



# HHS Public Access

Author manuscript

*Nat Neurosci.* Author manuscript; available in PMC 2022 September 24.

Published in final edited form as:

*Nat Neurosci.* 2022 April ; 25(4): 474–483. doi:10.1038/s41593-022-01032-6.

## Chromatin domain alterations linked to 3D genome organization in a large cohort of schizophrenia and bipolar disorder brains

Kiran Girdhar<sup>1,3,5,†</sup>, Gabriel E. Hoffman<sup>1,3,5</sup>, Jaroslav Bendl<sup>1,2,3,4,5</sup>, Samir Rahman<sup>1,2,3,4,5</sup>, Pengfei Dong<sup>1,2,3,4,5</sup>, Will Liao<sup>8</sup>, Mads E. Hauberg<sup>1,2,3,4,20,21</sup>, Laura Sloofman<sup>7</sup>, Leanne Brown<sup>2</sup>, Olivia Devillers<sup>2</sup>, Bibi S. Kassim<sup>2</sup>, Jennifer R Wiseman<sup>2</sup>, Royce Park<sup>2</sup>, Elizabeth Zharovsky<sup>2</sup>, Rivky Jacobov<sup>2</sup>, Elie Flatow<sup>2</sup>, Alexey Kozlenkov<sup>4</sup>, Thomas Gilgenast<sup>17</sup>, Jessica S. Johnson<sup>1,3,5</sup>, Lizette Couto<sup>2</sup>, Mette A. Peters<sup>11</sup>, Jennifer E Phillips-Cremins<sup>17</sup>, Chang-Gyu Hahn<sup>15</sup>, Raquel E. Gur<sup>16</sup>, Carol A. Tamminga<sup>18</sup>, David A. Lewis<sup>19</sup>, Vahram Haroutunian<sup>2,4,6,13</sup>, PsychENCODE Consortium<sup>#</sup>, Stella Dracheva<sup>2,4,13</sup>, Barbara K. Lipska<sup>10</sup>, Stefano Marengo<sup>10</sup>, Marija Kundakovic<sup>2,4,6,9</sup>, John F. Fullard<sup>1,2,3,4,5</sup>, Yan Jiang<sup>2,4,6,14</sup>, Panos Roussos<sup>1,2,3,4,5,12,13,\*†</sup>, Shahram Akbarian<sup>2,4,6,\*†</sup>

<sup>1</sup>Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>2</sup>Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>3</sup>Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>4</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>5</sup>Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>6</sup>Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>7</sup>Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms> Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

<sup>†</sup>Correspondence and requests for materials should be addressed to P.R. or S.A. Correspondence to: Kiran.Girdhar@mssm.edu; Panagiotis.Roussos@mssm.edu; Shahram.Akbarian@mssm.edu.

<sup>#</sup>List of PsychENCODE consortium authors is provided in the appendix

\*Equally contributing senior authors.

Authors Contributions Statement

Wet lab work including tissue processing, sorting of nuclei and ChIP-seq and Hi-C library generation: Y.J., L.B., M.K., E.Z., R.J., J.R.W., R.P., B.S.K., L.C., O.D., S.R., J.F., E.F., A.K. Data processing and coordination: Y.J., M.K., M.A.P., J.S.J. Bioinformatics and computational genomics: K.G., G.E.H., J.B., S.R., T.G., J.P.-C., P.D., W.L., M.E.H., L.S., L.C. Provision of brain tissue and resources: C.A.T., S.M., B.K.L., D.A.L., V.H., C.-G. H., R.E.G., S.D., P.C. Conception of study and design: P.R., S.A., K.G., G.E.H., J.B. Writing of the paper: K.G., S.A., P.R.

Code availability:

All publicly available software utilized is noted in Methods. We have used decorate software to call CRDs <https://github.com/GabrielHoffman/decorate>.

**Additional information:** Supplementary information is available for this paper at <https://doi.org/10.7303/syn25710572>.

Competing Interests Statement

The authors declare no competing financial interests.

<sup>8</sup>New York Genome Centre, New York, NY 10013, USA

<sup>9</sup>Department of Biological Sciences, Fordham University, Bronx, NY 10458, USA

<sup>10</sup>Human Brain Collection Core, National Institute of Mental Health- IRP Bethesda, MD 20892-9663, USA

<sup>11</sup>Sage Bionetworks, Seattle, Washington 98121, USA

<sup>12</sup>Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

<sup>13</sup>Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, 10468, USA

<sup>14</sup>Institutes of Brain Science, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Fudan University, 200032, Shanghai, China

<sup>15</sup>Jefferson University, Philadelphia, PA 19107, USA

<sup>16</sup>The Karl and Linda Rickels Professor of Psychiatry, Univ of PA School of Medicine, Philadelphia, PA 19104, USA

<sup>17</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>18</sup>Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, TX 75390, USA

<sup>19</sup>Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213, USA

<sup>20</sup>iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark

<sup>21</sup>Department of Biomedicine, Aarhus University, Aarhus, Denmark

## Abstract

Chromosomal organization, scaling from the 147 base pair nucleosome to megabase-ranging domains encompassing multiple transcriptional units including heritability loci for psychiatric traits, remains largely unexplored in the human brain. Here, we construct promoter and enhancer enriched nucleosomal histone modification landscapes for adult prefrontal cortex (PFC) from H3-lysine 27 acetylation and H3-lysine 4 trimethylation profiles, generated from (n=739) 388 controls and 351 subjects diagnosed with schizophrenia (SCZ) or bipolar disorder (BD). We mapped thousands of cis-regulatory domains (CRDs), revealing fine-grained,  $10^4$ - $10^6$  bp chromosomal organization, firmly integrated into Hi-C topologically associating domain (TAD) stratification by open/repressive chromosomal environments and nuclear topography. Large clusters of hyperacetylated CRDs were enriched for SCZ heritability, with prominent representation of regulatory sequences governing fetal development and glutamatergic neuron signaling. Therefore, SCZ and BD brains show coordinated dysregulation of risk-associated regulatory sequences assembled into kilo- to megabase-scaling chromosomal domains.

## Introduction

Chromosomal organization scales from nucleosomes, or  $1.47 \times 10^2$  base pairs (bp) of DNA wrapped around a histone octamer, to functional and structural domains extending across  $10^3$ - $10^7$  bp, with highly interdependent regulation across scales. This includes transcription-associated nucleosomal histone modifications in fibroblasts and peripheral myeloid cells, including mono- and tri-methyl-H3K4 (H3K4me) and acetyl-H3K27 (H3K27ac), which are tightly linked to chromatin structures defined by local chromosomal conformations, including the megabase-scaling ‘self-folded’ topologically associating domains (TADs) and other features of 3D genome organization<sup>1</sup>. Whether or not such type of acetyl- and methyl-histone defined higher order chromatin exists in the human brain, including cell type-specific regulation and disease-associated alterations, remains unexplored. To date, virtually all conventional brain epigenomic maps present transcriptional histone marks (including H3K4me and H3K27ac) as isolated ‘peaks’ confined to short nucleosomal arrays, typically covering an average of 3.6-3.8 kb<sup>2,3</sup> in the human brain, with only a very small portion of peaks showing some degree of confluence by merging into super-enhancers important for cell-specific gene expression programs<sup>4</sup>.

Interestingly however, regulators nucleosomal histone modifications, including H3K4me3 and H3K27ac confer heritable risk for schizophrenia (SCZ) and related co-heritable traits, including bipolar disorder (BD) by genome-wide association and exome sequencing<sup>5,6,7</sup> and furthermore, in the adult human frontal lobe, SCZ and BD risk loci are enriched for active neuronal promoters and enhancers and other regulatory elements tagged by open chromatin-associated histone marks<sup>2,6-8</sup>. Unfortunately, representative genome-scale histone modification studies in diseased brain disease are lacking. It is not known whether changes in acetylation and methylation landscapes affect the general population of SCZ and BD subjects<sup>9</sup>, and whether such type of alteration could reveal broader changes in chromosomal organization beyond the classical analysis ‘peak-by-peak’ based analysis for nucleosomal histone modifications.

Here, we generated 739 ( $N_{H3K4me3}=230$ ,  $N_{H3K27ac}=260$  from neurons and  $N_{H3K27ac}=249$  from bulk tissue) ChIP-Seq (chromatin immunoprecipitation followed by deep sequencing) libraries from prefrontal cortex (PFC) of adult SCZ, BD and control brains. Using population-scale correlational analysis and cell type-specific chromosomal conformation mapping, we define acetylation and methylation landscapes by the coordinated regulation of sequentially arranged histone peaks constrained by local chromosomal conformations and nuclear topographies. We report widespread disease-associated alterations affecting the neuronal H3K27ac acetylome, but not the trimethyl H3K4 (H3K4me3) methylome. On a genome-wide scale, hundreds of kilo- to megabase-scale chromosomal domains are altered in disease, with converging alignments by genetic risk, cell type, developmental function, nuclear topography, and active vs. repressive chromosomal environments. Our findings, reproducible across two independent brain cohorts, identify higher order chromatin alterations representative of the broader population of SCZ and BD subjects, and link cognitive disease to altered organization of neuronal genomes in the prefrontal cortex.

## Results

### Acetyl-histone peaks show disease specific dysregulation

We first generated 490 ChIP-Seq genome-wide maps of H3K4me3 and H3K27ac from  $\sim 3\text{-}5 \times 10^5$  neuronal NeuN<sup>+</sup> nuclei/sample isolated from dorsolateral PFC via fluorescence activated nuclear sorting (FANS,) from 321 demographically matched SCZ and non-psychiatric control brains that are part of the CommonMind Consortium collection<sup>10-12</sup> (hereafter referred to as study-1). We then generated an additional set of H3K27ac ChIP-Seq libraries (N=249) prepared from unsorted nuclei extracted from bulk dorsolateral PFC tissue of SCZ, BD and control brains, contributed by the Human Brain Collection Core (HBCC) at the NIMH (hereafter referred to as study-2) (Figure 1A, Table 1, Table S1A–B).

Samples were processed with our in-house version of the encode ChIP-Seq pipeline<sup>2</sup>; H3K4me3 NeuN<sup>+</sup> peaks had a narrower genomic coverage (3.1%; mean peak width  $\sim 1425$  bp) compared to H3K27ac (genomic coverage 12.8% in NeuN<sup>+</sup> and 17.1% in bulk tissue; Figure 1A; mean peak width  $\sim 3364$  bp NeuN<sup>+</sup> and  $\sim 3583$  bp bulk tissue) (see QC metrics in Figure S1A; consensus peak sets in Tables S2A–C), with > 60% of H3K4me3 and 75% of H3K27ac peaks distributed among distal intergenic, exonic, intronic and UTR elements (Figure 1A). Importantly, each dataset showed high concordance (Jaccard similarity coefficients  $\sim 0.7$ ) to previously generated PFC NeuN<sup>+</sup> and bulk tissue H3K4me3 and H3K27ac datasets of brains not included in the present study (Figure S1B).

After various technical factors related to tissue processing and sequencing were regressed out (Methods, Figure S1C), we obtained three sets of normalized histone peak activity matrices (64,254 peaks  $\times$  230 H3K4me3 NeuN<sup>+</sup>, 114,136 peaks  $\times$  260 H3K27ac NeuN<sup>+</sup> from study-1 and 143,092 peaks  $\times$  249 H3K27ac Tissue from study-2; Table S3A–C). Furthermore, for each study-2 sample, cell type heterogeneity in bulk tissue was adjusted by estimating the proportion of oligodendrocytes, and glutamatergic and GABAergic neurons using cell type specific ChIP-Seq data from an independent reference set<sup>13</sup>.

We explored H3K4me3 NeuN<sup>+</sup> (mostly promoter-associated) ‘peak’-based epigenomic aberrations in SCZ study-1. Surprisingly, none of the 64,254 peaks (Table S3A) survived multiple testing corrections after differential (cases vs controls) analysis, indicating that this methylation mark is not consistently affected. Next, we evaluated H3K27ac (promoter and enhancer associated) peaks. Altogether, 11,471 of the 114,136 H3K27ac NeuN<sup>+</sup> peaks were dysregulated (FDR 5%) in SCZ study-1 (Table S3B), and similarly, 5,656/143,092 H3K27ac tissue peaks were significantly affected in SCZ study-2 (Figure 1B, Table S3C), with 559 dysregulated peaks present in both studies (Figure S2A). However, there was a significant correlation between cases-controls effect sizes of SCZ study-1 H3K27ac NeuN<sup>+</sup> at FDR 5% and SCZ study-2 H3K27ac Tissue peaks ( $n=9,951$  peaks, Spearman’s  $\rho=0.36$ ,  $P = 4.9 \times 10^{-295}$ ) (Methods, Figure 1C). Having shown that histone acetylation changes in SCZ PFC are broadly reproducible across independent brain collections (Table S1), we next combined the differential histone peak effects sizes and p-values from study-1 H3K27ac NeuN<sup>+</sup> and study-2 H3K27ac Tissue datasets (Methods), yielding a consensus set of 46,294 H3K27ac Meta NeuN<sup>+</sup> peaks each with 90% overlap of peak extension across the two studies. Of these, 6,219 peaks were dysregulated in SCZ (FDR 5%) (Figure 1D, Table S3D).

We applied a similar differential analysis workflow to determine BD specific epigenomic aberrations in study-2 H3K27ac Tissue and identified 1,809/143,092 dysregulated peaks (FDR 5%) (Table S3E), with 630 dysregulated peaks shared with H3K27ac Tissue SCZ and 158 shared with H3K27ac NeuN+ SCZ (Figure S2B). Furthermore, there was significant correlation ( $n=5,656$  peaks, Spearman's  $\rho=0.87$ ,  $P < .05$ ) of BD vs. controls effect sizes with SCZ vs. controls effect sizes at FDR 5% in H3K27ac Tissue within the study, and across studies ( $n=9,951$  peaks, Spearman's  $\rho=0.19$ ,  $P = 1.9 \times 10^{-78}$ ) with H3K27ac NeuN+ (Figure S3A), suggesting shared epigenomic dysfunction in these two common types of psychiatric disorders.

