MES
H1
NPC
TRO
Hippo
DLPFC
Adult
GZ
CP
Fetal
IMR90
MSC
MSC
IMR90
Fetal
CP
GZ
Adult
DLPFC
Hippo
TRO
NPC
H1
MES
AO
OV
BL
LG
PO
LV
RV
SB
GM12878
SX
AD
LI
PA

Figure S3a

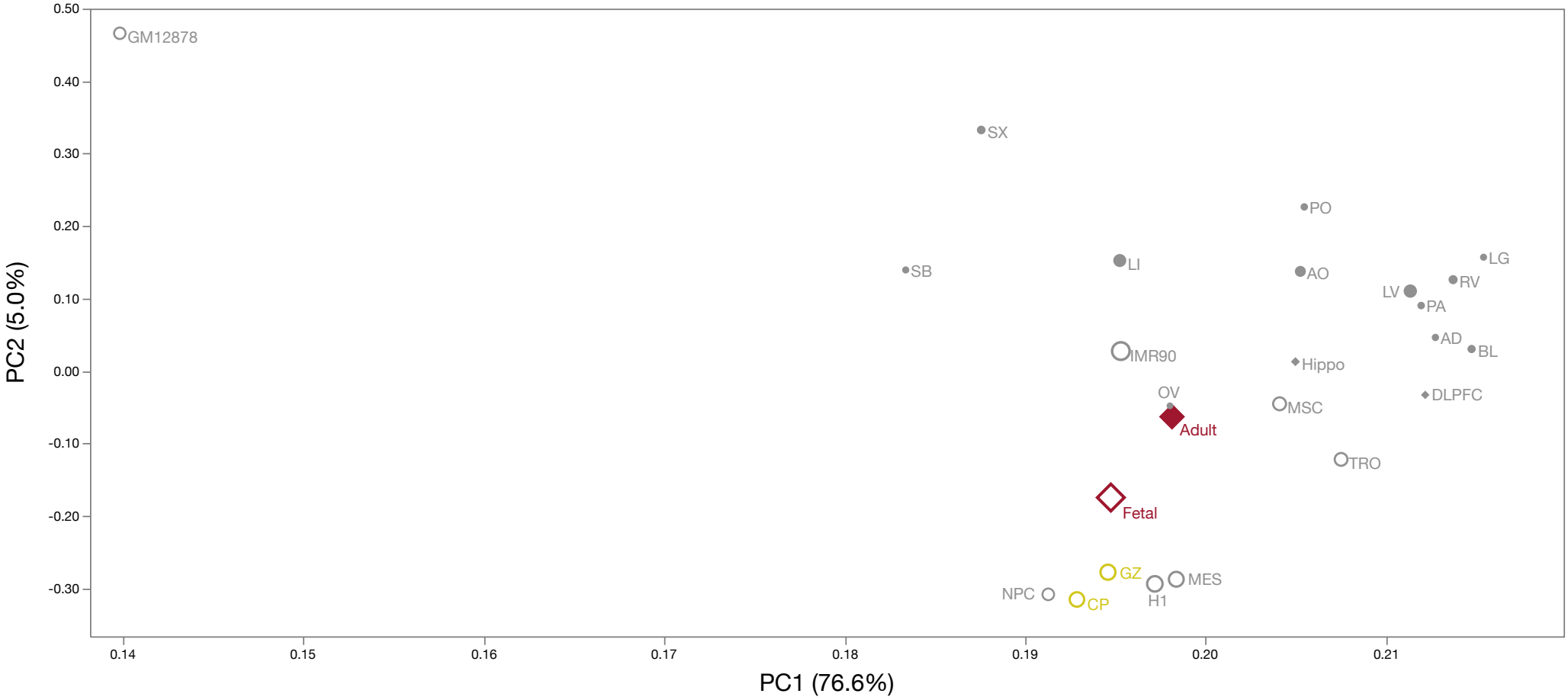

0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
PC2 (5.0%)
0.14
0.15
0.16
0.17
0.18
0.19
0.20
0.21
PC1 (76.6%)
GM12878
SX
PO
SB
LI
AO
LG
RV
LV
PA
AD
BL
IMR90
Hippo
OV
DLPFC
MSC
Adult
TRO
Fetal
GZ
MES
NPC
CP
H1