Indeed, gene set enrichment analysis of dysregulated peaks were consistent for immune responses across our SCZ and BD cohorts, and neuronal (including neuron development) signaling and synaptic plasticity pathways ranked top among gene ontologies in SCZ sensitive H3K27ac peaks (Figure S4). A representative example for H3K27ac peak-based alterations in our SCZ PFC datasets (study-1, study-2, meta) includes the 0.25Mb wide *SYNTAXIN 1A (STX1A)* psychiatric susceptibility locus, encoding a regulator of synaptic vesicle docking<sup>14</sup> (Figure 1E). Finally, we compared our H3K27ac NeuN+ (tissue) peaks in PFC showing >10% sequence overlap with PFC tissue open chromatin regions (OCRs)<sup>15</sup> generated from a cohort partially congruent with study-1. Using  $\pi 1$  statistics, the proportion of true positive SCZ sensitive H3K27ac peaks that overlapped with OCRs ranged from 27.3% (tissue) to 30.6% (NeuN+).

Because the majority of our diseased brains were exposed to antipsychotic drugs (APD) prior to death (Table S1A–B), we assessed the potential impact of medication, by studying the subset of  $N=116$  (36=Yes, 80=No) study-1 cases with documented exposure to typical (D2-like receptor antagonists) and  $N=117$  (52=Yes, 65=No) atypical/mixed receptor profile APD in the month prior to death. However, disease-associated H3K27ac changes showed almost null (atypical, Spearman's  $\rho=0.0006$ ,  $P = .0031$ ) or negative (typical, Spearman's  $\rho=-0.27$ ,  $P < 2.2 \times 10^{-16}$ ) correlation with APD. Therefore, medication is not a driver (atypical APD) or even anticorrelated (typical APD) with H3K27ac alterations in diseased PFC NeuN+ (Figure S5).

### Hyperacetylated peaks are enriched for SCZ risk variants

To better understand these disease-associated aberrations in PFC H3K27ac peaks and their link with directionality in acetylation, we stratified peak alterations into hyperacetylation “ SCZ $\uparrow$ ” and hypoacetylation “ SCZ $\downarrow$ ” based on  $\log_2$  fold change (cases/controls)  $>0$  and  $<0$ , respectively, and computed their enrichment for genetic variants associated with SCZ and related psychiatric traits using stratified LD score regression<sup>16</sup>. Interestingly, SCZ heritability coefficients were driven significantly by the group of hyper- but not hypo-acetylated peaks in all three of our SCZ case control comparisons, an effect particularly striking for the H3K27ac Meta NeuN+ dataset (Figure 2A). These changes were highly specific to psychiatric traits because non-psychiatric traits, such as height, or medical conditions including autoimmune and cardiac disease completely lacked association with our disease-associated PFC peaks (Figure 2B, Table S4).

To further assess enrichment of common variants of SCZ and other behavioral traits by the type of regulatory element, we stratified peak alterations into promoters ( $< \pm 3\text{Kb}$  from TSS) and enhancers ( $> \pm 3\text{Kb}$  from TSS). The coefficient of SCZ heritability was of higher magnitude in enhancers than promoters, an effect specific to hyperacetylated peaks (Figure S6A). Next, with genomic coverage of dysregulated BD peaks underpowered to run LDSc regression, we instead annotated differentially acetylated BD study-2 peaks to genes and checked for enrichment in SCZ/BD GWAS variants. We found a significant association of hyperacetylated “BD $\uparrow$ ” peaks with SCZ (but not BD) genetics (MAGMA  $P < 0.05$ ) (Figure S6B). Importantly, this effect was again significant for (SCZ) risk-associated enhancers in BD and BD $\uparrow$ , in contrast to dysregulated promoters in BD, BD $\uparrow$  and BD $\downarrow$  peaks (Figure S6C). Therefore, genetic risk for SCZ tracks genomic loci that are hyperacetylated in diseased SCZ and BD PFC, an effect consistent across all our disease cohorts.

### Histone peak correlations reveal chromosomal architecture

After identifying alterations in the activity of PFC histone peaks, we investigated the impact of disease on the structural organization of PFC chromatin by characterizing the modular architecture of coordinated histone peaks in the brain epigenome. We hypothesized that the structure of coordinated histone peaks could be particularly important in disease context. This hypothesis is plausible, given recent reports from peripheral cells with coordinated regulation of multiple cis-regulatory elements sequentially organized along the linear genome<sup>17</sup>. Additionally, we observed that pairwise correlation between PFC histone peaks within chromosomal loopings in Hi-C NeuN<sup>+</sup> from an independent set of PFCs (N=6; 3F/3M)<sup>18,19</sup> was substantially higher as compared to peaks of equivalent distance located outside of chromosomal loop contacts (Figure S8A) indicating the presence of correlation structure in histone peaks within Hi-C defined loops. Furthermore, studies on hundreds of lymphoblastoid and fibroblast cultures, leveraging ‘population-scale’ interindividual correlations between histone peaks, successfully uncovered coordinated regulated regions, or ‘cis-regulatory domains’ (CRDs), with spatial clustering of CRD histone peaks ranging from  $10^4$ - $10^6$  bp of linear genome and integrated into local chromosomal conformation landscapes<sup>20,21</sup>. Similar approaches have been applied to OCRs in Alzheimer’s postmortem brains<sup>18</sup>.

Here, we developed a systematic workflow (Methods, Figure S7) by combining the previously developed software *decorate*<sup>22</sup> with additional steps of statistical analyses to identify CRDs on our population-scale H3K27ac and H3K4me3 datasets encompassing 739 PFC ChIP-seq libraries. The pipeline applied adjacency constrained hierarchical clustering<sup>22,23</sup>, across each of our three ChIP-seq datasets (H3K4me3 NeuN+, H3K27ac NeuN+, H3K27ac Tissue) to identify sequentially aligned clusters of peaks as a strongly correlated structure (Methods, Figure S8B–C). Altogether, 39% (H3K4me3 NeuN+), 65% (H3K27ac NeuN+) and 68% (H3K27ac Tissue) of peaks assembled into 2,721, 6,389 and 8,239 CRDs respectively (Figure 3A, Table S5A–C), with H3K27ac (H3K4me3) CRDs encompassing on average  $\sim 11.7$  ( $\sim 9.3$ ) histone peaks (Figure S8D).

Comparison of study-1 and -2 H3K27ac CRDs showed higher similarity (Jaccard  $J$  of 0.39) as compared to study-1 H3K4me3 CRDs and H3K27ac CRDs ( $J = 0.22$ ) (Figure 3B).

Furthermore, 78-79% of H3K27ac CRD peaks were putative enhancers (i.e.  $> \pm 3\text{Kb}$  from TSS), in contrast to ~61% of H3K4me3 CRD peaks (Figure S8E). Promoters comprised the remaining peak populations and, in H3K27ac (H3K4me3) CRDs, were linked with an average of ~4 (1.6) enhancer peaks.

Next, we wanted to explore the potential link between our CRD and higher order chromosomal conformations, such as the topologically associating domains (TADs) computed from Hi-C libraries from ensembles of PFC NeuN<sup>+</sup> nuclei (Methods). Megabase-scaling TADs, and the smaller subTADs hierarchically nested into them, are thought to represent a type of conformation defined by dynamic chromosomal loop extrusions of individual chromatin fibers, constrained by strong boundary elements at TAD peripheries, and weaker in-TAD boundaries demarcating subTADs<sup>24,25</sup>.

Indeed, visual examination of Hi-C maps and H3K27ac CRD structure reveals CRDs located within TADs. A representative example (Figure 3C) shows the 2MB *GATB* (*Glutamyl-TRNA Amidotransferase Subunit B*) locus linked to cognitive traits and educational attainment<sup>26</sup>. Importantly, CRDs, with a median length of 120-168kb (Figure S8F), were significantly more likely to be inside TADs as compared to any random sequence of the same width as CRDs (Fisher's exact test: OR $>1$ , p value  $<.05$ ), an effect that was particularly pronounced (OR ~3-4) for acetylated CRDs (Figure S9). A detailed analysis revealed up to 77.4%(81.4%) and 94.3%(95.7%) acetylated CRDs were within subTADs and TADs respectively. Also, a substantial proportion of 81.5% (83.5%) subTADs and 59.1%(64.1%) TADs covered full H3K27ac CRD in PFC NeuN<sup>+</sup> (tissue) (Figure S10). Interestingly, acetylated CRDs consistently showed, both in H3K27ac NeuN<sup>+</sup> (study 1) and H3K37ac tissue (study 2), maximum density at the center of subTADs and TADs (Figure S11). In contrast, methylated CRDs were enriched at TAD boundaries (Figure S12), resonating with earlier reports on H3K4me3 enrichment at TAD boundaries<sup>1,27,28</sup>. Furthermore, both histone CRD and TAD borders were strongly enriched for occupancies of the structural protein, CTCF (Figure 3D), affirming that CRD modules are heavily constrained by the boundaries of their local TAD. Taken together, our studies reveal CRDs as structural units inserted into TADs of the adult PFC, with H3K27ac CRDs primarily representing enhancer-associated transcriptional domains localizing towards TAD centers while the topology of H3K4me3 CRDs indicates more a diverse function at TAD peripheries.

### Reproducible alterations of acetylated CRDs in diseased PFC

Having shown that individual histone peaks organize into CRDs as structural subunits within chromosomal domains, we then wanted to explore genome-wide CRD alterations in diseased brain. To this end, we applied a two step stage-wise statistical test (Methods, Figure S7) to identify dysregulated CRDs (CRD) and dysregulated histone peaks (CRD<sub>peaks</sub>) inside them. There were 1,010/6,389 (15.8%) significantly hyper- and 953 (14.9%) hypo-acetylated at FDR 5% in PFC NeuN<sup>+</sup> SCZ study-1, with proportions of CRDs somewhat lower in PFC tissue SCZ study-2 with 563 (6.8%) hyper- and 521 (6.3%) hypoacetylated (Figure 4A, Table S6). However, there was significant correlation between SCZ CRD, quantified as log<sub>2</sub>FC of peaks inside CRDs, from SCZ study-1 and -2 ( $\rho=0.28$ , p-value=4.4e-55, from 375(2138) and 367(2645) CRD(peaks) of study-1 and study-2 respectively, Figure S12A).

Similarly, we counted 203 (2.5%) hyper- and 251 (3.0%) hypoacetylated CRDs for PFC tissue BD (study 2), with significant correlation ( $\rho=0.69$ ,  $p\text{-value}=1.1\text{e-}266$ , from 126(1918) CRD(peaks), Figure S12B) between  $\log_2\text{FC}$  of peaks inside CRDs of SCZ and BD of study-2. Furthermore, ~10-12% of in-CRD H3K27ac peaks were significantly dysregulated (CRD Peaks) (Figure 4A and Figure S13).

To determine SCZ genetic variant enrichment in CRDs, we applied LDSc regression analysis and found higher SCZ heritability coefficients in hyperacetylated CRDs over hypoacetylated CRDs and over all CRDs, a highly consistent effect across study-1 and study-2 (Figure 4B). Since the genomic coverage of BD CRDs was insufficient for LDSc computation, we instead estimated enrichment for common risk variants in-CRD genes using MAGMA, and, like for the SCZ-sensitive CRDs, observed significantly higher coefficient of genetic association for SCZ (not BD) in hyperacetylated BD CRDs as compared to the total set of CRDs, and no genetic association in hypoacetylated CRDs (Figure S14). Interestingly, SCZ and BD CRD Peaks showed enrichment for neuronal signaling and metabolic functions (Table S7), with even higher heritability coefficients compared to all study-1 and study-2 CRDs (Figure 4B, Figure S14).

Next, to assess a potential link between local ‘peak’-level (from Figure 1) and CRD-level dysregulation, we first evaluated the odds of peaks to be in-CRD, with focus on disease-associated H3K27ac differences using glm (binomial generalized linear model) (Figure S15). Strikingly, in 3/3 datasets (study-1 PFC NeuN+ SCZ, study-2 PFC tissue SCZ and BD), disease-sensitive peaks were significantly more likely ( $\text{OR} > 1$ ,  $P < .05$ ) to fall inside CRDs as compared to outside of CRDs (Figure S15). Moreover, dysregulated peaks showed a strong tendency towards in-CRD clustering (poisson-based glm model,  $\text{OR}=1.65\text{-}3.48$ ,  $P < .05$ ) (Figure S15). Furthermore, differentially expressed genes across SCZ and controls from the CMC RNASeq cohort<sup>29</sup> were more likely to be in-CRDs than non-dysregulated CRDs (Poisson-based glm model,  $\text{OR}=1.3\text{-}1.5$ ,  $P < .05$ ) (Figure S15). These findings strongly suggest that in diseased SCZ PFC, alterations in histone acetylation manifest in a domain specific manner encompassing an array of peaks, potentially affecting transcription. A representative hyperacetylated CRD (Figure 4C) shows 0.4Mb of the chr. 5 GABA receptor *GABRA1/GABRG2* gene cluster and risk locus, encompassing ten H3K27ac peaks, including 1(6) hyperacetylated peaks from Figure 1 Peak (Figure 4A CRD Peaks).

### Dysregulated CRDs are aligned by chromosomal organization

Having shown that SCZ/BD PFC harbors alterations in structural domains, or dysregulated CRDs, we then asked whether disease-sensitive CRDs show evidence for coordinated (‘trans-CRD’) regulation in higher chromatin structure. We quantified each CRD as the mean of in-CRD H3K27ac peak levels followed by its correlation as (diseased) CRD contact matrix ( $m$  CRDs X  $m$  CRDs). Indeed, principal component analysis of the CRD contact matrix (Figure S16) revealed stratifications by the HiC-defined A and B compartments along with hyper- and hypoacetylation across component-1. This finding suggested that CRDs are aligned by directionality (hyper- vs. hypo-acetylation) and chromatin structure, including ‘A’ permissive vs ‘B’ repressed/condensed compartments.

For a more detailed analysis on CRD stratification, we applied the Bayesian information criterion (BIC) to identify the optimal number of clusters in every CRD contact matrix (K-means, see Methods). We identified  $k=3$  in SCZ PFC NeuN+ study-1 and  $k=2$  in SCZ, BD PFC tissue study-2 as optimal number of clusters (Figure S17A–C). We then created a resource of functional annotation of CRDs, including a) cell type-specific PFC reference sets including H3K27ac for glutamatergic projection neurons, GABAergic interneurons and oligodendrocytes (Table S8)<sup>13</sup>, b) NeuN+ Hi-C chromosomal A and B compartments<sup>18</sup> and c) developmental (fetal vs adult) stage, defined from the epigenetic trajectory of human cortical development<sup>30</sup> (Methods, see Figure S18A for the distribution of annotated CRDs).

Of note, two out of three clusters in the study-1 CRD contact matrix (Figure 5A) were primarily comprised of hyperacetylated CRDs representing GLU projection neurons, with chromosomal A:B compartmentalization further differentiating into cluster-1 A:B~1:2 and cluster-3 A:B~2:1. In striking contrast, cluster-2, overwhelmingly composed of hypoacetylated CRDs (91.2%), showed a 10-fold over-representation of interneuron-specific ‘GABAergic’ CRDs (Figure 5B). Furthermore, cluster 3 which showed the highest proportion of A-compartment, showed an overall higher magnitude of gene expression, as compared to clusters 1 and 2 (Figure S20).

Similarly, study-2 SCZ and BD specific CRDs contact matrices again showed stratification by hyper- vs. hypoacetylation and chromosomal compartmentalization as A vs. B (Figure S19A–B). Analysis of enrichment of SCZ GWAS variants in dysregulated CRDs (using LDSc) by annotated CRDs in each cluster revealed that coefficient of SCZ heritability, as determined by LDsc, was highest in magnitude for fetal as compared to adult annotated CRDs. This finding, consistent across study-1 (coef.= $2.6e-07 \pm 9.1e-08$ , p value= $2.6e-03$  in cluster-1; Figure 5C) and study-2 (coef.= $1.4e-07 \pm 9.9e-08$ , p value=.085 in cluster-2; Figure S21A), indicates the presence of neurodevelopmental signatures in SCZ, an effect that was particularly strong in PFC NeuN+. Furthermore, in PFC NeuN+ study-1, there was a strong cell specific effect with higher heritability for GLU (as compared to GABA) annotated CRDs in hyperacetylated clusters 1 & 3. However, due to the additional signal from non-neuronal cell types, study-2 SCZ PFC tissue lacked clear cell-specific heritability coefficients (Figure S21A), while in BD PFC tissue, genetic association of BD-risk genes (MAGMA) was observed specifically for GLU CRDs in hyperacetylated cluster-2; Figure S21B). Furthermore, SCZ heritability was present in both A/B compartment (study-1) or A only (study-2) (Figure 5C, Figure S21A).

### Nuclear topography of hyperacetylated CRDs

Having shown that histone CRDs comprise a type of structural subunit embedded within the chromosomal TADs, with stratification of disease-sensitive CRDs aligning with facilitative vs. repressive chromosomal environment and hyper- vs. hypoacetylation, we then explored nuclear topography and spatial 3D genome organization of the dysregulated CRDs. We utilized TAD coordinates of PFC NeuN+ Hi-C reference sets in *chrom3D*, a Monte Carlo-type algorithm for spherical genome modeling<sup>31,32</sup>. Indeed, pairwise Euclidean distances between TAD coordinates of PFC NeuN+ that overlapped with the genomic coordinates of A-compartment rich disease clusters defined by hyperacetylated CRDs,

revealed significantly higher TAD proximity and connectivity, when compared to chrom3D connectivity of all CRDs ( $p$  value  $< .05$ ) (Figure 6A). This 3D genome phenotype was remarkably consistent across all three disease cohorts, including cluster-3 from SCZ PFC NeuN+ study-1 and cluster-2 from SCZ and BD PFC Tissue study-2, respectively (Figure 6B). Therefore, diseased CRDs show distinct differences in spatial organization, including high chromosomal interactions between the TADs from hyperacetylated clusters.

## Discussion

The present study mapped active promoter- and enhancer-associated histone methylation and acetylation profiles in PFC of 563 brain donors, providing to date the largest histone modification dataset for SCZ and BD. Our histone peaks based analyses linked histone hyperacetylation to regulatory sequences for neuronal signaling and development, and to SCZ genetic risk, with hyperacetylated enhancers disproportionately enriched (compared to promoters) for risk-associated variants. These findings strongly suggest that epigenomic alterations in SCZ and BD brain are tracking the underlying genetic risk architecture.

Our finding that acetylated (but not methylated) chromatin shows disease-sensitive changes in PFC neurons is interesting given that the frontal lobe of SCZ and BD subjects is reportedly affected by alterations in histone deacetylase enzyme (HDAC) activity, according to *in vivo* imaging<sup>33,34</sup>, and postmortem expression<sup>35</sup> studies. Likewise, in the animal model, transgene-derived HDAC expression in PFC neurons alters cognition and behavior<sup>36,37</sup>, and furthermore, negative interference with PFC HDAC expression and activity exerts a therapeutic effect in psychosis<sup>38–40</sup>. Furthermore, according to the present study, while dysregulation of H3K27ac acetylation in (adult) PFC is representative for the broader population of subjects diagnosed with SCZ and BD, altered histone methylation, or at least H3K4me3, is not. However, regulation of H3K4 methylation is highly dynamic during the extended period of human PFC development and maturation<sup>41</sup>, and furthermore, according to animal systems modeling disrupted fetal development in SCZ, brain-specific alterations in H3K4me3 are transient and antecede the emergence of defective cognition and behavior in the adult<sup>42</sup>. Therefore, considering the neurodevelopmental etiology of common psychosis including SCZ and BD, it is possible that the PFC of our (adult) disease cases was transiently affected by H3K4 methylation changes during a much earlier (including prenatal) period, and could include many of the regulatory H3K4me3-tagged sequences that are associated with heritable risk<sup>2</sup> (Figure S6).

In the second part of this study, we constructed CRD chromosomal domains by estimating the inter-individual correlations between histone peaks. We show that acetylated and methylated CRDs are firmly embedded into TAD and subTAD (self-folded) domains of chromosomal conformations, but at much finer resolution. Histone CRDs, like the TADs and their nested subTADs, showed enrichment of CTCF structural protein at domain boundaries. This finding, together with our observation that H3K27ac-CRDs, comprised of arrays of active enhancers and promoters, are primarily located in the TAD center while H3K4me3-CRDs tend to locate towards the TAD periphery, underscores that CRDs are a type of chromosomal modular unit linked to transcriptional activity and organized on a smaller scale than the chromosomal conformation-defined TADs. The CRD concept could open new

avenues in neurogenomics, with the combination of Hi-C and CRD analyses, as presented here, offering novel insights into the finer grained architecture of functional organization of chromosomes.

Notably, according to our study, regulatory sequences affected by H3K27ac ‘peak’ alterations in diseased PFC from SCZ and BD subjects were much more likely to origin within CRDs, as compared to isolated peaks positioned outside of CRDs. This observation strongly speaks to the functional significance of acetylated CRDs for human cognition and behavior. Along these lines, our two-stage analysis, CRD Peaks in CRDs, confirmed that heritability risk for SCZ was highest for hyperacetylated H3K27ac peaks located in diseased CRDs. Because hyperacetylated CRDs were strongly enriched for regulatory sequences linked to excitatory (projection) neurons, and also harbored H3K27ac peaks with high coefficient of heritability in fetal annotated CRDs, such type of cluster-specific fingerprint could signal functional importance for many acetylated chromatin domains early in the disease process. These findings broadly resonate with the notion that enhancers and other cis-regulatory sequences of the fetal brain are disproportionally over-represented among the set of common risk variants linked to schizophrenia<sup>6,43,44</sup>. Therefore, it is plausible to hypothesize that a subset of hyperacetylated CRDs in diseased PFC neurons are vestiges of an early occurring neurodevelopmental disease process. Such type of epigenetic pathology in developing PFC could extend beyond the level of histone acetylation, given that alterations in DNA cytosine methylation profiles in adult SCZ PFC frequently encompass regulatory sequences defined by dynamic methylation drifts during the transition from the pre- to the postnatal period<sup>43</sup>.

Furthermore, the strong GLU neuron-specific fingerprint in our hyperacetylated CRD clusters is in excellent agreement with recent single nucleus-level transcriptome profilings, reporting up-regulation of glutamatergic neuron-specific expression modules in cortical layers of SCZ PFC<sup>8,45</sup> in addition to increased composite measures for glutamatergic transcripts in PFC of SCZ subjects<sup>46</sup>.

The final part of our analyses was focused on spatial genome organization of disease-associated CRDs, revealing, inside the virtual 3D sphere of a PFC neuron nucleus, an overall increased connectivity of TADs harboring hyperacetylated CRDs, an effect that was particular pronounced for CRD (clusters) stratified by a high proportion of the ‘A’ facilitative chromosomal compartment. This includes an overall higher inter-domain connectivity score in the *chrom3D* simulated nuclear sphere. Which types of molecular mechanisms could drive the nuclear topography of disease-relevant chromosomal domains, including the structural convergence of functionally inter-related hyperacetylated domains, as reported here? Interestingly, chromosomal contacts in brain and other tissues preferentially occur between loci targeted by the same transcription factors<sup>47</sup>, with convergence on intra- and inter-chromosomal hubs sharing a similar regulatory architecture among the interconnected enhancers<sup>48–51</sup>. In any case, based on the work presented here, we propose a longitudinal 3D or ‘4D nucleome’ model for the epigenomics of SCZ and BD. According to this model, H3K27ac peaks that became dysregulated as early as in the fetal period of (PFC) development could subsequently serve as ‘seed points’ ultimately spreading epigenomic dysregulation, specifically hyperacetylation, across an

entire functional chromosomal domain, or H3K27ac CRD. Therefore, SCZ and BD could ultimately be the manifestation of hyperacetylation events progressing from risk-associated histone peaks to their chromosomal domains and eventually further spread by nuclear topography.

From this work, we provide unique resources (1) of dysregulated histone peaks in a large brain cohort of SCZ and BD and (2) of genome-wide cis regulatory domains (CRDs) that delimit the highly connected histone peaks from unconnected histone peaks, and (3) a workflow to integrate epigenomics into 3D nuclear organization from *cis* to *trans* level of interactions between histone peaks to investigate the impact of disease at population-scale, and (4) domain-specific disease-sensitive peaks (CRD<sub>Peaks</sub>) as critical seed points impacting gene regulation. We expect these resources will provide a roadmap for future studies with even larger cohorts of SCZ and BD brains, aimed at gaining a deeper understanding exploring the emerging link between circuit-specific dysfunction and genome organization in SCZ and BD. These could include polygenic liabilities affecting distinct dimensions of psychosis including disorganization of thought process, delusions and hallucinations, and social withdrawal and other negative symptoms<sup>52</sup>.

## Methods

### Brains (postmortem):

All tissue donors of study-1 were from the Icahn School of Medicine at Mount Sinai (MSSM), University of Pennsylvania (PENN) and University of Pittsburgh (PITT) brain bank. All tissue donors of study-2 were from the Human Brain Collection Core (HBCC) at the national institute of mental health. Demographics of the brain cohort, toxicology and neuropathology reports are summarized in Table 1 and Table S1. No statistical methods were used to pre-determine sample sizes.

### ChIP-Seq library preparation and sequencing:

From the total set of 739 histone ChIP-seq datasets presented here, 28% (100 control and 109 SCZ cases) had been included in a recent PsychENCODE genomics reference paper for the adult human brain<sup>53</sup>, the remaining 530 ChIP-seq datasets had not been presented before.

Nuclei were extracted from approximately 300mg aliquots of frozen frontal (dorsolateral prefrontal) cortex tissue, immuno-tagged with Anti-NeuN-Alexa488 (Cat# MAB377X, EMD Millipore) antibody which robustly stains human cortical neuron nuclei<sup>54,55</sup> for subsequent fluorescence-activated nuclei sorting. Next, chromatin of sorted nuclei was digested with micrococcal nuclease and subsequently pulled down with anti-histone antibodies, followed by library preparation and sequencing. Two histone antibodies, anti-H3K4me3 (Cat# 9751BC, lot 7; Cell Signaling, Danvers, MA) and anti-H3K27ac (Cat# 39133, Lot# 01613007; Active Motif, Carlsbad, CA) were used for immunoprecipitation. Antibody specificity was tested using peptide binding assays and immunoblotting of nuclear extracts from human postmortem cortical tissue. A commercially available histone H3 peptide array (Cat# 16-667; Millipore) containing 46 peptides representing 46 different

histone H3 posttranslational modifications was used as previously described<sup>54</sup>. All procedures were performed as described in the recent PsychENCODE methods paper, providing a detailed description of the protocol<sup>54</sup>. For each cell-type specific ChIP-assay, a minimum of 400,000 sorted neuronal (NeuN+) nuclei was required as starting material. For selected gene promoters ChIP-PCR was conducted to validate cell-type specific peak profiles. Furthermore, quality controls for nuclei post-FACS included visual inspection under the microscope as described<sup>54</sup>. Of note, due to our stringent FACS gating criteria with maximized specificity (not sensitivity), 100% of sorted nuclei in the neuronal fraction showed green fluorescence confirming NeuN+ status, while 100% of sorted nuclei in the non-neuronal fraction only showed blue DAPI stain, confirming NeuN- status. Additional ChIP-seq studies were conducted with homogenized dorsolateral prefrontal cortex as input. To this end, frozen human postmortem brain tissue (approximately 20–200mg) was homogenized in lysis buffer and the total nuclei were purified. The nuclei solution was resuspended in 300ul of douncing buffer, treated with 2uL of micrococcal nuclease (0.2U/uL) for 5 minutes at 28 degrees Celsius, followed by 30uL of 500mM of EDTA to stop the reaction. After this initial procedure for nuclei preparation and digestion, the sample was processed in the same manner as described for the FACS sorted nuclei samples.