Figure S3b

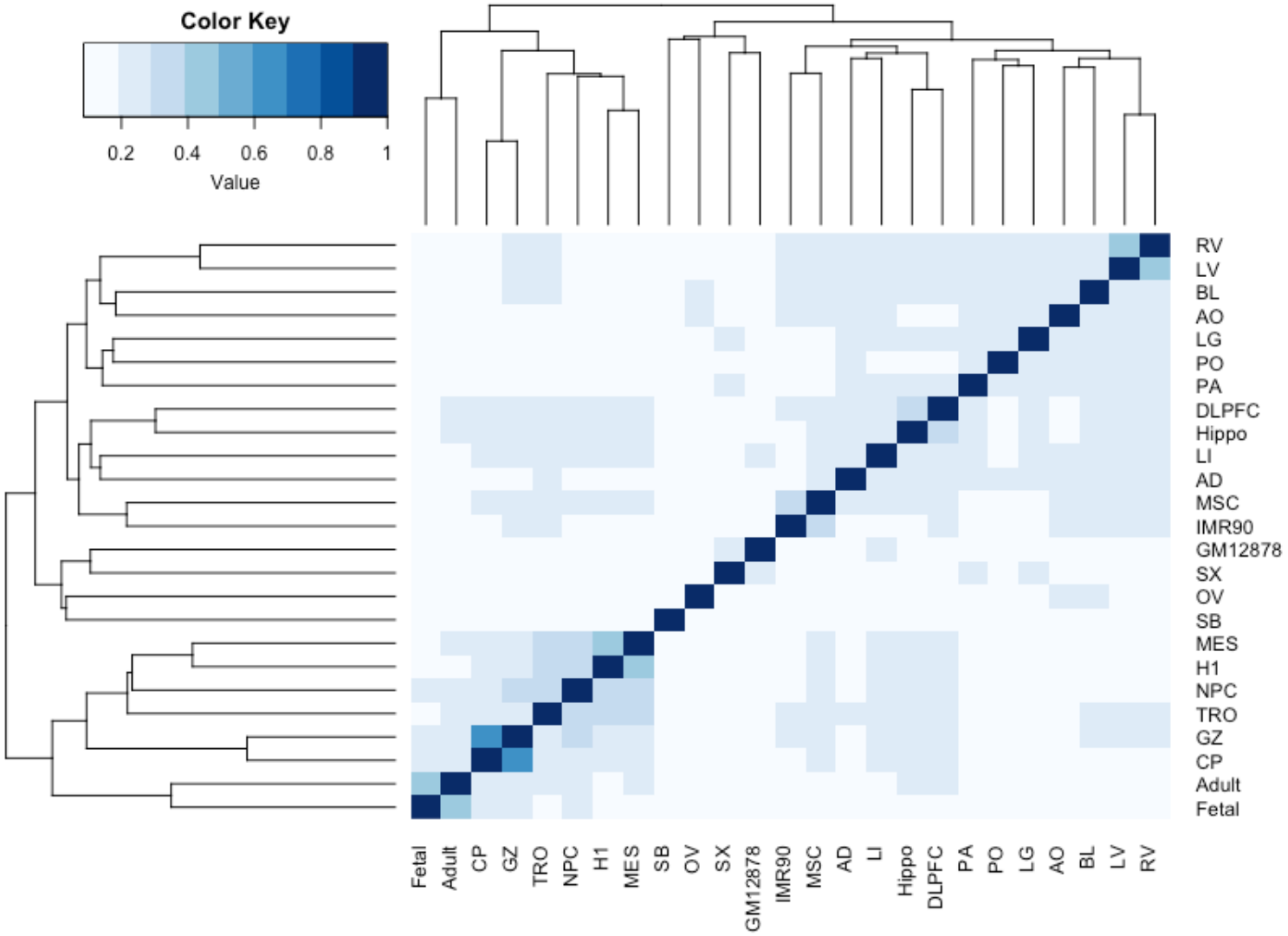

Color Key
0.2
0.4
0.6
0.8
1
Value
RV
LV
BL
AO
LG
PO
PA
DLPFC
Hippo
LI
AD
MSC
IMR90
GM12878
SX
OV
SB
MES
H1
NPC
TRO
GZ
CP
Adult
Fetal
Fetal
Adult
CP
GZ
TRO
NPC
H1
MES
SB
OV
SX
GM12878
IMR90
MSC
AD
LI
Hippo
DLPFC
PA
PO
LG
AO
BL
LV
RV