#### Randomization and blinding:

To avoid batch effects and other confounds, samples underwent repeated rounds of randomization, including (i) chromatin immunoprecipitation procedures and (ii) library preparation. Blinding was not relevant to this study; analysts were aware of data generation, processing and donor metadata.

#### Adapter sequences removal:

First the raw fastq files were corrected for adaptor pair end sequences using trimming tool called *Trimmomatic (v0.36)*<sup>56</sup> with the following settings: *ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36*.

#### Alignment, filtering, quality control and consolidation of BAM files:

Trimmed fastq files from each study were aligned to Hg38 (GRCh38) human genome using the Burrows-Wheeler Aligner (*BWA-0.7.8-r455*) method with default settings<sup>57</sup>. The output files were exported as BAM files. For quality control steps of BAM files, we implemented ENCODE pipeline workflow, which is as follows, 1) remove unmapped reads, mates and low quality mapping reads (mapq=30), 2) remove orphan reads and reads that were mapped to different chromosomes and 3) remove PCR duplicates using *picard (v2.2.4)* tool (<http://broadinstitute.github.io/picard>).

All BAM files from above step were tested for ENCODE quality control parameters for ChIP-Seq files: normalized strand coefficient (NSC>1.0) and relative strand coefficient (RSC>1) using *phantompeakqualtools (v2.0)*<sup>58</sup>. Figure S1A shows the frequency of NSC and RSC of samples from study-1 and study-2. We provide the NSC and RSC of each sample (Data availability).

After filtering out the BAM files based on ChIP-Seq qc parameters, we prepared the files for the next step which is consolidation of bam files separately for each dataset. The objective was to subsample each ChIP-Seq library to a fixed number of mapped reads and consolidate the subsampled libraries into one file. To obtain fixed number mapped reads for subsampling of bam files, we took minimum of number of mapped reads from each study; H3K4me3 NeuN+=12M, H3K27ac NeuN+=22M and H3K27ac Tissue=23M. We obtained (median) ~30, ~60 & ~59 million of mapped paired-end reads (2 x 75bp) for H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue respectively (Figure S1A). A similar procedure was followed to create a consolidated input-control file for NeuN+ study-1 and tissue study-2

### Mislabelling and contamination of samples check:

For samples mismatch and contamination check we used *QTLtools (v1.3) mbv*<sup>59</sup> (Match BAM to VCF) option. *MBV* takes as input a VCF file containing the genotype data for study-1 and study-2 samples and a mapped BAM file from the above section (Alignment, filtering and consolidation of BAM files). We did this step using the merged vcf file of genotypes of study-1 and study-2 separately. None of our samples were mismatched or contaminated. We provide a summary of *QTLtools (v1.3) mbv*<sup>59</sup> results of all samples (Data availability).

### Peak Calling:

Narrow peak regions were called on a consolidated file of H3K4me3 histone mark dataset using *macs2 (v2.2.6)*<sup>60</sup> with Poisson p-value = 0.01 with --keep-dup all --nomodel --extsize = 150. Similarly, broad peak regions were called on study-1 and study-2 consolidated files of H3K27ac histone mark datasets using *macs2 (v2.2.6)*<sup>60</sup> with P-value cutoff = .01, --extsize = 150. We used a NeuN+ consolidated input control and a tissue consolidated control file separately as control inputs for peak calling on each study. All peaks were filtered from blacklisted<sup>61</sup> region peaks for downstream analysis.

### Quantification of ChIP-Seq signal:

ChIP-Seq signal was quantified for every sample and every consensus peak obtained from the above section using *featureCounts (v1.5.0)* software<sup>62</sup>. The objective is to count the number of reads overlapping the genomic coordinates of peaks. This step results into a matrix of  $m_{peaks} \times n_{samples}$  66,163 peaks X 230 H3K4me3 NeuN+, 124,054 peaks X 260 H3K27ac NeuN+ and 207,866 peaks X 249 H3K27ac Tissue, Tables S2A–C

From these matrices, peaks with the low expression were filtered out using cpm of histone peaks >1 in at least 10% of samples as a threshold resulting into 64,254 peaks X 230 H3K4me3 NeuN+, 114,136 peaks X 260 H3K27ac NeuN+ and 143,092 peaks X 249 H3K27ac Tissue, Tables S3. Next, the read counts were corrected for library size using the trimmed mean of M-values (TMM) method from *edgeR* library<sup>63</sup> and converted into the voom-normalized matrices.

### Estimation of proportion of cell types in H3K27ac Tissue:

To account for cell type heterogeneity in H3K27ac Tissue samples, we estimate the proportion of glutamatergic, gabaergic and oligodendrocytes using *dtangle (v2.0.9)*

software<sup>64</sup>. Each tissue was modeled as a mixture of glutamatergic, gabaergic and oligodendrocytes. The reference samples and peaks of glutamatergic, gabaergic and oligodendrocytes were created using the previously published H3K27ac dataset from the PFC brain region<sup>13</sup>. We provide a vector of % of cell types for each sample in the metadata table (Data availability).

### Covariates model selection:

To estimate the technical and biological noise at sample level, we employ a 2-step approach.

1) We first identify the number of principal components using principal component analysis (PCA) method on the normalized read counts to identify the number of components that had variance of at least 1% of variance in the data. For each dataset, we take the correlation of all technical and biological covariates with the identified principal components and shortlisted the ones with FDR<20%

2) BIC: To identify the optimal number of covariates to have a good average model of histone peaks expression, we apply a Bayesian information criterion (BIC) approach<sup>65</sup> which introduces a penalty term for the number of parameters in the model. We start with “Diagnosis+Sex” as a base model and test all covariates one by one identified in the PCA step. Selection criterion of a covariate in the model is at least 5% of peaks should have  $(BIC_{\text{Diagnosis+Gender+Covariate}} - BIC_{\text{Diagnosis+Gender}})$  per histone peak  $\geq 2$

Other covariates are added sequentially in this model until they fail to meet the criterion of BIC threshold. Following are the covariates that were used to correct the voom-normalized matrices for each study.

$$\text{H3K4me3 NeuN}^+ = \text{Sex} + \text{Diagnosis} + \text{GC content} + \text{FRIP} + \text{Magnetic Beads}$$

$$\text{H3K27ac NeuN}^+ = \text{Sex} + \text{Diagnosis} + \text{GC content} + \text{FRIP} + \text{Age}$$

$$\text{H3K27ac Tissue} = \text{Sex} + \text{Diagnosis} + \text{Neurons proportion} + \text{GC content} + \text{FRIP} + \text{FRIP}^2 + \text{Age}$$

FRIP: Fraction of Reads In Peaks

Figure S1C shows the distribution of variance explained by each covariate in the ChIP-Seq peaks activity matrix of each study. For a complete list of covariates, see the metadata table in the Data availability section.

### Annotating ChIP-seq histone peaks regions:

Genes and genomic Context: The Ensembl 95 genes were used for all analyses in this paper. To annotate the genomic region of a histone peak as TSS, exon, 5' UTR, 3' UTR, intronic or intergenic, we used *ChIPSeeker* (v.1.18.0)<sup>66</sup>. The transcript database used for the annotation

is “TxDb.Hsapiens.UCSC.hg38.knownGene”. We used a threshold of  $\pm 3$  kb distance from TSS of a gene for promoter annotation. Figure 1A shows the distribution of peaks annotated to categories 1) promoters, 2) introns 3) distal intergenic and 4) exon and UTRs using the hg38 transcript database imported using *ChIPSeeker* package.

### Overlap with previously published datasets:

We calculated the Jaccard index to measure the concordance of histone peaks in study-1 and study-2 with existing datasets of REP<sup>67</sup> and EpiMap<sup>2</sup>. Jaccard index is measured as the intersection of base pairs divided by union of base pairs. Figure S1B shows the pairwise similarity of datasets REP, EpiMap and study-1,2.

### Peaks analysis

**Disease differential analysis:** To identify SCZ and BD sensitive peaks, we performed differential analysis on covariates corrected (Covariates model selection section) matrices from H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue using *limma (v4.1)*<sup>68</sup> pipeline. Table S3 provides differential analysis results from the above mentioned studies. Figure S2A shows the overlap of SCZ-sensitive H3K27ac NeuN+ peaks that overlap at least one base pair with H3K27ac Tissue peaks whereas Figure S2B shows the overlap of SCZ and BD sensitive peaks in H3K27ac Tissue.

**Meta analysis of H3K27ac NeuN+ and H3K27ac Tissue:** Next, we combined the differential analysis results from H3K27ac NeuN+ and H3K27ac Tissue to obtain the consensus peaksets using fixed effect analysis<sup>69</sup>. We first created the consensus peakset by taking the set of H3K27ac NeuN+ peaks that had at least 90% overlap of its width with H3K27ac NeuN+ peaks. Then, we take the differential analysis table of overlapping peaksets of both NeuN+ and tissue to run fixed effect analysis using *rma* function from the R *metafor* package (*v2.0*)<sup>69</sup>. Figure 1D shows the proportion of differential peaks and the volcano table of *rma* analysis.

**Pathway analysis of histone peaks:** To interpret the disease specific signatures in dysregulated H3K27ac NeuN+, Meta NeuN+ and Tissue peaks, we used the GREAT approach to assign peaks to genes. We examined the biological function of nearby genes for these non-overlapping peak regions using the Genomic Regions Enrichment of Annotations Tool (GREAT)<sup>70</sup>. The settings for GREAT are as follows: proximal 5.0 kb upstream, 5.0 kb downstream and plus Distal: up to 100 kb. Figure S4 shows the pathway enrichment of SCZ dysregulated peaks from H3K27ac NeuN, Meta NeuN+ and Tissue and BD dysregulated peaks from H3K27ac Tissue.

**Antipsychotics differential analysis:** To estimate the variance explained by antipsychotic treatment in SCZ and BD sensitive peaks, we performed differential analysis on covariates corrected (Covariates model selection section) matrices from H3K27ac NeuN+ using *limma (v4.1)*<sup>68</sup> pipeline. We had information on antipsychotics typical (e.g. Haldol, chlorpromazine) from 116 SCZ (36=Yes, 80=No) and antipsychotics-atypical (e.g. risperidone, clozapine) from 117 SCZ (52=Yes, 65=No) patients. We conducted differential analysis in expression matrix from 116(117) SCZ patients across a) antipsychotics typical

(yes) vs. no antipsychotics typical and b) antipsychotics atypical (yes) vs. no antipsychotics typical. Table S3 provides the results from differential analysis.

**LDscore enrichment analysis:** To estimate the enrichment of brain and non-brain related GWAS in all identified histone peaks and disease sensitive peaks from H3K27ac NeuN+, Meta NeuN+ and Tissue we used *LD-score partitioned heritability (v.1.0.0)*<sup>16</sup>. Figure 2, Table S5 and Figure S6 show the LDscore enrichments of SCZ sensitive peaks from H3K27ac NeuN+, Meta NeuN+ and Tissue.

For the traits, we used the European only version of the summary statistics when available. As a consequence, all GWAS results were based on individuals of European ancestry. The broad MHC-region (hg19:chr6:25-35MB) was excluded due to its extensive and complex LD structure, but, otherwise, default parameters were used for the algorithm. We ran LD-score- analyses only with sets of histone peaks covering 0.05% or more of the human genome.

**MAGMA association trait analysis:** Owing to the genomic coverage BD associated peaks < 0.05%, we used Multimarker Analysis of GenoMic Annotation (MAGMA)<sup>71</sup>, version 1.06b, to measure the association with schizophrenia risk peaks.

### Cis-regulatory domains (CRD)

**Genome wide CRD calling:** We identify cis-regulatory domains (CRD) separately on H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue by leveraging the inter-individual correlations of samples. Here we discuss in detail the stepwise workflow of CRD calling and identification of disease specific CRDs as shown in Figure S7.

**Removal of low correlation structure:** We first corrected for global effects of covariates to retain the correlation structure using PEER (probabilistic estimation of expression residuals) residualization<sup>72</sup> of histone peaks normalized expression from each study. A total of 18 PEER-corrected matrices  $m_{peak\_PEER\_i} \times n_{samples}$  ( $i = \{1, 5, 10, 15, 20, 25\}$ ) were produced (6 PEER-corrected H3K4me3 NeuN+, 6 PEER-corrected H3K27ac NeuN+, and 6 H3K27ac Tissue). CRDs were called on 18 matrices individually using the following R functions from the *decorate (v1.0.14)*<sup>22</sup> package.

```
CRDlist = runOrderedClusteringGenome (mpeak_PEERj × nsamples, peaks, method.corr = "Spearman")
```

```
CRDClusters = createClusters (CRDlist, method = "meanClusterSize", meanClusterSize = c(10, 25, 50, 80, 100))
```

```
CRDScore = scoreClusters(CRDlist, CRDClusters)
```

The output from above mentioned commands were 18 *CRDScore* objects (6 for each study datasets) containing a table of histone peaks assigned to CRDs, their mean correlation, and

lead eigen factor (LEF). LEF of a CRD is a fraction of variance explained by the first eigenvalue of the correlation matrix  $[m \times m]$  of histone peaks located within a CRD. Larger LEF values (i.e. >10%) can be interpreted as strongly correlated peaks whereas smaller values correspond to weaker correlations of peaks located within a CRD. Filtering out the CRDs with weaker correlations is an important step because it substantially reduces the burden of multiple testing in differential CRD analysis.

**CRD filtering and merging:** To filter out CRDs with weaker correlations, histone peak positions were shuffled per chromosome for all samples to create permuted matrices  $m_{Permutation\_j\_peaks\_PEER\_i} \times n_{samples}$  (where  $i = \{1, 5, 10, 15, 20, 25\}$  and  $j = 1-10$ ). A total of 180 PEER-corrected matrices  $m_{peak\_PEER\_i} \times n_{samples}$  ( $i = \{1, 5, 10, 15, 20, 25\}$ ) were produced (10 permutations  $\times$  6 PEER-corrected matrices  $\times$  3 datasets;  $m_{Permutation\_j\_peaks\_PEER\_i} \times n_{samples}$ ). CRDs were called on 180 matrices individually using the following R functions from the *decorate* package.

CRDs calling on 180 matrices (10 permutations  $\times$  6 PEER-corrected matrices  $\times$  3 datasets;  $m_{Permutation\_j\_peaks\_PEER\_i} \times n_{samples}$ ) on permuted matrices followed the same workflow as explained. Lastly,  $LEF_{cutoff}$  was obtained as vectors were combined from 10 *CRDScore* lists obtained from CRD calling on  $m_{Permutation\_j\_OCR\_PEER\_i} \times n_{samples}$  ( $i = 1-10$ )

$$LEF_{permuted} = [LEF_1, LEF_2, LEF_3, \dots, LEF_{10}], \text{ where } LEF_i = N \times 1$$

$$LEF_{cutoff} = Pr(LEF_{permuted} = \text{threshold}), \text{ where } \text{threshold} = 1 - 0.10$$

Final table of CRDs was obtained by keeping all CRDs with  $LEF_{measured} > LEF_{cutoff}$  in *CRDScore* list of  $m_{peaks\_PEER\_i} \times n_{samples}$ . Figure S8B shows an example of the distribution of  $LEF_{measured}$  and  $LEF_{permuted}$ .

Next, overlapping CRDs of different sizes were merged to obtain discrete CRDs for downstream analysis. To decide the optimal number of PEER factors, we measured the  $LEF_{cutoff}$  of CRDs called on the input matrix histone peaks matrix residualized by various numbers of PEER factors; we tested  $\{1, 5, 10, 15, 20, 25\}$  PEER factors (Figure S8C). The number of peaks within the CRDs are shown in Figure S8D, while the final lists of coordinates of CRDs of study-1 H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue are provided in Table S5.

**In-silico biological validation of CRDs:** To validate 3D interactions captured by CRDs with Hi-C dataset, we used CTCF ChIP-seq peak list from ENCODE human neural cells <sup>73</sup> (Data and materials availability). We quantified the density of CTCF sites in 200 bins (each bin size equals to 1kb) around CRD boundaries (Figure 3E). In order to quantify how many *in-silico* 3D interactions captured by CRDs are within the 3D interactions measured as Topologically Associated Domains (TADs) from PFC NeuN+ Hi-C experiments <sup>18</sup>, we measured the number of CRDs overlapping with PFC NeuN+ Hi-C TADs stratified by number of TADs ( $N = \{0, 1, 2, 3, >=4\}$ ). Next, we measured how many PFC NeuN+ Hi-C

TADs are within the CRDs stratified by the number of CRDs ( $N=\{0,1,2,3,\geq 4\}$ ). We measured the correlation of histone peaks inside the Hi-C loops and outside the Hi-C loops to show that peaks inside the Hi-C loops have more correlation than the peaks outside the Hi-C loops (Figure S10).

**Glutamatergic, GABAergic, and oligodendrocyte ChIP-seq data:** The data were obtained from <sup>13</sup>. H3K27ac peaks were called with DFilter <sup>74</sup> using the following parameters: “-f=bam -pe -ks=60 -lpval=4”. For each cell type, H3K27ac peak lists for replicate samples were then overlapped using a custom R script, and peaks which were present in at least half of replicates were preserved for further analysis (peak numbers: 44,519 in GABA neurons, 46,580 in Glu neurons, 45,963 in OLIG cells). Peaks detected in Glu, GABA and OLIG cells were further overlapped using the bedtools package to obtain Glu-specific (19,697), GABA-specific (16,297), and OLIG-specific (26,975) peaks (Table S8).

**Annotation of CRDs:** CRDs annotation to 1) fetal/adult category as fetal and adult <sup>30</sup>, 2) cell types as glutamatergic (GLU), GABAergic (GABA) and oligodendrocytes (OLIG) from (Table S8) <sup>13,75</sup>, 3) active compartment as A and outside A compartment and 4) inactive compartment as B and outside B compartment using the PFC NeuN+ HiC data <sup>18</sup>. Every CRD was assigned to a specific category, if the fraction of peaks coverage in a CRD in a given assay matches the testing dataset (using the data resources as explained above) and is significantly different from the fraction of peaks coverage in all other CRDs as a background dataset using Fisher’s Exact Test at Pvalue < .05.

For fetal/adult annotation (1), we ran the annotation test to assess whether a given CRD in an assay is enriched for H3K27ac fetal specific peaks vs. all other CRDs. From this test, all CRD with OR>1 and <1 are annotated as Fetal and Adult respectively whereas all non significant CRDs are annotated as N.S. fetal.

For cell types annotation (2) in study-1 from neuronal assay, we ran the annotation test for GLU and GABA only. All non significant CRDs from this test are annotated as N.S. GABA/GLU. In study-2 H3K27ac tissue, we first assessed whether a given CRDs has enrichment in oligodendrocytes (OLIG) or GABAergic/glutamatergic cell types. CRDs that were significantly enriched for oligodendrocytes are annotated as OLIG and not significant as N.S. OLIG. Next, we take CRDs that show gabaergic/glutamatergic cell types and assess their enrichment in gabaergic vs glutamatergic cell type as explained above. Overall, we obtained five categories here, 1) GABA, 2) GLU 3) N.S. GABA/GLU 4) OLIG and 5) N.S. OLIG.

For chromosomal environment annotation (3), a fraction of full CRD coverage was used for the annotation test instead of using coverage of peaks within the CRD and all non significant CRDs from this test are annotated as N.S. A/B. Figure S18 shows the final counts of CRDs annotated to each category of cell type, development and compartments and Table S6 lists the annotation of CRDs from three disease groups.

**Disease specific CRDs:** In this section, we show how differential analysis of CRDs was done and how we identified the relationship between structure and activity of CRDs.

**Differential CRD analysis:** We apply a two-stage testing procedure using stageR package<sup>76</sup> that identifies significant CRDs using aggregated CRDs-level P values in the screening stage and in the confirmation stage, individual hypotheses are assessed to determine dysregulated peaks for CRDs that pass the screening stage. Hence, it has the advantage of improving the resolution in stage II by providing dysregulated peaks in dysregulated CRDs from stage I.

For stage 1, we used the peak differential analysis table as input (Table S3) and aggregated p values using the equation 1. Here we show calculation of P-value and  $\log_2(\text{fold change})$  for one CRD ( $CRD_x$ ) that is linked to  $k$  Peaks.

$CRD_x = [Peak_1, Peak_2, Peak_3, \dots, Peak_k]$ , where  $CRD_x$  has  $k$  histone peaks

$$M = \min\{P \text{ value}_{Peak = i}; i = 1, 2, 3, \dots, k\}$$

$$PValue_{CRD_x} = 1 - (1 - M^k)$$

$$\log_2 FC_{CRD_x} = \frac{1}{k} \sum_{i=1}^k \log_2 FC (Peak_i)$$

Table S6 is the final differential analysis table of SCZ H3K27ac NeuN+, SCZ and BD H3K27ac Tissue.

**Model fitting and hypothesis testing:** To assess the link between the disease associated peaks from “disease differential peak analysis” and CRDs, we fit the logistics regression to predict the status of peak inside or outside CRD using T statistics from H3K27ac “Disease differential analysis” section. We ran this regression in R using the equations below:

$$glm(\text{peak inside CRD} \sim \text{Peaks Tstat}^2, \text{data} = \text{df}, \text{family} = \text{binomial})$$

$$\text{Peaks inside CRD} = 0/1\{\text{peaks outside/inside CRD}\}$$

$$\text{Peaks T stat}^2 = \text{square of T statistics from differential table of H3K27ac peaks}$$

$$\text{df} = \text{cbind}(\text{Peaks Inside CRD}, \text{Peaks T stat}^2)\{\text{data frame}\}$$

To test if differential peaks tend to be clustered inside dysregulated CRDs we applied poisson regression in which the predictor variable is the number of differential peaks from “Disease differential analysis” inside dysregulated CRDs from “*Differential CRD analysis*” accounting for the number of peaks inside CRDs as an offset in the equation.

$$glm(N_{\Delta peaks} \sim offset(log log(N_{peaks in CRD})) + N_{\Delta CRD}, data = df, family = poisson(link = log))$$

$$N_{\Delta peaks} = counts of differential peaks in every CRD \{m \times 1\}$$

$$N_{peaks in CRD} = number of peaks inside CRD \{m \times 1\}$$

$$N_{\Delta CRD} = 0/1 for non dysregulated/dysregulated CRD \{m \times 1\}$$

$$df = cbind(N_{\Delta peaks}, N_{peaks in CRD}, N_{\Delta CRD}) \{data frame\}$$

Next, we tested if differential genes tend to be clustered inside dysregulated CRDs to impact the gene regulation. To do this, we first annotated CRDs to genes by taking the genes that reside inside CRDs. After that, we applied Poisson regression in which the predictor variable is the number of differential genes using the differential analysis table of CMC RNA-Seq cohort. The R functions used are explained below.

$$glm(N_{\Delta genes} \sim offset(log log(N_{genes in CRD})) + N_{\Delta CRD}, data = df, family = poisson(link = log))$$

$$N_{\Delta genes} = counts of differential genes in every CRD \{m \times 1\}$$

$$N_{genes in CRD} = No of genes in every CRD \{m \times 1\}$$

$$N_{\Delta CRD} = 0/1 for non dysregulated/dysregulated CRD \{m \times 1\}$$

$$df = cbind(N_{\Delta genes}, N_{genes in CRD}, N_{\Delta CRD}) \{data frame\}$$

**CRD Contact Matrix:** Next we quantified the expression of CRD for  $m$  CRDs and  $n$  samples as CRD contact matrix by taking the mean of peaks that are within the CRD per sample as shown in the equation below.

$$CRD \text{ expression matrix} = \sum_{i=1}^k CPM_{peaks}$$

$$CRD \text{ contact matrix} = \text{corr}(CRD \text{ expression matrix} \times CRD \text{ expression matrix})$$

We applied K-means clustering<sup>77</sup> on the disease sensitive CRD contact matrix and evaluated the optimal number of clusters using the equation below.

$$BIC = D + \log(n) * m * k$$

$$D = \text{sum of squares within clusters}; n = \text{number of CRDs}, m = \text{number of CRDs}$$

$$k = \text{number of clusters in kmeans function}$$

Figure S17 shows the BIC value from kmeans clustering as a function of  $k(1:10)$  for CRD contact matrix SCZ sensitive H3K27ac NeuN+, H3K27ac Tissue and BD sensitive H3K27ac Tissue.

**LDscore enrichment analysis of CRDs:** To estimate the enrichment of brain and non-brain related GWAS in all identified CRD and disease sensitive CRD, we tested the genomic regions of histone peaks within the CRDs from H3K27ac NeuN+, Meta NeuN+ and Tissue and applied *LD-score partitioned heritability (v.1.0.0)*<sup>16</sup> as explained in LDscore enrichment analysis of peaks section. Figure 5 and Figure S21 show the LDscore enrichments of SCZ and BD sensitive CRDs from H3K27ac NeuN+ and H3K27ac Tissue.

**Modeling chromatin conformation in 3D:** Hi-C data from PFC NeuN+ was used to infer chromatin conformation structure in 3D. We used bulk Hi-C data from PFC NeuN+ from four adults. Primary processing was performed with the HiC-Pro pipeline<sup>78</sup> at 50kb and 1Mb resolution. In order to improve the accuracy of 3D modeling, we combined data from different donors for the PFC NeuN+ to increase sequencing depths. Contact matrices produced by HiC-Pro were converted to cooler format using HiCEXplorer<sup>79</sup>, balanced using the cooler suite of tools<sup>80</sup>, excluding the ENCODE v3 blacklisted regions (<https://www.encodeproject.org/files/ENCFF356LFX/>) from balancing with the `cooler balance --blacklist` parameter. Topologically associated domains (TADs) were called at 50kb using the `diamond-insulation` algorithm implemented in the cooltools suite (<https://cooltools.readthedocs.io/en/latest/>). Hi-C contact matrices and TAD calls were preprocessed

to `gtrack` files as input to Chrom3D, as previously described<sup>31,32</sup>. For more details on HiC data generation and processing see methods section on HiC<sup>18</sup>. We restricted our analysis to diploid autosomal interactions, 50kb for intrachromosomal and 1Mb for interchromosomal. Chrom3D was run with a nucleus radius of 5.0 for 2E6 iterations, '--radius 5.0 --iterations 2000000'. XYZ-coordinates were parsed from the output 'cmm' files.

We took the coordinates of SCZ sensitive H3K27ac NeuN+ and Tissue CRDs and BD sensitive H3K27ac Tissue CRDs and overlapped with the PFC NeuN+ TAD coordinates obtained from above. To test the presence of localized SCZ or BD sensitive CRDs in the 3D genome, we measured the pairwise 3D distance of the TADs that overlapped with diseased CRDs stratified by clusters as shown in Figure 6.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank late Pamela Sklar for her numerous contributions in the early phase of this project and Prashanth Rajarajan and Sergio Espeso-Gil for helpful discussions. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. We are extremely grateful to J. Ochando, C. Bare and other personnel of the Icahn School of Medicine at Mount Sinai's Flow Cytometry Core for providing and teaching cell sorting expertise and to Lora Bingman in the Division of Neuroscience and Basic Behavioral Science (DNBBS) at the National Institute of Mental Health (NIH) for logistical support in context of the PsychENCODE consortium.