Figure S4a

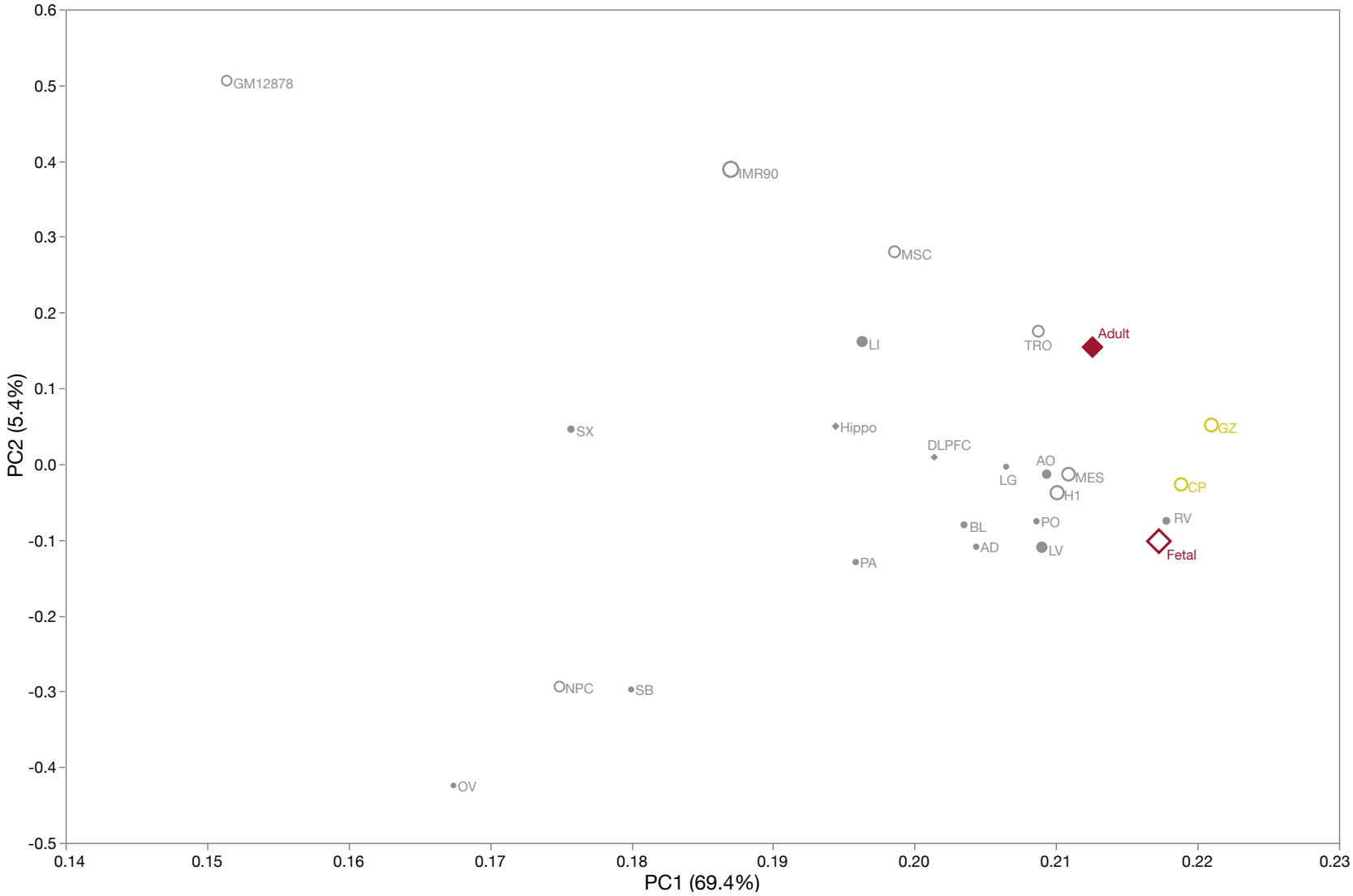

GM12878
IMR90
MSC
LI
TRO
Adult
SX
Hippo
GZ
DLPFC
AO
LG
MES
CP
H1
RV
BL
PO
AD
LV
Fetal
PA
NPC
SB
OV
PC2 (5.4%)
PC1 (69.4%)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
0.14
0.15
0.16
0.17
0.18
0.19
0.20
0.21
0.22
0.23