This project was supported by NIH U01DA048279 (S.A., P.R.) and R01MH106056 (S.A.). PsychENCODE Consortium -- Data were generated as part of the first phase of the PsychENCODE Consortium supported by: U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to: Shahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale), Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (UIC), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman (Yale), Kevin White (UChicago) and Peter Zandi (JHU). The HBCC is funded by the NIMH-IRP through project ZIC MH002903. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Data availability:

Raw (FASTQ files) and processed data (BigWig files, metadata, peaks, and raw / normalized count matrices) has been deposited in synapse under synID syn25705564 <https://www.synapse.org/#!/Synapse:syn25705564>. Browseable UCSC genome browser tracks of our processed ChIP-seq data are available as a resource at: [EpiDiff Phase 2](#).

External validation sets used in the study are: H3K27ac ChIP-seq fetal specific peaks: Spatio-temporal enrichment of H3K27ac peaks table from <http://development.psychencode.org/#>, RoadMap Epigenome Project (REP) H3K27ac, H3K4me3 tissue ChIP-seq peaks, chromHMM states on E073 and fetal male E081 and fetal female E082 <https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/> and CTCF ChIP-seq on human neural cell (GEO GSE127577). TruSeq3-PE.fa file was downloaded from the adaptor folder under

the trimmomatic repository. <https://github.com/timflutre/trimmomatic/blob/master/adapters/TruSeq3-PE.fa>

The source data described in this manuscript are available via the PsychENCODE Knowledge Portal (<https://psychencode.synapse.org/>). The PsychENCODE Knowledge Portal is a platform for accessing data, analyses, and tools generated through grants funded by the National Institute of Mental Health (NIMH) PsychENCODE program. Data is available for general research use according to the following requirements for data access and data attribution: (<https://psychencode.synapse.org/DataAccess>).

## References:

1. Dixon JR et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336 (2015). [PubMed: 25693564]
2. Girdhar K et al. Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* 21, 1126–1136 (2018). [PubMed: 30038276]
3. Cheung I et al. Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proc Natl Acad Sci USA* 107, 8824–8829 (2010). [PubMed: 20421462]
4. Khan A, Mathelier A & Zhang X Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers. *Epigenetics* 13, 910–922 (2018). [PubMed: 30169995]
5. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* 18, 199–209 (2015). [PubMed: 25599223]
6. Roussos P et al. A role for noncoding variation in schizophrenia. *Cell Rep.* 9, 1417–1429 (2014). [PubMed: 25453756]
7. Fullard JF et al. An atlas of chromatin accessibility in the adult human brain. *Genome Res.* 28, 1243–1252 (2018). [PubMed: 29945882]
8. Hauberg ME et al. Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat. Commun.* 11, 5581 (2020). [PubMed: 33149216]
9. Smigielski L, Jagannath V, Rössler W, Walitza S & Grünblatt E Epigenetic mechanisms in schizophrenia and other psychotic disorders: a systematic review of empirical human findings. *Mol. Psychiatry* 25, 1718–1748 (2020). [PubMed: 31907379]
10. Fromer M et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19, 1442–1453 (2016). [PubMed: 27668389]
11. Hoffman GE et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data* 6, 180 (2019). [PubMed: 31551426]
12. Hauberg ME et al. Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls. *Mol. Psychiatry* 24, 1685–1695 (2019). [PubMed: 29740122]
13. Kozlenkov A et al. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Sci. Adv.* 4, eaau6190 (2018). [PubMed: 30263963]
14. Wong AHC et al. Association between schizophrenia and the syntaxin 1A gene. *Biol. Psychiatry* 56, 24–29 (2004). [PubMed: 15219469]
15. Bryois J et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* 9, 3121 (2018). [PubMed: 30087329]
16. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235 (2015). [PubMed: 26414678]
17. Madani Tonekaboni SA, Mazrooei P, Kofia V, Haibe-Kains B & Lupien M Identifying clusters of cis-regulatory elements underpinning TAD structures and lineage-specific regulatory networks. *Genome Res.* 29, 1733–1743 (2019). [PubMed: 31533978]

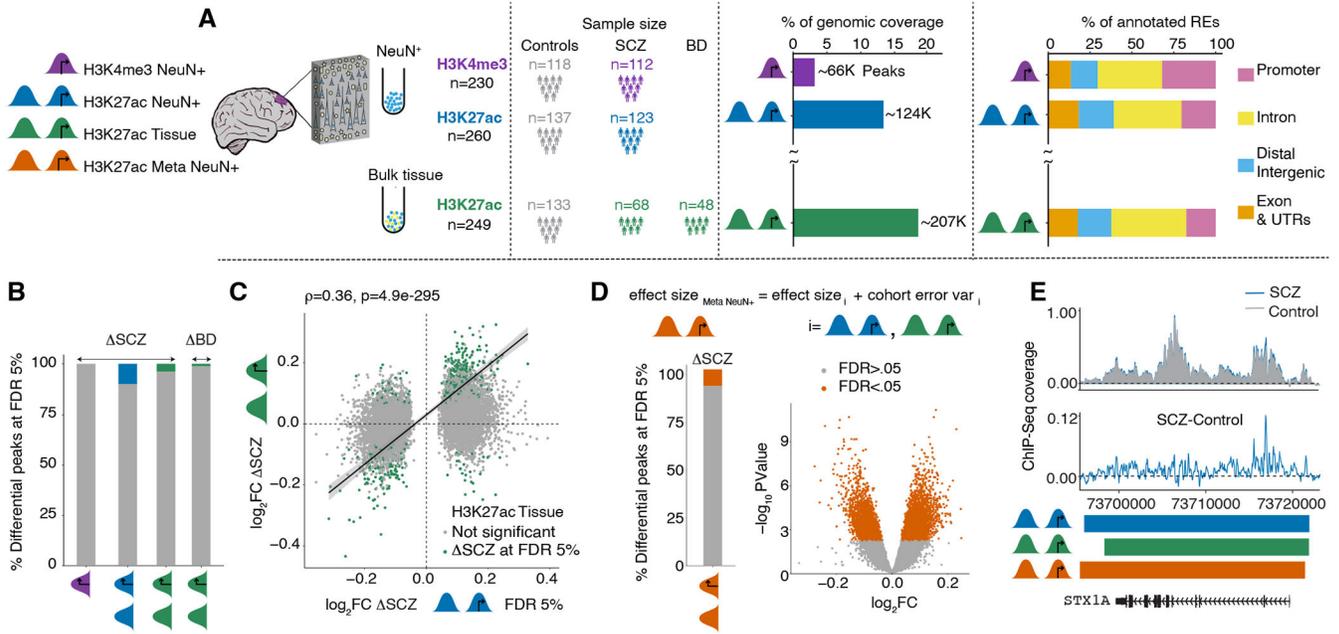
18. Bendl J et al. The three-dimensional landscape of chromatin accessibility in Alzheimer's disease. *BioRxiv* (2021) doi:10.1101/2021.01.11.426303.
19. Dong P et al. Population-level variation of enhancer expression identifies novel disease mechanisms in the human brain. *BioRxiv* (2021) doi:10.1101/2021.05.14.443421.
20. Delaneau O et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* 364, (2019).
21. Waszak SM et al. Population variation and genetic control of modular chromatin architecture in humans. *Cell* 162, 1039–1050 (2015). [PubMed: 26300124]
22. Hoffman GE, Bendl J, Girdhar K & Roussos P decorate: differential epigenetic correlation test. *Bioinformatics* 36, 2856–2861 (2020). [PubMed: 32003784]
23. Ambrose C, Dehman A, Neuvial P, Rigail G & Vialaneix N Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics. *Algorithms Mol. Biol.* 14, 22 (2019).
24. Beagan JA & Phillips-Cremens JE On the existence and functionality of topologically associating domains. *Nat. Genet.* 52, 8–16 (2020). [PubMed: 31925403]
25. Nuebler J, Fudenberg G, Imakaev M, Abdennur N & Mirny LA Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci USA* 115, E6697–E6706 (2018). [PubMed: 29967174]
26. Kichaev G et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* 104, 65–75 (2019). [PubMed: 30595370]
27. Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012). [PubMed: 22495300]
28. Lazar NH et al. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res.* 28, 983–997 (2018). [PubMed: 29914971]
29. Hoffman GE et al. Sex differences in the human brain transcriptome of cases with schizophrenia. *Biol. Psychiatry* 91, 92–101 (2022). [PubMed: 34154796]
30. Li M et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, (2018).
31. Paulsen J et al. Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol.* 18, 21 (2017). [PubMed: 28137286]
32. Paulsen J, Liyakat Ali TM & Collas P Computational 3D genome modeling using Chrom3D. *Nat. Protoc.* 13, 1137–1152 (2018). [PubMed: 29700484]
33. Tseng C-EJ et al. In vivo human brain expression of histone deacetylases in bipolar disorder. *Transl. Psychiatry* 10, 224 (2020). [PubMed: 32641695]
34. Gilbert TM et al. PET neuroimaging reveals histone deacetylase dysregulation in schizophrenia. *J. Clin. Invest.* 129, 364–372 (2019). [PubMed: 30530989]
35. Schroeder FA et al. Expression of HDAC2 but Not HDAC1 Transcript Is Reduced in Dorsolateral Prefrontal Cortex of Patients with Schizophrenia. *ACS Chem. Neurosci.* 8, 662–668 (2017). [PubMed: 27959513]
36. Bahari-Javan S et al. HDAC1 links early life stress to schizophrenia-like phenotypes. *Proc Natl Acad Sci USA* 114, E4686–E4694 (2017). [PubMed: 28533418]
37. Jakovcevski M et al. Prefrontal cortical dysfunction after overexpression of histone deacetylase 1. *Biol. Psychiatry* 74, 696–705 (2013). [PubMed: 23664640]
38. Schroeder FA, Lin CL, Crusio WE & Akbarian S Antidepressant-like effects of the histone deacetylase inhibitor, sodium butyrate, in the mouse. *Biol. Psychiatry* 62, 55–64 (2007). [PubMed: 16945350]
39. de la Fuente Revenga M. et al. HDAC2-dependent Antipsychotic-like Effects of Chronic Treatment with the HDAC Inhibitor SAHA in Mice. *Neuroscience* 388, 102–117 (2018). [PubMed: 30025863]
40. Thomas EA Histone posttranslational modifications in schizophrenia. *Adv. Exp. Med. Biol.* 978, 237–254 (2017). [PubMed: 28523550]
41. Sakaue S et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* 53, 1415–1424 (2021). [PubMed: 34594039]

42. Connor CM et al. Maternal immune activation alters behavior in adult offspring, with subtle changes in the cortical transcriptome and epigenome. *Schizophr. Res.* 140, 175–184 (2012). [PubMed: 22804924]
43. Jaffe AE et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* 19, 40–47 (2016). [PubMed: 26619358]
44. Hannon E et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* 19, 48–54 (2016). [PubMed: 26619357]
45. Ruzicka WB et al. Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv* (2020) doi:10.1101/2020.11.06.20225342.
46. Dienel SJ, Enwright JF, Hoftman GD & Lewis DA Markers of glutamate and GABA neurotransmission in the prefrontal cortex of schizophrenia subjects: Disease effects differ across anatomical levels of resolution. *Schizophr. Res.* 217, 86–94 (2020). [PubMed: 31296415]
47. Bonev B et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171, 557–572.e24 (2017). [PubMed: 29053968]
48. Lomvardas S et al. Interchromosomal interactions and olfactory receptor choice. *Cell* 126, 403–413 (2006). [PubMed: 16873069]
49. Quinodoz SA et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744–757.e24 (2018). [PubMed: 29887377]
50. Khanna N, Hu Y & Belmont AS HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* 24, 1138–1144 (2014). [PubMed: 24794297]
51. Ahanger SH et al. Distinct nuclear compartment-associated genome architecture in the developing mammalian brain. *Nat. Neurosci.* 24, 1235–1242 (2021). [PubMed: 34239128]
52. Legge SE et al. Associations between schizophrenia polygenic liability, symptom dimensions, and cognitive ability in schizophrenia. *JAMA Psychiatry* (2021) doi:10.1001/jamapsychiatry.2021.1961.

## References (Methods)

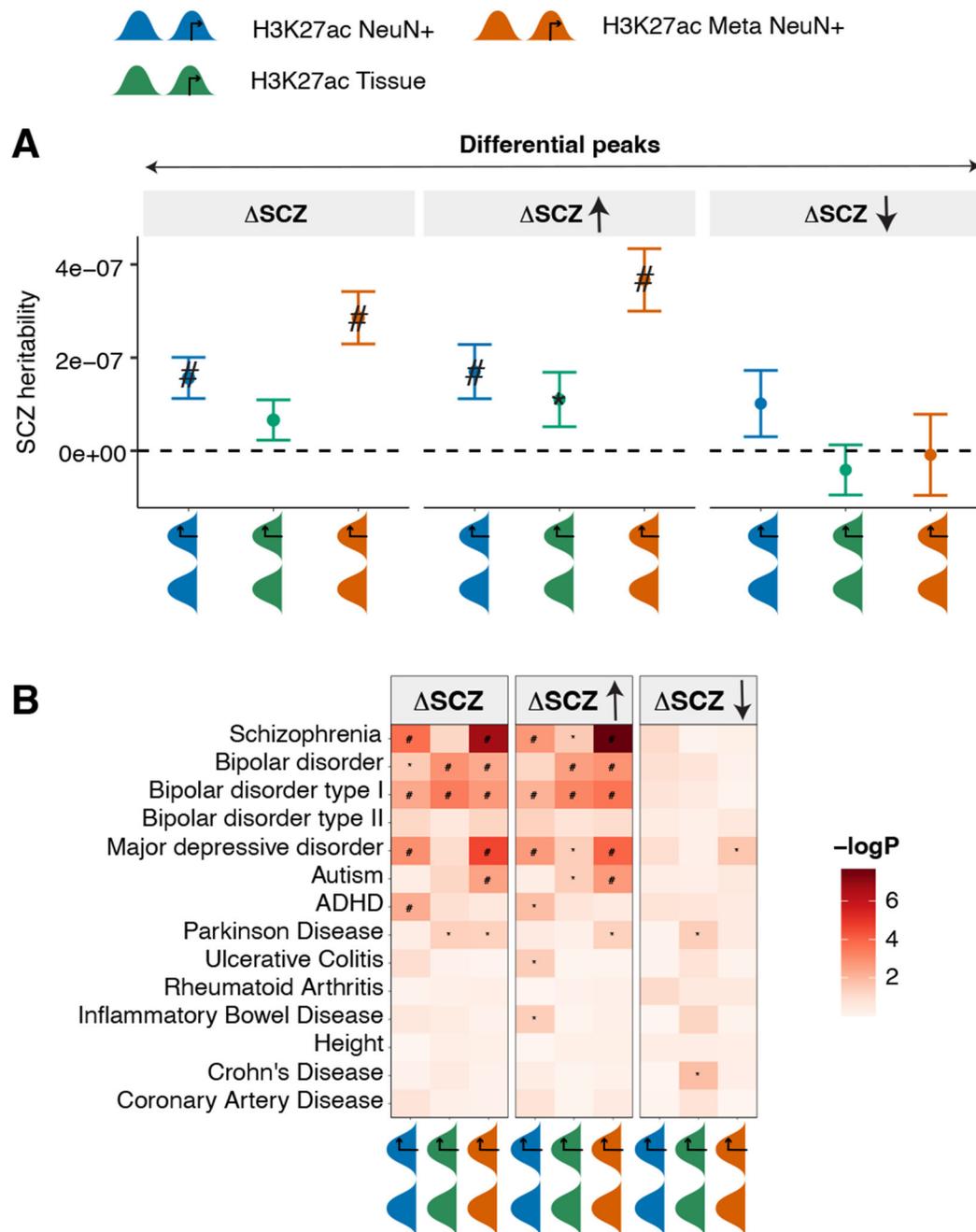
53. Wang D et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, (2018).
54. Kundakovic M et al. Practical Guidelines for High-Resolution Epigenomic Profiling of Nucleosomal Histones in Postmortem Human Brain Tissue. *Biol. Psychiatry* 81, 162–170 (2017). [PubMed: 27113501]
55. Jiang Y, Matevosian A, Huang H-S, Straubhaar J & Akbarian S Isolation of neuronal chromatin from brain tissue. *BMC Neurosci.* 9, 42 (2008). [PubMed: 18442397]
56. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014). [PubMed: 24695404]
57. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009). [PubMed: 19451168]
58. Landt SG et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831 (2012). [PubMed: 22955991]
59. Fort A et al. MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* 33, 1895–1897 (2017). [PubMed: 28186259]
60. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008). [PubMed: 18798982]
61. Amemiya HM, Kundaje A & Boyle AP The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* 9, 9354 (2019). [PubMed: 31249361]
62. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2014). [PubMed: 24227677]
63. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010). [PubMed: 19910308]

64. Hunt GJ, Freytag S, Bahlo M & Gagnon-Bartsch JA dtangle: accurate and robust cell type deconvolution. *Bioinformatics* 35, 2093–2099 (2019). [PubMed: 30407492]
65. Neath AA & Cavanaugh JE The Bayesian information criterion: background, derivation, and applications. *WIREs Comp Stat* 4, 199–203 (2012).
66. Yu G, Wang L-G & He Q-Y ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383 (2015). [PubMed: 25765347]
67. Ernst J & Kellis M Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* 12, 2478–2492 (2017). [PubMed: 29120462]
68. Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47 (2015). [PubMed: 25605792]
69. Viechtbauer W Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* 36, (2010).
70. McLean CY et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501 (2010). [PubMed: 20436461]
71. de Leeuw CA, Mooij JM, Heskes T & Posthuma D MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11, e1004219 (2015). [PubMed: 25885710]
72. Stegle O, Parts L, Piipari M, Winn J & Durbin R Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* 7, 500–507 (2012). [PubMed: 22343431]
73. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012). [PubMed: 22955616]
74. Kumar V et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* 31, 615–622 (2013). [PubMed: 23770639]
75. Kozlenkov A et al. Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucleic Acids Res.* 44, 2593–2612 (2016). [PubMed: 26612861]
76. Van den Berge K, Sonesson C, Robinson MD & Clement L stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biol.* 18, 151 (2017). [PubMed: 28784146]
77. Forgy E Cluster analysis of multivariate data : efficiency versus interpretability of classifications. undefined (1965).
78. Servant N et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259 (2015). [PubMed: 26619908]
79. Ramírez F et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* 9, 189 (2018). [PubMed: 29335486]
80. Abdennur N & Mirny LA Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* 36, 311–316 (2020). [PubMed: 31290943]



**Figure 1: Histone peak profiling in 739 ChIP-Seq datasets from two studies consisting of SCZ, BD and control subjects**

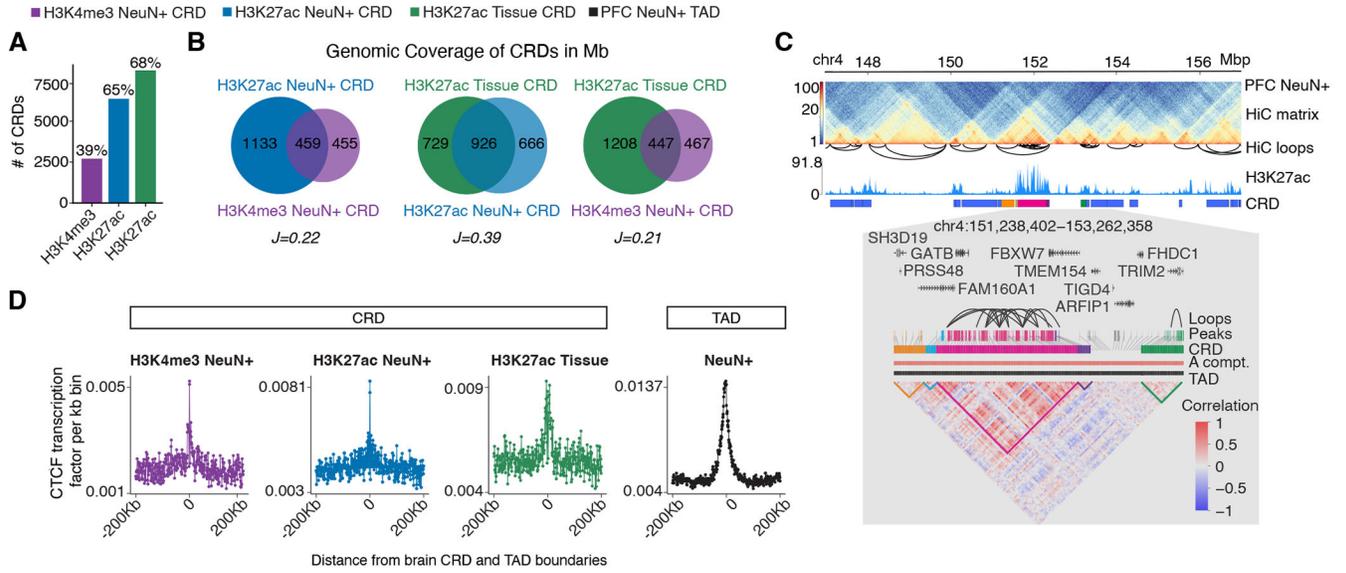
(A) (left) Datasets and studies; study-1, FANS isolated PFC NeuN<sup>+</sup> nuclei: H3K4me3 (purple), H3K27ac (blue). study-2, total (non-sorted) tissue PFC nuclei: H3K27ac (green). (middle) Bar plot, genomic coverage (%) of each ChIP-Seq dataset. Numbers of subjects as indicated and bar plot to show % of regulatory elements in each dataset. (B) Bar plots (colored sectors marking significantly different peaks) showing the proportion of differentially regulated histone peaks in our four datasets of case control comparisons, as indicated. See also Figure S2 for overlap of disease-sensitive peaks across the various H3K27ac datasets (C) Spearman correlation of effect sizes of H3K27ac NeuN<sup>+</sup> peaks altered (\*) in SCZ study-1, compared with effect sizes for corresponding peaks in SCZ study-2 subjects with bulk PFC Tissue as input. Green dots mark H3K27ac NeuN<sup>+</sup> at FDR < 5%. P value (p), Spearman rank correlation test. (D) Meta-analysis of H3K27ac NeuN<sup>+</sup> and H3K27ac Tissue: The bar shows differential SCZ-specific peaks of H3K27ac Meta NeuN<sup>+</sup> at FDR < 5%. Volcano plot of differentially modified H3K27ac Meta NeuN<sup>+</sup> peaks by fixed effect model. Orange dots represent peaks with FDR < 5%. P values for t-test of limma pipeline, with FDR correction (Benjamini & Hochberg) across all peaks. (E) (top) Visualization of the hyperacetylated region covering *STX1A* gene. H3K27ac NeuN<sup>+</sup> ChIP-Seq landscape from (gray) controls and (blue) SCZ individuals. (bottom) Differential peak profile, horizontal bars demarcate linear extension of disease-sensitive peak population for each of the three H3K27ac datasets, as indicated (blue) H3K27ac NeuN<sup>+</sup> study 1, (green) H3K27ac Tissue study-2, (orange) H3K27ac NeuN<sup>+</sup> meta-analysis.



**Figure 2: Enrichment of common SCZ risk variants in dysregulated peaks in NeuN+ and bulk tissue.**

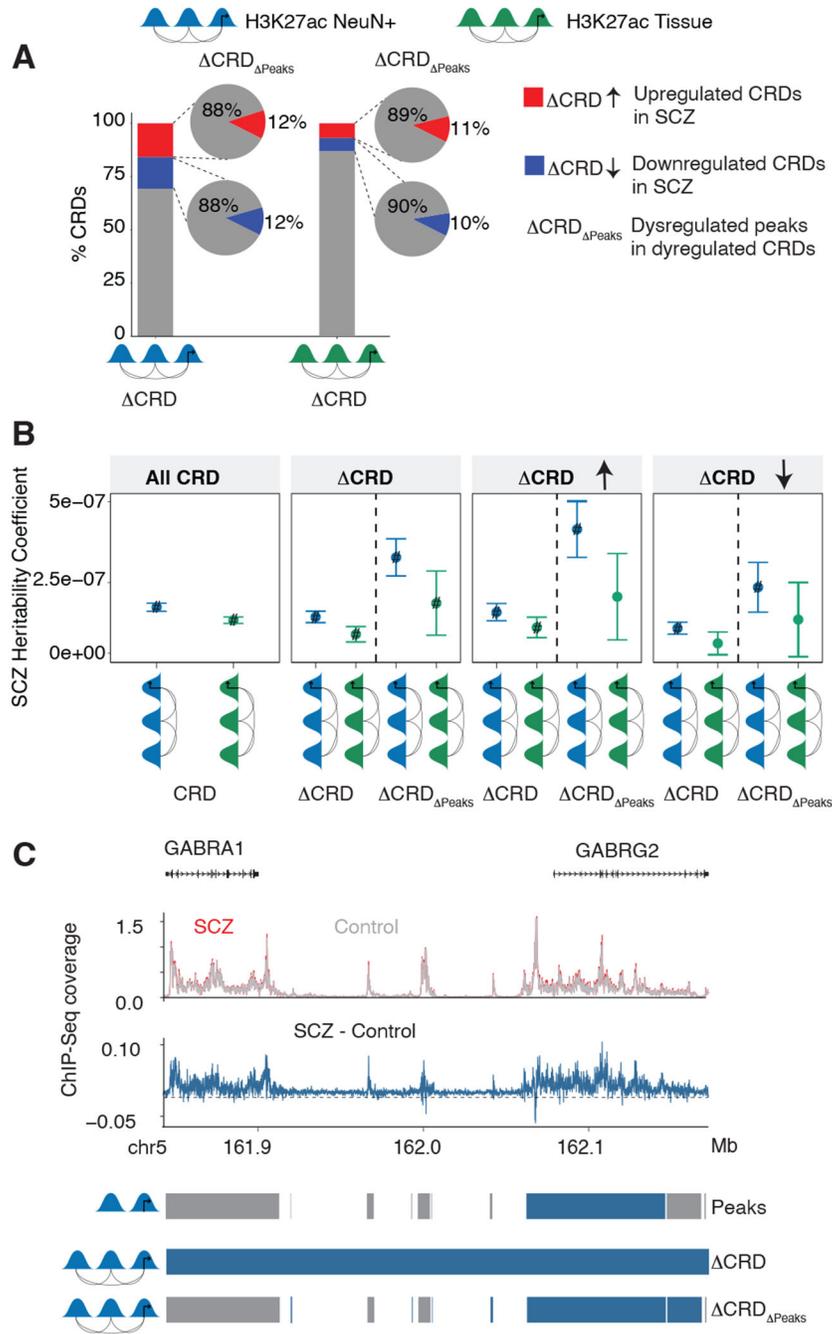
(A) SCZ heritability coefficients of genetic variants overlapping histone peaks from study-1 H3K27ac NeuN+, study-2 H3K27ac Tissue and H3K27ac Meta NeuN+ stratified by 1) “ $\Delta$ SCZ”: dysregulated peaks (n=3360, 5656 and 6219 peaks, respectively) 2) “ $\Delta$ SCZ $\uparrow$ ”: hyperacetylated dysregulated peaks (n=1918, 2681 and 4031 peaks, respectively) and 3) “ $\Delta$ SCZ $\downarrow$ ”: hypoacetylated dysregulated peaks (n=1442, 2975 and 2188 peaks, respectively) with  $\log_2FC$  (SCZ vs controls)  $>0$  and  $<0$  respectively. Error bars represent

standard error in SCZ heritability from LDSc regression (**B**) Heatmap of enrichment  $P$ -values of brain-related GWAS traits. The overlap of peaks with genetic variants was assessed using LD score regression. "#": Significant for enrichment in LD score regression after FDR correction of multiple testing across all tests in the plot (Benjamini & Hochberg test); "\*": Nominally significant for enrichment.