Figure S4b

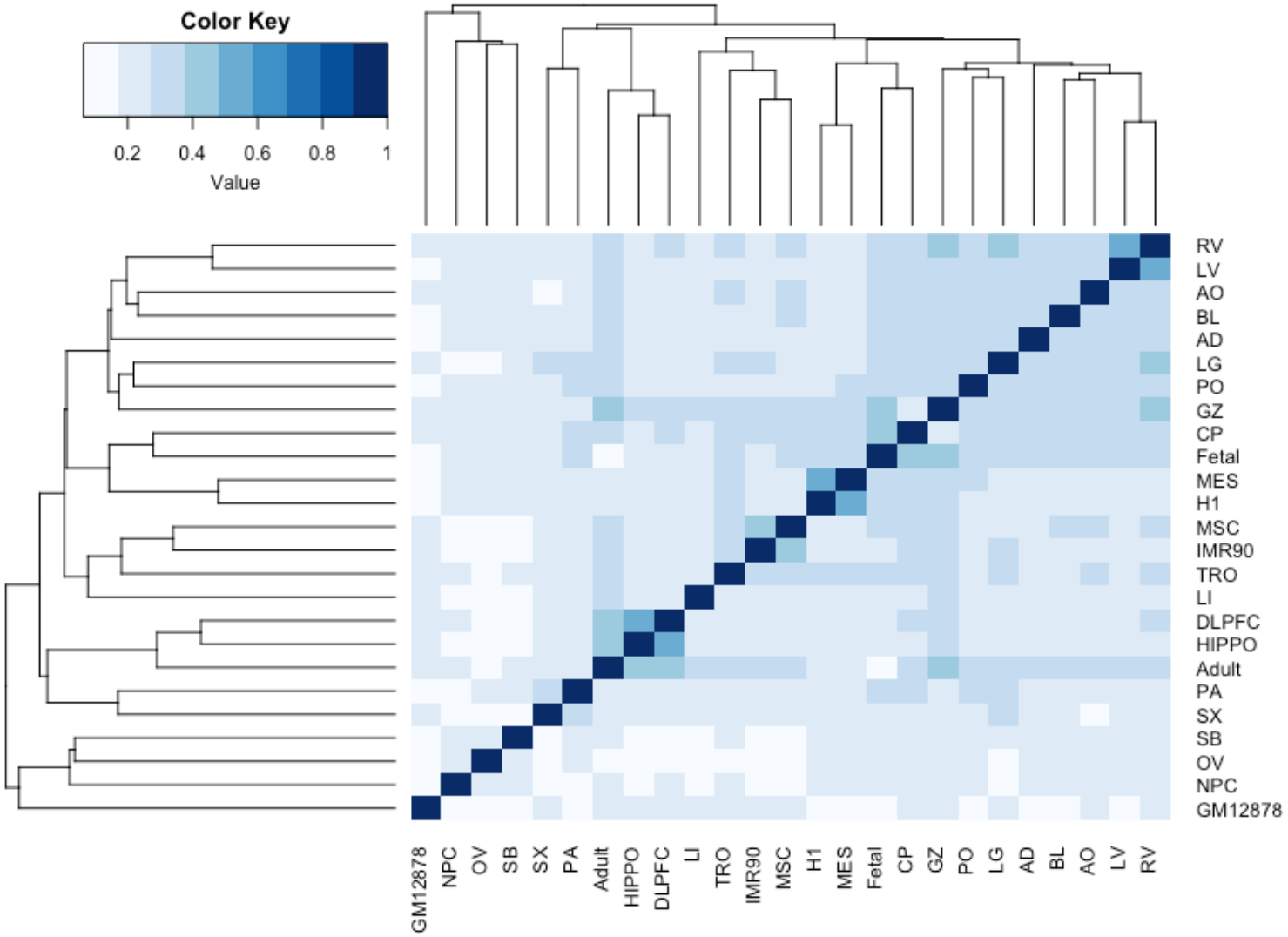

Color Key
0.2
0.4
0.6
0.8
1
Value
RV
LV
AO
BL
AD
LG
PO
GZ
CP
Fetal
MES
H1
MSC
IMR90
TRO
LI
DLPFC
HIPPO
Adult
PA
SX
SB
OV
NPC
GM12878
GM12878
NPC
OV
SB
SX
PA
Adult
HIPPO
DLPFC
LI
TRO
IMR90
MSC
H1
MES
Fetal
CP
GZ
PO
LG
AD
BL
AO
LV
RV

Figure S5

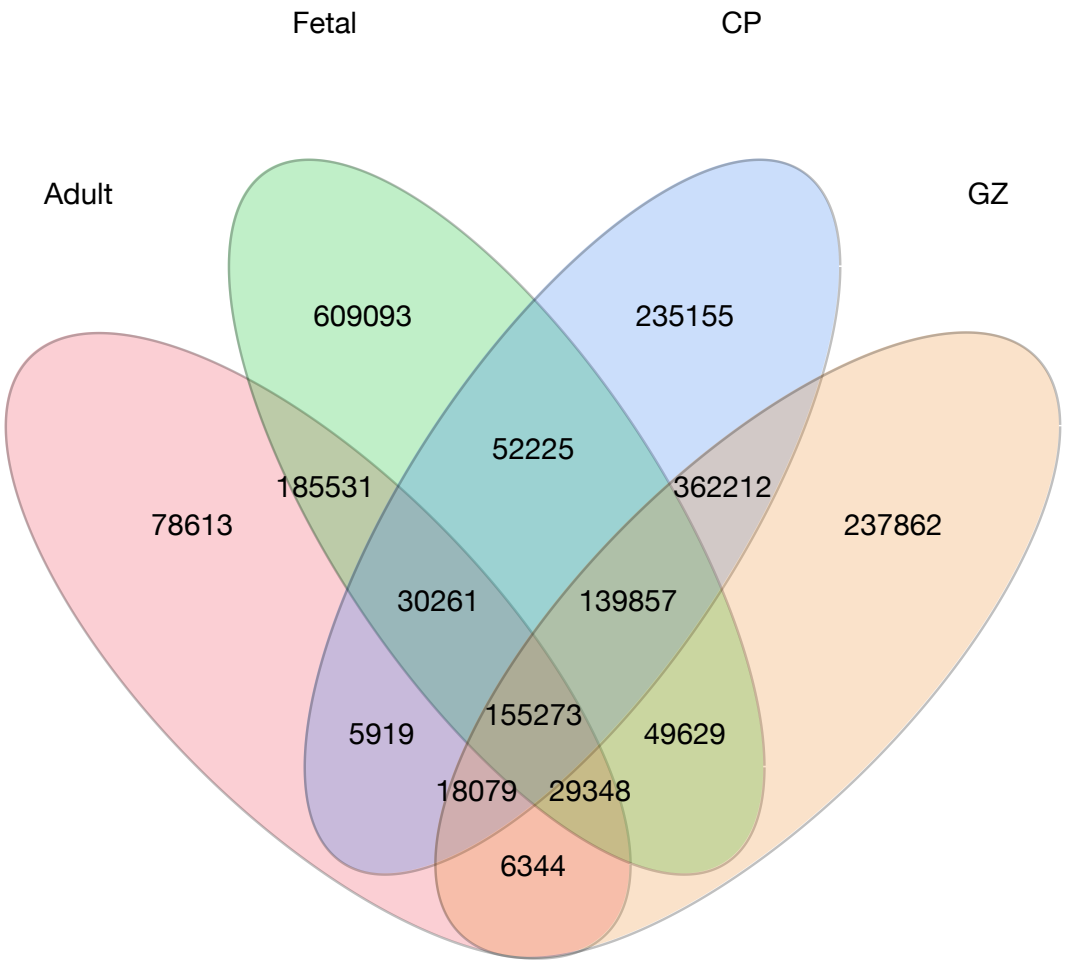

Fetal
CP
Adult
GZ
609093
235155
52225
185531
362212
78613
237862
30261
139857
155273
5919
49629
18079
29348
6344

Figure S6

Model 1: Hi-C data only

```
Call:
lm(formula = Diff.GeneExp ~ Diff.Loop + Diff.FIRE + Diff.IS +
    Diff.PC1, data = y)

Residuals:
     Min       1Q   Median       3Q      Max
-18.1610  -0.7991  -0.0167   0.6962  13.2941

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.001861   0.017243  -0.108  0.91405
Diff.Loop    0.007452   0.001460   5.103 3.39e-07 ***
Diff.FIRE    0.697242   0.099953   6.976 3.17e-12 ***
Diff.IS      0.770523   0.136819   5.632 1.82e-08 ***
Diff.PC1     0.014809   0.005377   2.754  0.00589 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.712 on 14951 degrees of freedom
Multiple R-squared:  0.01243,   Adjusted R-squared:  0.01216
F-statistic: 47.03 on 4 and 14951 DF,  p-value: < 2.2e-16
```

Model 2: ChIP-Seq data only

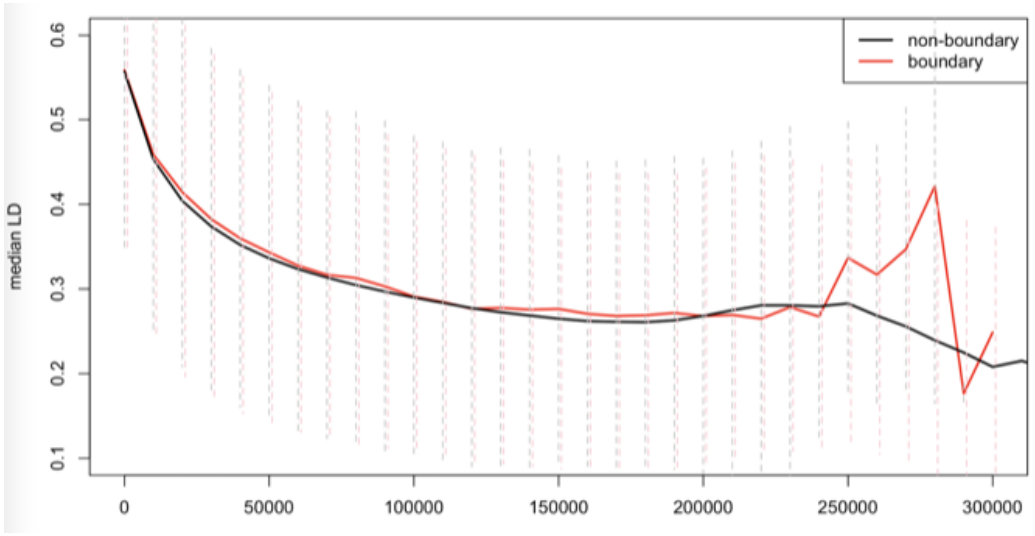
```
Call:
lm(formula = Diff.GeneExp ~ Diff.H3K27ac + Diff.H3K4me3, data = y)

Residuals:
     Min       1Q   Median       3Q      Max
-19.1258  -0.8037  -0.0011   0.7320  13.5769

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.07030    0.01422  -4.945 7.69e-07 ***
Diff.H3K27ac  0.62792    0.03660  17.158  < 2e-16 ***
Diff.H3K4me3  0.66888    0.04688  14.269  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.689 on 14953 degrees of freedom
Multiple R-squared:  0.03871,   Adjusted R-squared:  0.03858
F-statistic:   301 on 2 and 14953 DF,  p-value: < 2.2e-16
```

Model 3: Hi-C data + ChIP-Seq data

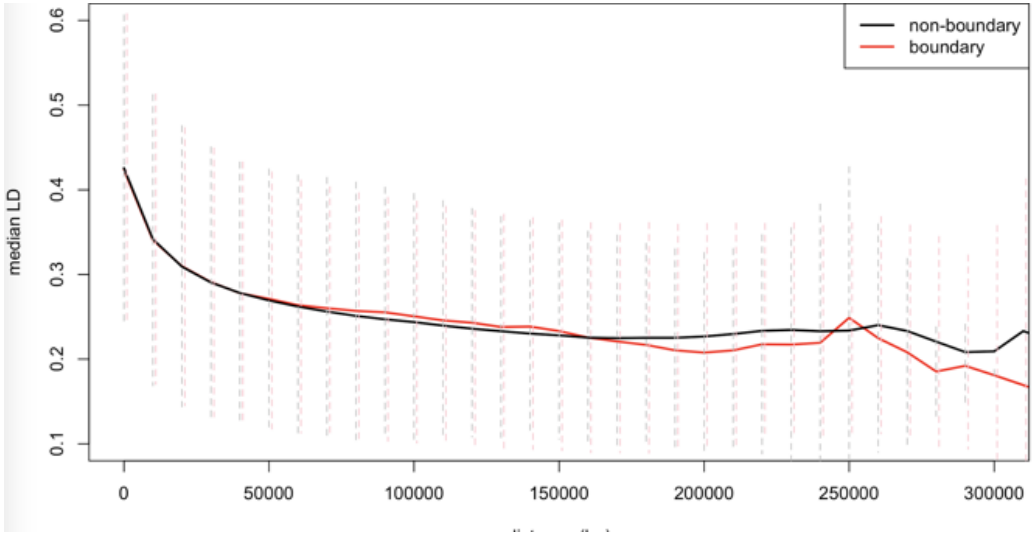
```
Call:
lm(formula = Diff.GeneExp ~ Diff.Loop + Diff.FIRE + Diff.IS +
    Diff.PC1 + Diff.H3K27ac + Diff.H3K4me3, data = y)

Residuals:
     Min       1Q   Median       3Q      Max
-18.2008  -0.8186  -0.0036   0.7430  13.3996

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.067217   0.017199  -3.908 9.34e-05 ***
Diff.Loop     0.007465   0.001435   5.203 1.98e-07 ***
Diff.FIRE     0.569230   0.098366   5.787 7.32e-09 ***
Diff.IS       0.621858   0.134558   4.621 3.84e-06 ***
Diff.PC1      0.009277   0.005290   1.754   0.0795 .
Diff.H3K27ac  0.581971   0.036697  15.859  < 2e-16 ***
Diff.H3K4me3  0.667712   0.046674  14.306  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.682 on 14949 degrees of freedom
Multiple R-squared:  0.04754,   Adjusted R-squared:  0.04716
F-statistic: 124.4 on 6 and 14949 DF,  p-value: < 2.2e-16
```

Figure S7

A


non-boundary
boundary
median LD
0.1
0.2
0.3
0.4
0.5
0.6
0
50000
100000
150000
200000
250000
300000

B


non-boundary
boundary
median LD
0.1
0.2
0.3
0.4
0.5
0.6
0
50000
100000
150000
200000
250000
300000