**Figure 3: PFC histone CRDs reveal subTAD chromosomal organization.**

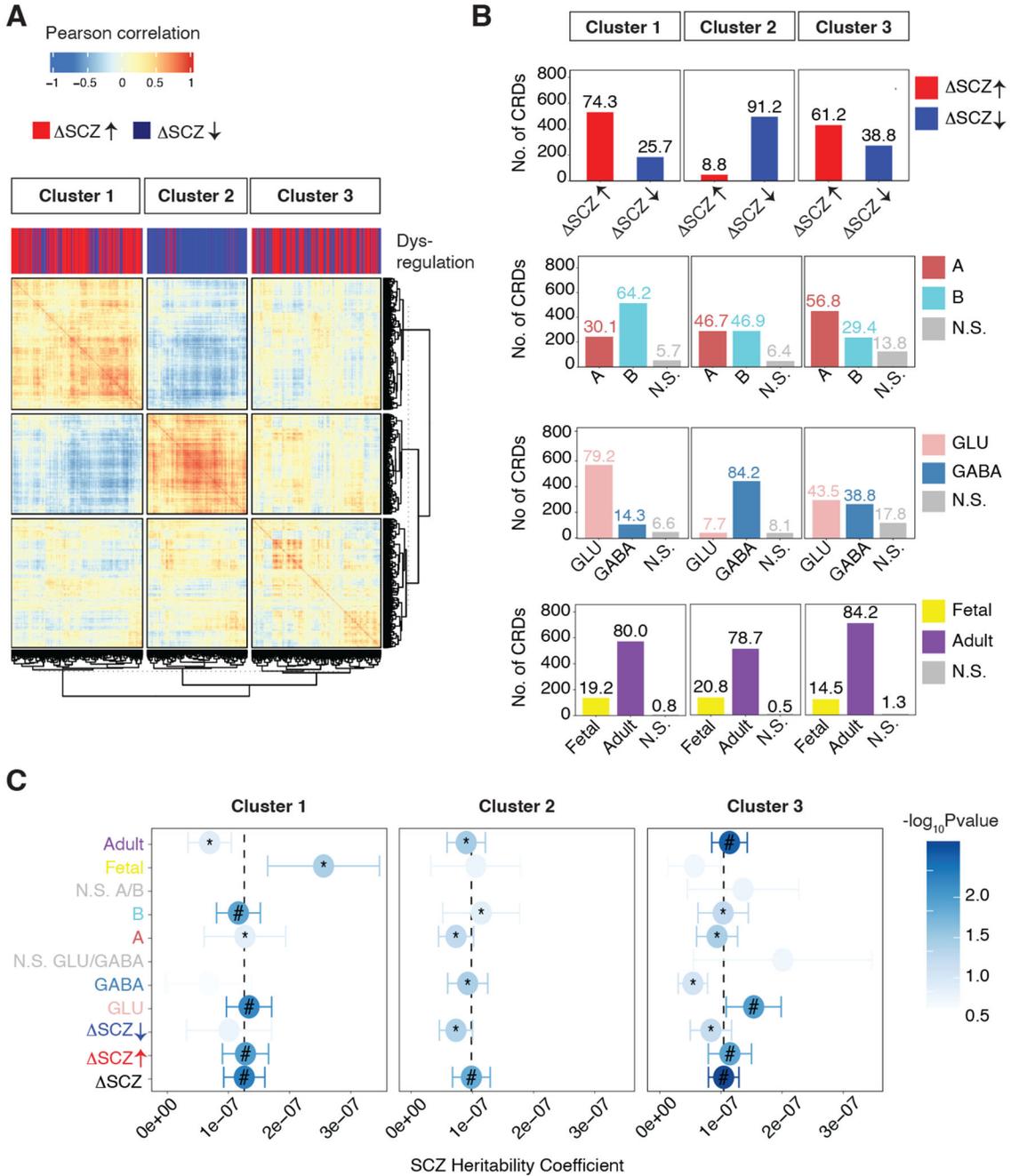
(A) CRD analyses were conducted separately for each of our three ChIP-seq datasets (H3K4me3 NeuN+, H3K27ac NeuN+, H3K27ac Tissue) (see also Figure 1A,B). The numbers next to bars indicate the proportion of total peak population integrated into CRD structures. (B) Venn diagrams summarizing genome-wide sequences, in Mb megabases, integrated into CRD structures, including overlap and Jaccard similarity index between different histones and cell population (NeuN+ or tissue). (C) (top) Representative 10Mb window of chromosome 4 showing PFC NeuN+ Hi-C TAD, chromosomal loop and H3K27ac landscape including CRD structure. (bottom, shaded in gray color) Higher resolution (2Mb) peak-to-CRD assignments and peak correlational structure expressed as an interaction matrix. (D) Neuronal CTCF chromatin occupancies (Y-axis, using CTCF ENCODE reference ChIP-seq from H1 stem cell-differentiated neuronal culture) in relation to distance from CRD (colored graphs) and TAD (black graph) boundaries.



**Figure 4: Acetylated CRDs dysregulated in SCZ.**

(A) Proportional representation of SCZ-sensitive H3K27ac NeuN+ CRDs stratified by hypoacetylation (blue), hyperacetylation (red) and not dysregulated (gray). Pie chart shows the proportion of dysregulated histone peaks  $\text{CRD}_{\text{Peaks}}$  inside CRD (B) SCZ heritability coefficients shown separately for (blue) H3K27ac NeuN+ and (green) H3K27ac Tissue, as indicated, by 1) “All CRD”: all peaks inside CRD (n=114,123 and 143,092 peaks in CRD), 2) “ $\Delta\text{CRD}$ ”: dysregulated CRD (n=28,866 and 15,787 peaks in CRD; 3,507 and 1,673 peaks in  $\text{CRD}_{\text{Peaks}}$ ), 3) “ $\text{CRD}\uparrow$ ”: hyperacetylated with mean  $\log_2\text{FC}$  (SCZ vs controls)

> 0 (n=14,710 and 7,770 peaks in CRD<sup>↑</sup>, 1,825 and 873 peaks in CRD<sup>↑</sup> Peaks), and 4) “CRD<sup>↓</sup>”: hypoacetylated with mean log<sub>2</sub>FC (SCZ vs controls) < 0 (n=14,156 and 8,017 peaks in CRD<sup>↓</sup>; 1,682 and 800 peaks in CRD<sup>↓</sup> Peaks) classified on x-axis as CRD for all peaks and CRD Peaks for only dysregulated histone peaks. The overlap of peaks within the dysregulated CRDs in clusters with SCZ risk variants was assessed using LD score regression. “#”: Significant for enrichment in LD score regression after FDR correction of multiple testing across all tests in the plot (Benjamini & Hochberg, multiple testing p value <0.05); “\*”: Nominally significant for enrichment (p value <0.05). Error bars show standard error in SCZ heritability from LDSc regression (C) Representative example of a genomic region which spans ten peaks. Horizontal bars (blue, FDR 5%, gray n.s.) mark (top row) peak level analysis with a single differential peak; (mid row) hyperacetylated CRD and (bottom row) CRD Peaks.



**Figure 5: Fingerprinting disease-sensitive CRDs.**

(A) CRD contact matrix of SCZ-sensitive H3K27ac NeuN+ CRDs clustered into three large clusters; notice striking separation of cluster 1 and cluster 3 representing hyperacetylated H3K27ac CRDs in red, and clusters 2 overwhelmingly defined by hypoacetylated H3K27ac CRDs in navy. (B) Composition of annotated CRDs by cell type (GABA in light blue, GLU in pink), compartments (A in cyan, B in indianred), dysregulation (hypo vs hyperacetylation in red and navy) and development (fetal in yellow, adult in purple). For every annotation, not significant or not annotated CRDs are shown in gray. (C) Coefficients of

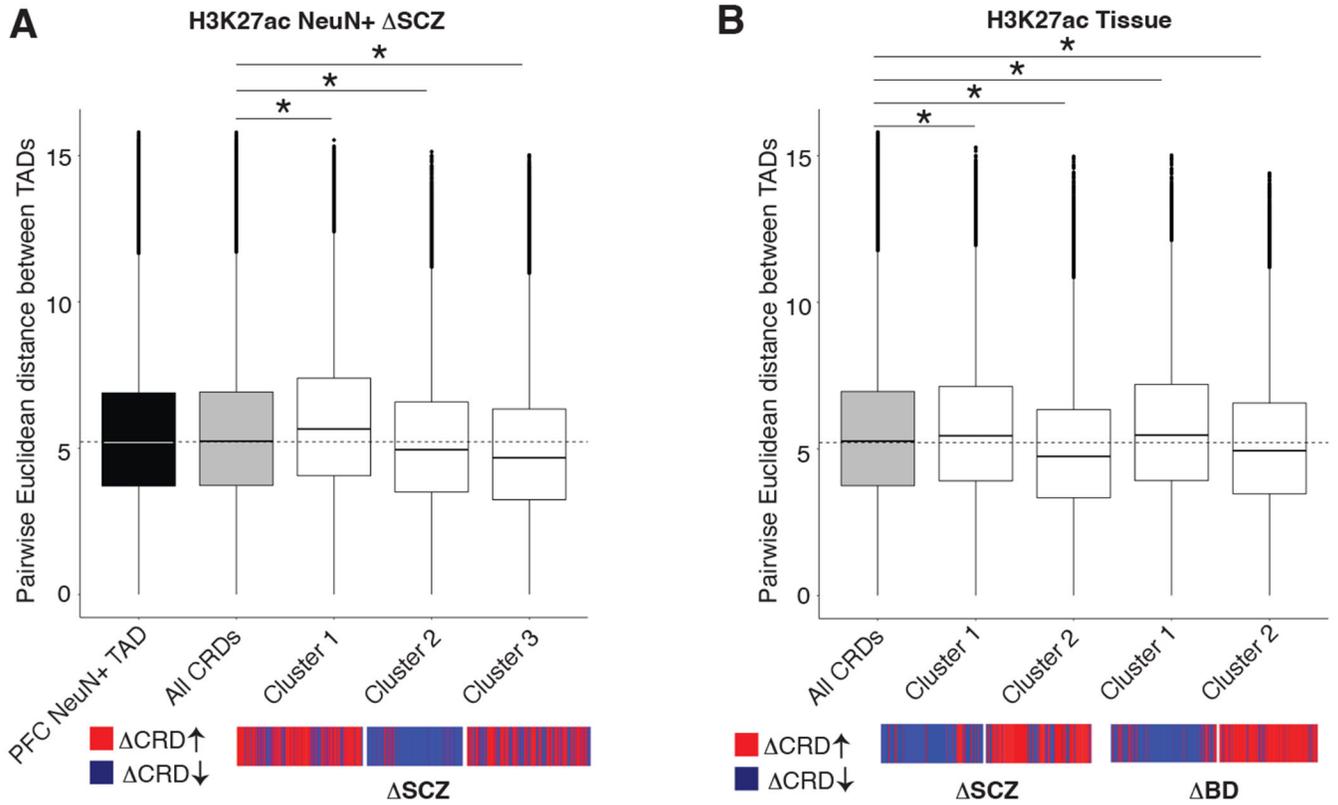
heritability of SCZ by cluster and annotation. The overlap of peaks within the dysregulated CRDs in clusters with genetic variants was assessed using LD score regression. P Values from LDSc regression. "#": Significant for enrichment in LD score regression after Benjamini and Hochberg FDR correction for multiple testing across all tests in the plot (FDR < 5%). "\*\*": Nominally significant for enrichment (p value<.05). Error bars show standard error in SCZ heritability from LDSc regression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 6: Spatial organization of diseased CRDs in the virtual chrom 3D model of the neuronal nucleus.**

(A) Box plots show for SCZ H3K27ac NeuN+ cohort, showing for diseased CRDs the pairwise Euclidean distance of PFC NeuN+ TAD, and all CRDs, and diseased CRDs from cluster-1, cluster-2 and cluster-3 (from figure 5) with  $n=11,103$ ,  $8,690$  and  $9,073$  peaks for each clusters as indicated. (B) Box plots for H3K27ac Tissue cohort, separated by SCZ sensitive and BD sensitive CRDs, with  $n = n=8,553$  in cluster 1 and  $7,234$  peaks in cluster 2 (from figure 5). Barplots in the bottom in red and navy show the dysregulation status of clusters. The center shows the median, the box shows the interquartile range, whiskers indicate the highest/lowest values within  $1.5\times$  the interquartile range, and potential outliers from this are shown as dots. \* represents  $p$  value  $< .05$  while  $p$  values are estimated using the Wilcoxon test.

**Table 1:**

Summary of metadata of samples in study-1 and study-2

Study	Histone Mark	PFC cell type	Brain Bank	Diagnosis (N)	Sex (N)	Age (yrs)	Ethnicity (N)
Study-1	H3K4me3	NeuN+	MSSM PENN PITT	112 SCZ 118 Ctrl	92F 138M	66.33±18.42	183 EU, 34 AA, 2 EA, 10 HISP, 1 Multiracial
	H3K27ac			123 SCZ 137 Ctrl	98F 162M	67.32±18.02	203 EU, 38 AA, 2 EA, 16 HISP, 1 Multiracial
Study-2	H3K27ac	Tissue	HBCC	68 SCZ 48 BD 133Ctrl	88F 161M	42.27±17.59	121 EU, 117 AA, 6 EA, 5 HISP

SCZ: Schizophrenia, BD: Bipolar Disorder, Ctrl: Control F: Female, M: Male EU: European, AA:African-American, EA: East Asian, HISP: Hispanic

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript