nature portfolio

https://doi.org/10.1038/s41593-022-01032-6

Supplementary information

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

In the format provided by the authors and unedited

Supplementary Information

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

This PDF file includes:

Captions for Supplementary Figures S1-S22
Supplementary Figs. 1-22

Other Supplementary Materials/ Data Files for this manuscript include the following:

Supplementary Tables 1-8 (.xlsx)

Supplementary Figures

Figure S1 | Mapped reads and phantompeakqualtools quality metrics.

Figure S2 | Overlap of differential peaks.

Figure S3 | Concordance analysis between the SCZ and BP effect sizes across and within two studies.

Figure S4 | Pathway analysis of SCZ and BD dysregulated peaks.

Figure S5 | Concordance analysis between SCZ dysregulated peaks and antipsychotic typical/atypical dysregulated peaks.

Figure S6 | Enrichment of SCZ and BD dysregulated peaks in common risk variants of SCZ and BD.

Figure S7 | Schematic workflow of the steps in CRD analyses.

Figure S8 | CRD filtering and merging.

Figure S9 | Enrichment of CRDs inside subTADs and TADs.

Figure S10 | Proportion of CRDs inside subTADs/TADs and vice-versa.

Figure S11 | Density of CRDs inside subTADs/TADs and vice-versa.

Figure S12 | Concordance between sets of dysregulated CRDs across and within the studies.

Figure S13 | Summary table of differential CRDs.

Figure S14 | Enrichment of dysregulated BD H3K27ac Tissue CRDs in common risk variants of SCZ.

Figure S15 | Summary table of coefficients of generalized linear model.

Figure S16 | Demonstration of nuclear organization in dysregulated CRDs.

Figure S17 | Evaluation of optimal number of clusters in CRD contact matrices.

Figure S18 | Annotated CRDs.

Figure S19 | Fingerprinting disease-sensitive CRDs in H3K27ac Tissue.

Figure S20 | RNAseq expression of disease-specific CRDs.

Figure S21 | Cluster 2 is enriched for SCZ and BD risk variants.

Figure S22 | Chromatin immunoprecipitation followed by deep sequencing quality controls.

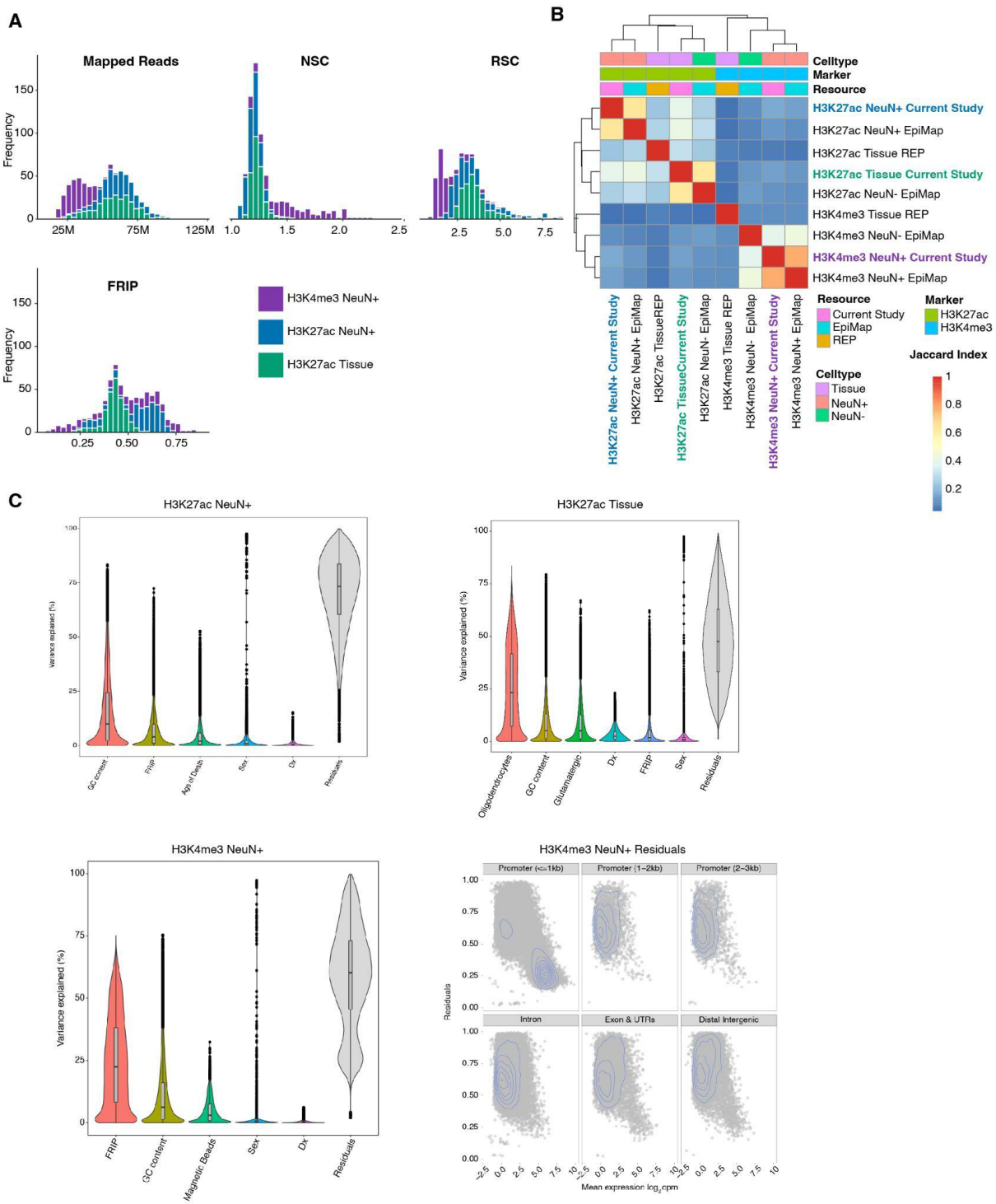
A
Mapped Reads
NSC
RSC
Frequency
25M
75M
125M
1.0
1.5
2.0
2.5
5.0
7.5
FRIP
0.25
0.50
0.75
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
B
Celltype
Marker
Resource
H3K27ac NeuN+ Current Study
H3K27ac NeuN+ EpiMap
H3K27ac Tissue REP
H3K27ac Tissue Current Study
H3K27ac NeuN- EpiMap
H3K4me3 Tissue REP
H3K4me3 NeuN- EpiMap
H3K4me3 NeuN+ Current Study
H3K4me3 NeuN+ EpiMap
Current Study
EpiMap
REP
H3K27ac
H3K4me3
Tissue
NeuN+
NeuN-
Jaccard Index
1
0.8
0.6
0.4
0.2
C
H3K27ac NeuN+
Variance explained (%)
GC content
FRIP
Age of Death
Sex
Dx
Residuals
H3K27ac Tissue
Oligodendrocytes
GC content
Glutamatergic
Dx
FRIP
Sex
Residuals
H3K4me3 NeuN+
FRIP
GC content
Magnetic Beads
Sex
Dx
Residuals
H3K4me3 NeuN+ Residuals
Promoter (<=1kb)
Promoter (1–2kb)
Promoter (2–3kb)
Intron
Exon & UTRs
Distal Intergenic
Residuals
Mean expression log2cpm

Figure S1 | Mapped reads and phantompeakqualtools quality metrics. (A) The distribution of mapped reads, Normalized Strand Cross-correlation coefficient (NSC) distribution, Relative Strand Cross-correlation coefficient (RSC) and Fraction of Reads In Peak (FRIP) for 230, 260 and 249 samples stratified by studies (H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue). **(B)** Heatmap to show Jaccard index between pairs of datasets 1) study-1 H3K27ac NeuN+, 2) EpiMap H3K27ac NeuN+, 3) REP H3K27ac Tissue, 4) study-2 H3K27ac Tissue, 5) EpiMap H3K27ac NeuN-, 6) study-1 H3K4me3 NeuN+ and 7) EpiMap-H3K4me3 NeuN+. **(C)** Violin plots of (upper) H3K27ac NeuN+ Tissue and (lower left) H3K4me3 NeuN+ to show the distribution of variance in each peak explained by the identified confounds plotted as x-axis across all three studies. (Lower right) plot of H3K4me3 NeuN+ residuals vs. mean expression of log2cpm of peaks stratified by regulatory elements to illustrate the lower hump of residuals distribution if mainly from promoter peaks (<=1kb TSS)

Figure S2 | Overlap of differential peaks. Venn diagrams to show **(A)** H3K27ac NeuN+ ΔSCZ peaks that overlapped at least one base pair with H3K27ac Tissue ΔSCZ **(B)** H3K27ac Tissue ΔSCZ peaks that overlapped with H3K27ac Tissue ΔBD and **(C)** H3K27ac NeuN ΔSCZ with at least one base pair overlap with H3K27ac Tissue ΔBD.

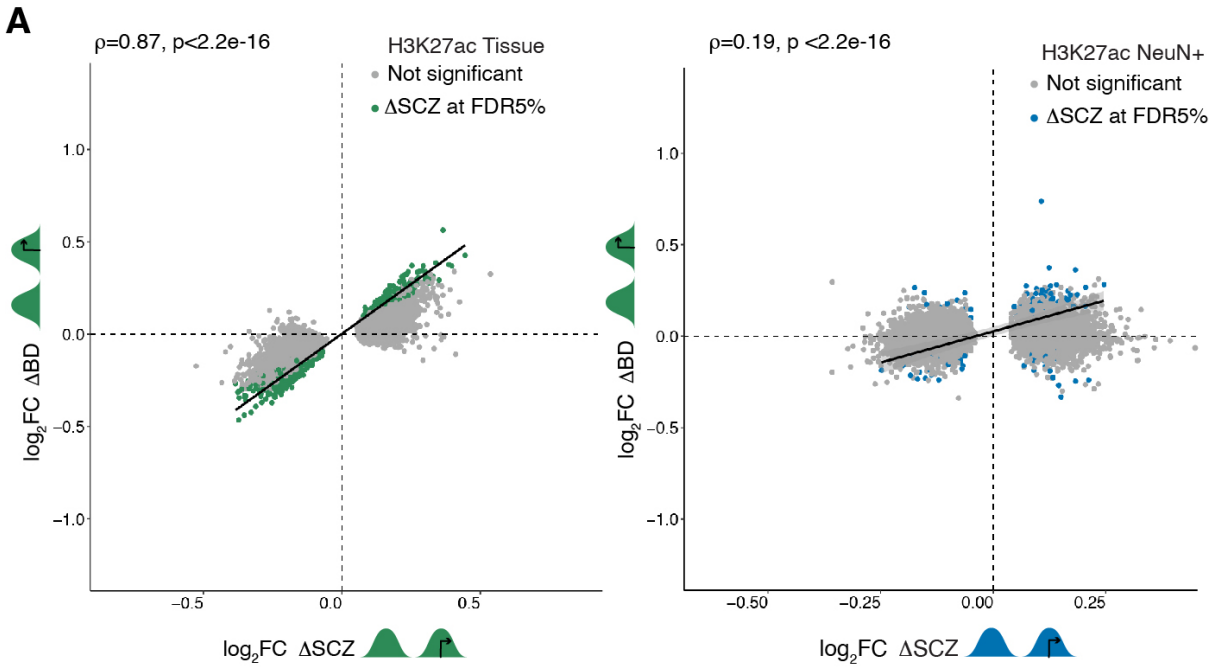

Figure S3 | Concordance analysis between the SCZ and BP effect sizes across and within two studies. (A) Left plot shows the spearman correlation between log_2FC (SCZ vs. controls) H3K27ac NeuN+ and log_2FC (BD vs. controls) H3K27ac Tissue. Green colored points are significant in SCZ vs. controls and BD vs. controls. The right plot shows correlation between log_2FC (SCZ vs. controls) H3K27ac NeuN+ at FDR 5% and log_2FC (BD vs. controls) H3K27ac Tissue. Blue colored points are significant in both studies

A

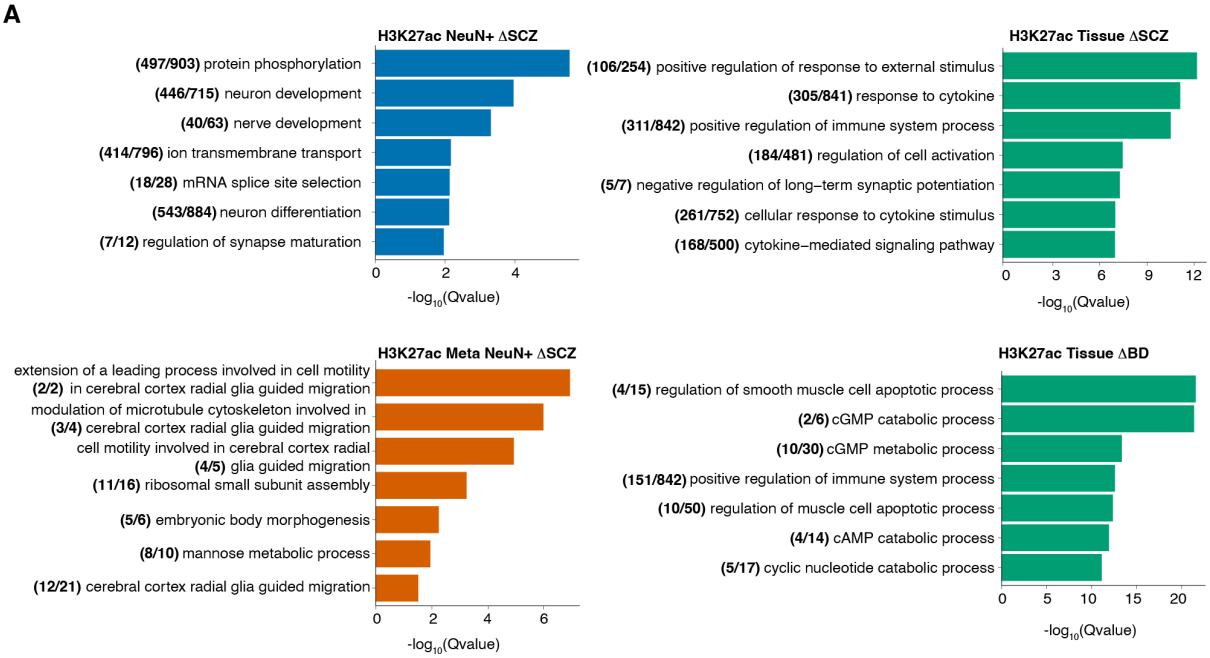

Figure S4 | Pathway analysis of SCZ and BD dysregulated peaks. (A) Gene set enrichment analysis of dysregulated peaks at FDR 5 % identified from differential analysis of SCZ vs. controls in H3K27a NeuN+ (blue), H3K27ac Tissue (green) and H3K27ac Meta NeuN+(orange) and BD vs. controls in H3K27ac Tissue (green). After peak annotation to genes by GREAT software, no of genes enriched/tested are shown in bold in parenthesis.

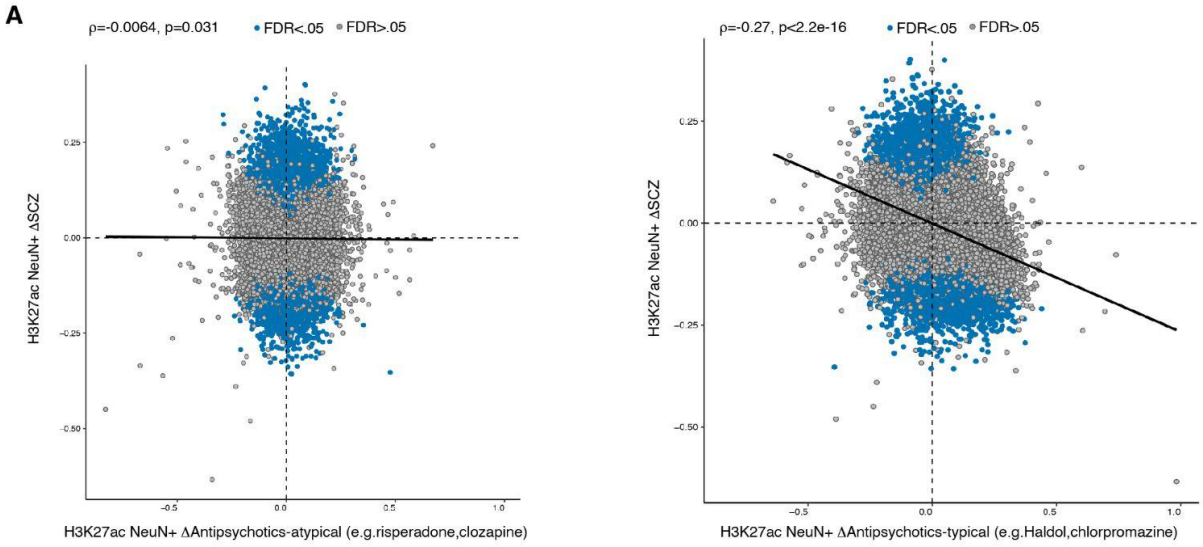

Figure S5 | Concordance analysis between SCZ dysregulated peaks and antipsychotics typical/atypical dysregulated peaks. Left plot shows the spearman correlation between **effect size that shows (A)** atypical antipsychotics and **(B)** typical antipsychotics associated differences with SCZ associated differences quantified in H3K27ac Tissue. Blue colored points in left(right) plots are significant at FDR 5% in atypical antipsychotics vs. controls(typical antipsychotics vs. controls) and SCZ vs. controls comparison.

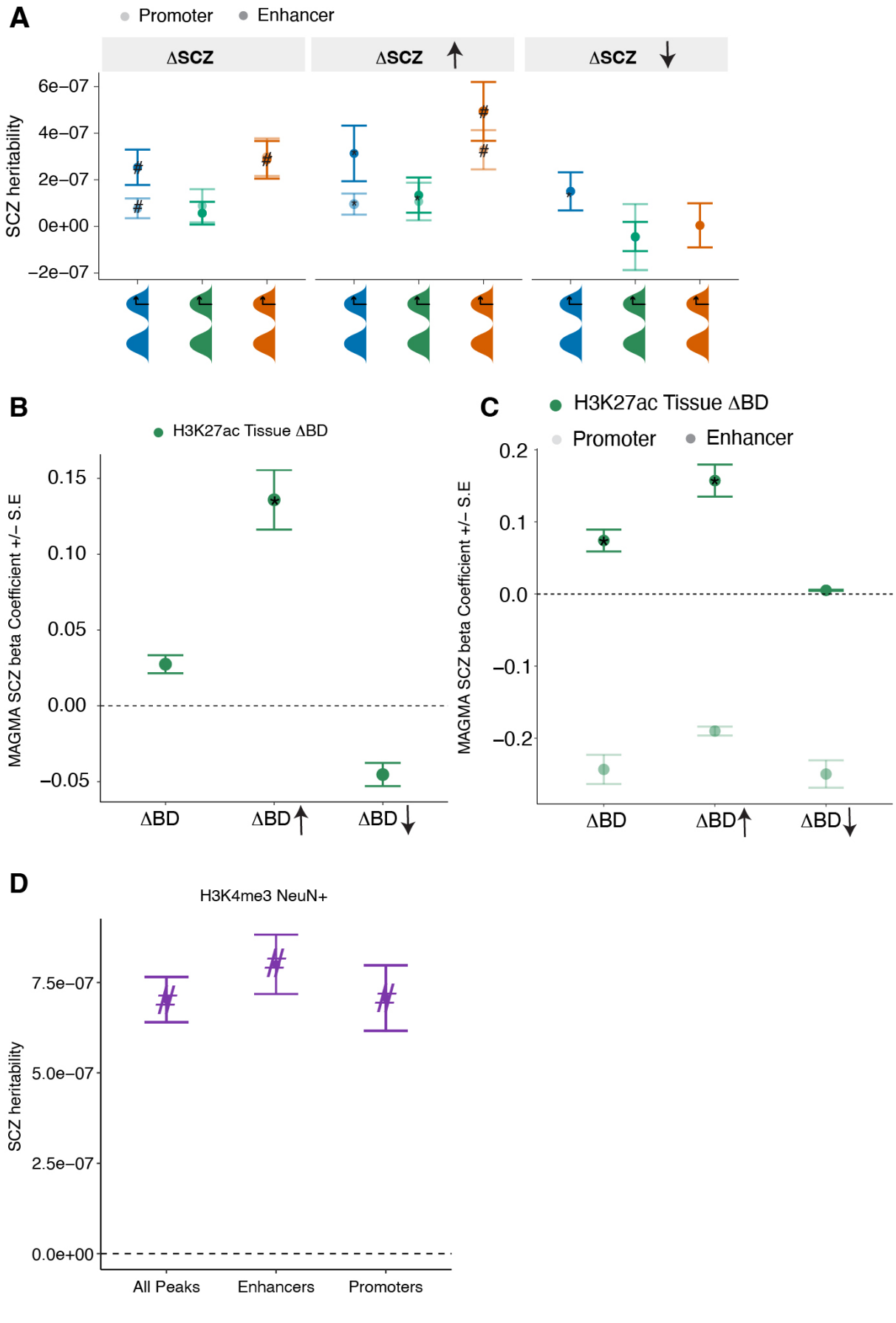
A
Promoter
Enhancer
ΔSCZ
ΔSCZ
ΔSCZ
SCZ heritability
6e−07
4e−07
2e−07
0e+00
−2e−07
B
H3K27ac Tissue ΔBD
MAGMA SCZ beta Coefficient +/− S.E
0.15
0.10
0.05
0.00
−0.05
ΔBD
ΔBD
ΔBD
C
H3K27ac Tissue ΔBD
Promoter
Enhancer
MAGMA SCZ beta Coefficient +/− S.E
0.2
0.1
0.0
−0.1
−0.2
ΔBD
ΔBD
ΔBD
D
H3K4me3 NeuN+
SCZ heritability
7.5e−07
5.0e−07
2.5e−07
0.0e+00
All Peaks
Enhancers
Promoters

Figure S6 | Enrichment of SCZ and BD dysregulated peaks in common risk variants of SCZ and BD. (A) SCZ heritability coefficients of common SCZ risk variants for each of the three cohorts H3K27a NeuN+ (blue), H3K27ac Tissue (green) and H3K27ac Meta NeuN+(orange), as indicated. From left to right 1) "ΔSCZ": dysregulated CRD (n=1883, 1377 and 2223 promoter and 9588, 4279 and 3996 enhancers for each cohort, respectively), 2) "ΔSCZ↑": hyperacetylated with log_2FC (SCZ vs. controls) > 0 (n=1183, 980 and 1969 promoter and , 4681, 1701, and 2039 enhancers for each cohort respectively) and 3) "ΔSCZ↓": hypoacetylated with log_2FC (SCZ vs. controls) < 0 (n= 700, 397, 0 promoters and 4907, 2578 and 1957 enhancers for each cohort, respectively) stratified by annotations as promoters (< ±3Kb from TSS) shown in lighter shade and enhancers (> ±3Kb from TSS). **(B)** Coefficients of enrichment of common SCZ risk variants using MAGMA in genes annotated to 1) "ΔBD": dysregulated peaks (n=1311 genes mapped to peaks), 2) "ΔBD↑": hyperacetylated with log_2FC (BD vs controls) > 0 (n=558 genes mapped to peaks) and 3) "ΔBD↓": hypoacetylated with log_2FC (BD vs controls) < 0 (n=772 genes mapped to peaks). **(C)** stratified by annotations as promoters (< ±3Kb from TSS) shown in lighter shade and enhancers (> ±3Kb from TSS) (n=177, 145 and 33 genes mapped to promoters and 1174, 530, 660 genes mapped to enhancers for ΔBD, ΔBD↑ and ΔBD↓ respectively). **(D)** SCZ heritability coefficients of all peaks (n=64,224), enhancer (n=43,683) and promoters (n=20,541) of H3K4me3 NeuN+ peaks. "#": Significant for enrichment (A, D) in LD score regression in and (B) in MAGMA after FDR correction of multiple testing across all tests in the plot (Benjamini & Hochberg); "*": Nominally significant for enrichment. Error bars show standard error in the beta coefficient from MAGMA (B,C) and coefficient of heritability from LDSc (A,D).

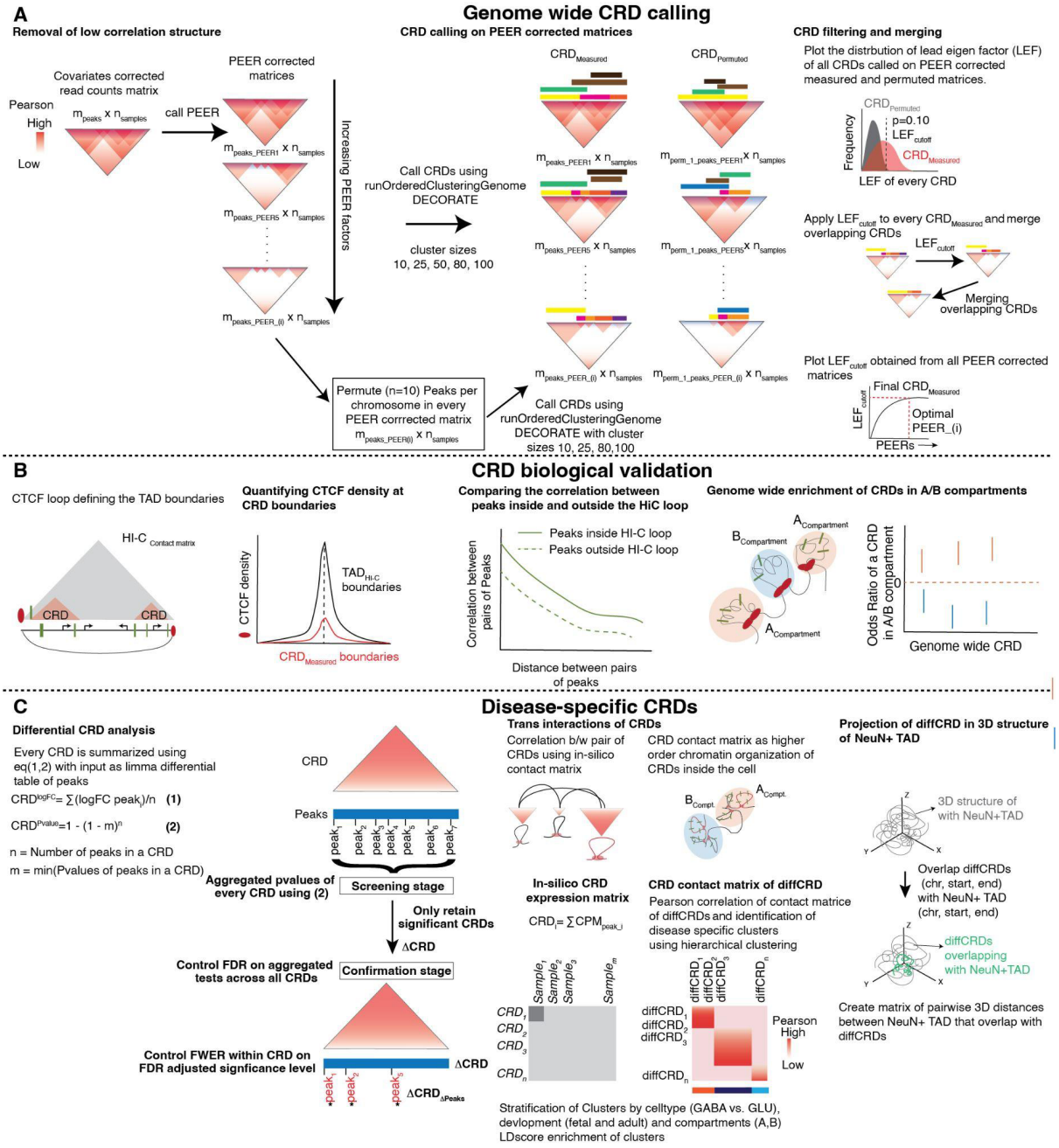

Figure S7 | Schematic workflow of the steps in CRD analyses. (A) Genome wide CRD calling includes data preprocessing, genome wide CRD calling and filtering and merging of CRDs. **(B)** *In-silico* biological validation of CRDs including the CTCF enrichment at the boundaries of CRDs, comparison of correlation of peaks that are inside and outside the HiC-loop and enrichment of CRDs in A and B compartments .**(C)** Identification of disease specific CRDs including two stage differential analysis of CRDs, CRD contact matrix from correlation of differential CRDs and finally projection of identified dysregulated CRDs in 3D genome and finally, 3D genome was simulated using *chrom3D*.

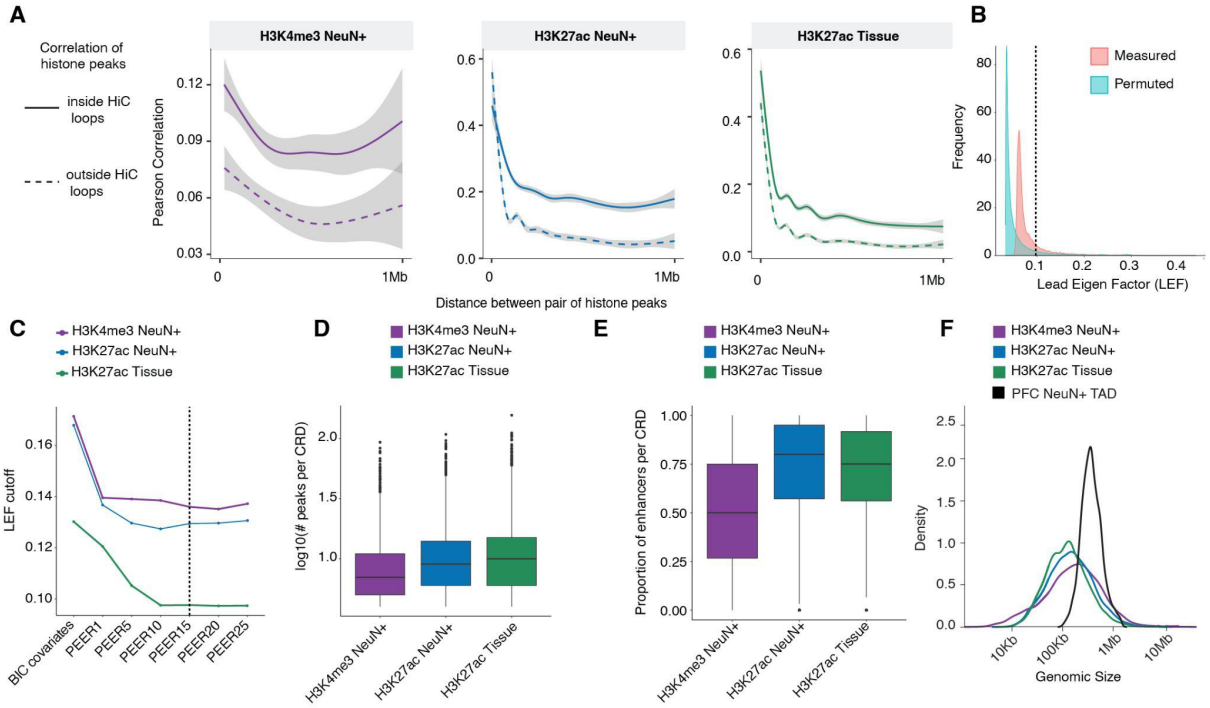

Figure S8 | CRD filtering and merging. (**A**) Genome-scale peak-to-peak correlations constrained by Hi-C loop and outside HiC-loop structures. **(B)** An example illustrating filtering step using distribution of Lead Eigen Factor (LEF) of CRDs called on $m_{peaks\_PEER} \times n_{samples}$ and $m_{Permutation\_i\_OCR\_PEER} \times n$ samples ($i=10$ permutations were performed). Final LEF_{cutoff} was obtained at P-value=0.10 from the distribution of LEF of permuted CRD. (B) Distribution of **(C)** Lead eigen factor of CRD called on $m_{peaks\_PEER} \times n_{samples}$ after filtering using LEF_{cutoff} obtained from the previous step (A), stratified by three datasets H3K4me3 NeuN+, H3K27ac NeuN+ and H3K27ac Tissue. **(D)** Boxplot depicting the log of number of peaks in CRDs called on $m_{peaks\_PEER15} \times n_{samples}$ matrices of H3K4me3 NeuN+ (n=25,031), H3K27ac NeuN+ (n=74,217) and H3K27ac Tissue (n=97,505). **(E)** Distribution of proportion of enhancers (peaks that are located > ±3Kb from TSS) within the CRDs called on $m_{peaks\_PEER15} \times n_{samples}$ matrices of H3K4me3 NeuN+ (n=12,509), H3K27ac NeuN+ (n=55,354) and H3K27ac Tissue (n=72,341). The center shows the median, the box shows the interquartile range, whiskers indicate the highest/lowest values within 1.5x the interquartile range, and potential outliers from this are shown as dots in the box plot in (D,E). (**F**) Genome-wide base pair length distribution (x-axis log scale 10^4-10^7 base pairs) of CRDs (colored curves) compared to neuronal TADs (black curve).

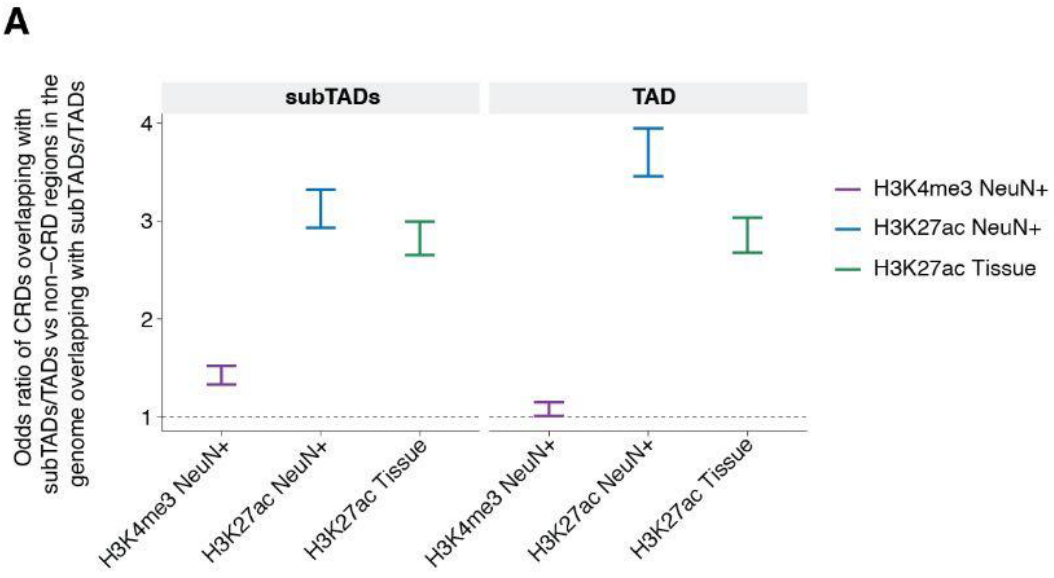

Figure S9 | Enrichment of CRDs inside subTADs and TADs. Concordance between sets of dysregulated CRDs across and within the studies. **(A)** Odds ratios of CRD genomic regions overlapping with subTADs/TADs vs. non-CRD regions.

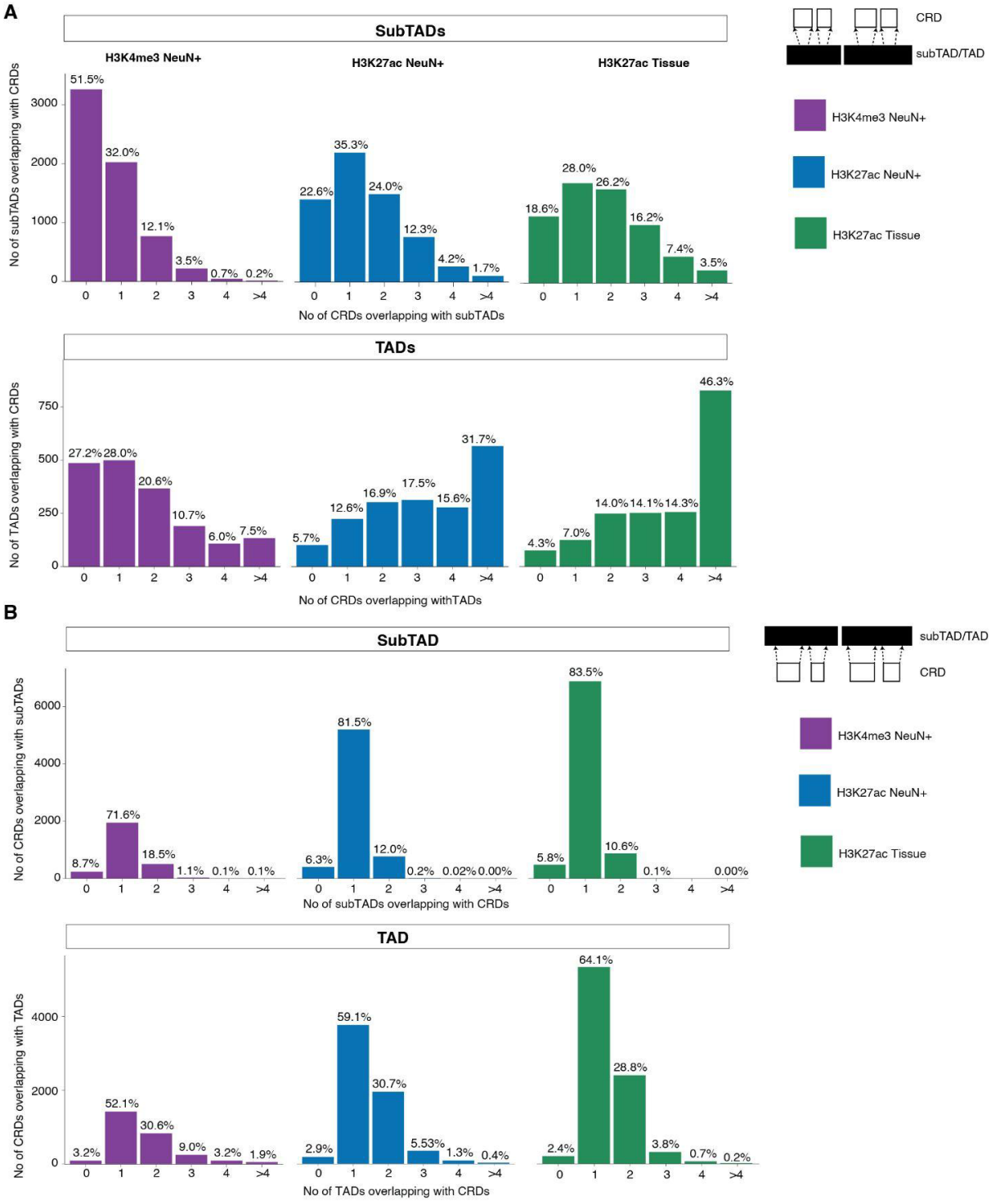

Figure S10 | Proportion of CRDs inside subTADs/TADs and vice-versa. (A) Number and percentage of CRDs overlapping with subTADs and TADs stratified by the number of overlapped subTADs and TADs. **(B)** Number and percentage of subTADs and TADs overlapping with CRDs stratified by the number of overlapped CRDs.

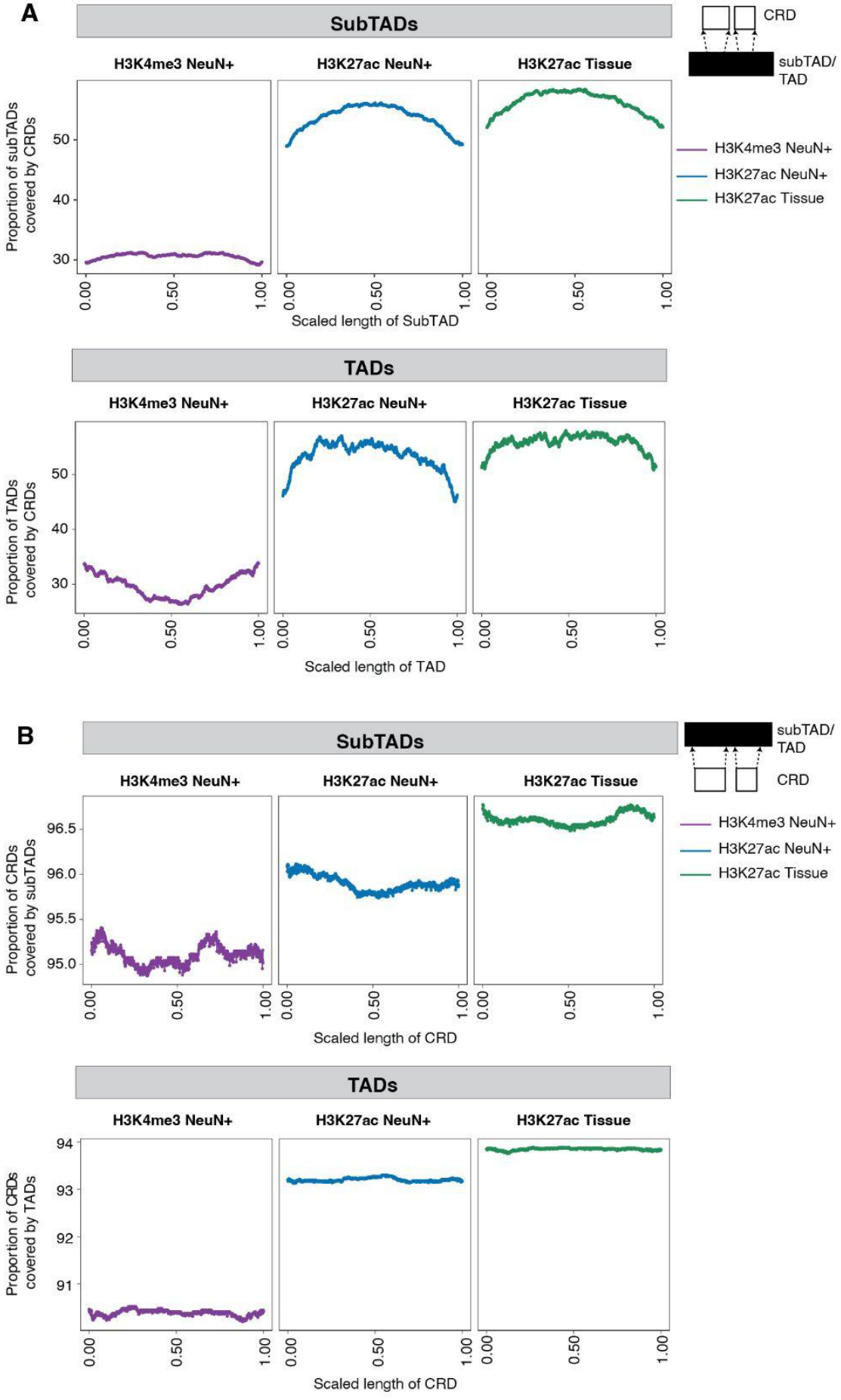
A
SubTADs
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
CRD
subTAD/
TAD
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
Proportion of subTADs covered by CRDs
50
40
30
0.00
0.50
1.00
Scaled length of SubTAD
TADs
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
Proportion of TADs covered by CRDs
50
40
30
0.00
0.50
1.00
Scaled length of TAD
B
SubTADs
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
subTAD/
TAD
CRD
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
Proportion of CRDs covered by subTADs
96.5
96.0
95.5
95.0
0.00
0.50
1.00
Scaled length of CRD
TADs
H3K4me3 NeuN+
H3K27ac NeuN+
H3K27ac Tissue
Proportion of CRDs covered by TADs
94
93
92
91
0.00
0.50
1.00
Scaled length of CRD

Figure S11 | Density of CRDs inside subTADs/TADs and vice-versa. Density of CRDs inside **(A)** subTADs/TADs and **(B)** CRDs inside subTADs and TADs. The width of each **(A)** subTADs, TADs and **(B)** CRDs have been normalized between 0 and 1 in 1K,10K and 40K equidistant points respectively.

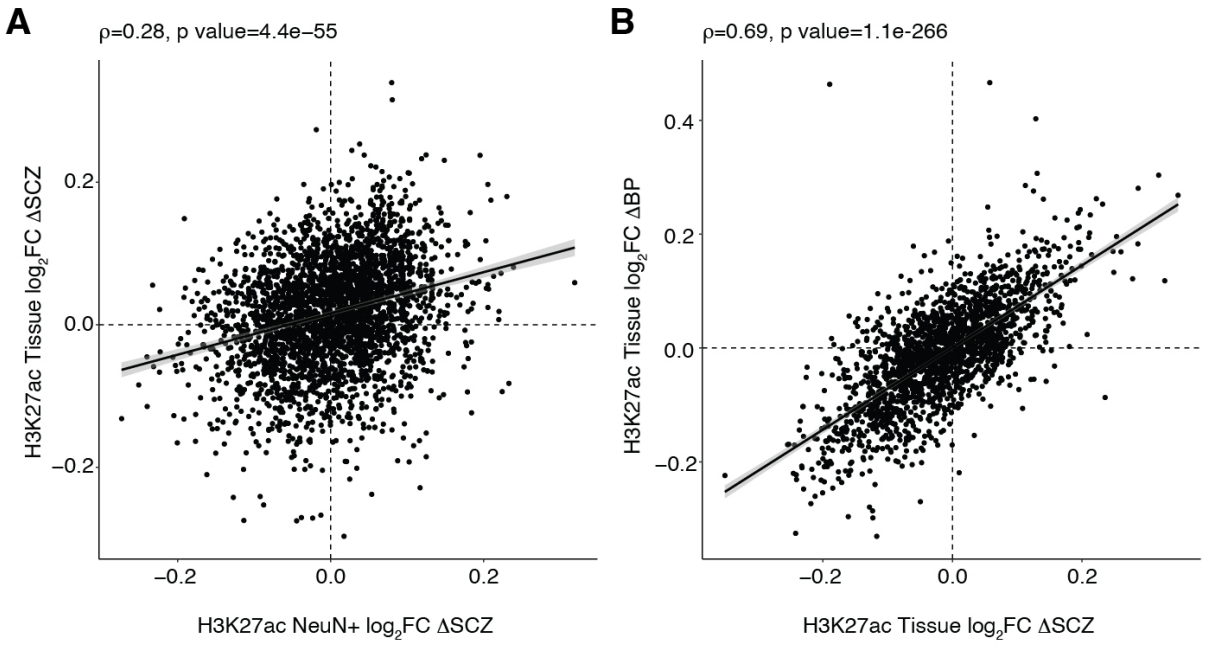

Figure S12 | Concordance between sets of dysregulated CRDs across and within the studies. (A) Left plot shows the spearman correlation between log_2FC (SCZ vs. controls) H3K27ac NeuN+ CRDs and log_2FC (SCZ vs. controls) H3K27ac Tissue at FDR 5% **(B)** Correlation between log_2FC (SCZ vs. controls) H3K27ac Tissue at FDR 5% and log_2FC (BD vs. controls) H3K27ac Tissue. P Values (p) in (A,B) computed for Spearman rank correlation.

A

Summary table of two step differential CRD analysis

Study	ΔCRD	ΔCRD_{Peaks}	$\Delta CRD_{\Delta Peaks}$
H3K27ac NeuN+ ΔSCZ	1,963	28,866	3,507
H3K27ac Tissue ΔSCZ	1,084	15,787	1,673
H3K27ac Tissue ΔBP	454	6,387	675

Figure S13 | Summary table of differential CRDs. Number of differential CRDs at FDR 5% and peaks inside them in ΔCRD_{Peaks} after stage -1 and $\Delta CRD_{\Delta Peaks}$ after stage-2 in all three groups.

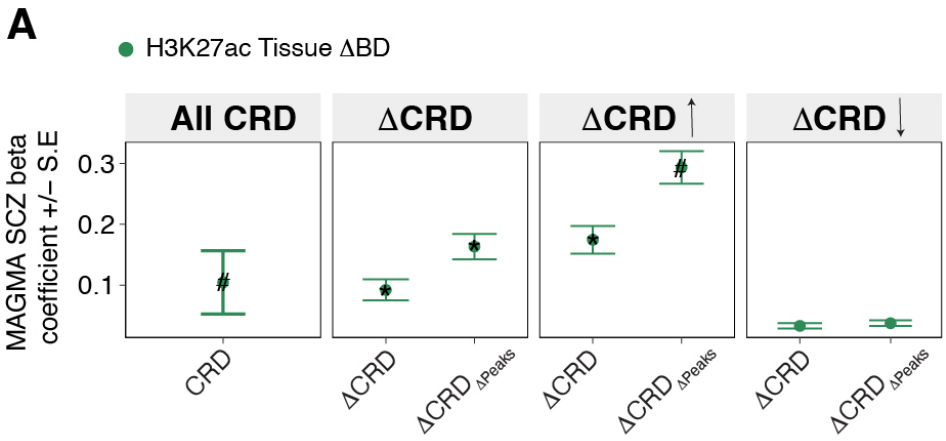

Figure S14 | Enrichment of dysregulated BD H3K27ac Tissue CRDs in common risk variants of SCZ. (A) Coefficients of enrichment of common SCZ risk variants using MAGMA by 1) "All CRD": all peaks inside CRD (n=20,644 genes), 2) "ΔCRD": dysregulated CRDs (n=1315 genes in ΔCRD and 463 genes in $\Delta CRD_{\Delta Peaks}$), 3) "ΔCRD↑": hyperacetylated with mean log_2FC (BD vs controls) > 0 (n=617 genes in ΔCRD, 214 genes in $\Delta CRD_{\Delta Peaks}$) and 4) "ΔCRD↓": hypoacetylated with mean log_2FC (BD vs controls) < 0 (n=704 genes in ΔCRD and 251 genes in $\Delta CRD_{\Delta Peaks}$) classified on x-axis as ΔCRD for all peaks and $\Delta CRD_{\Delta Peaks}$ for only dysregulated histone peaks. "#": Significant for enrichment in MAGMA after FDR correction of multiple testing across all tests in the plot (Benjamini & Hochberg); "*": Nominally significant for enrichment. Error bars show standard error in the beta coefficient from MAGMA.

A

Summary table of coefficients of glm models to interpret dyregulated peaks inside CRDs

Model	Dataset	$e^{estimate}$	Pvalue
Peak inside CRD ~ Peaks_Tstat^2	H3K27ac NeuN+ ΔSCZ	1.02	1.39e-27
	H3K27ac Tissue ΔSCZ	1.02	1.26e-23
	H3K27ac Tissue ΔBP	1.04	9.45e-55
$N_{\Delta Peak} \sim offset(log(N_{Peaks\_in\_CRD})) + N_{\Delta CRD}$	H3K27ac NeuN+ ΔSCZ	1.65	8.47e-11
	H3K27ac Tissue ΔSCZ	2.27	3.06e-13
	H3K27ac Tissue ΔBP	3.48	2.89e-08

Summary table of coefficients of glm models to interpret dyregulated genes inside CRDs

Model	Dataset	$e^{estimate}$	Pvalue
$N_{\Delta Genes} \sim offset(log(N_{Genes\_in\_CRD})) + N_{\Delta CRD}$	H3K27ac NeuN+ ΔSCZ	1.54	1.9e-4
	H3K27ac Tissue ΔSCZ	1.34	7.1e-3
	H3K27ac Tissue ΔBP	1.2	0.2605

ΔGenes: differential meta analysis of RNASeq from CMC cohort

Figure S15 | Summary table of coefficients of generalized linear model. (**A)** Coefficients and p values of the binomial and poisson generalized linear model estimated using the equation shown in the first column.

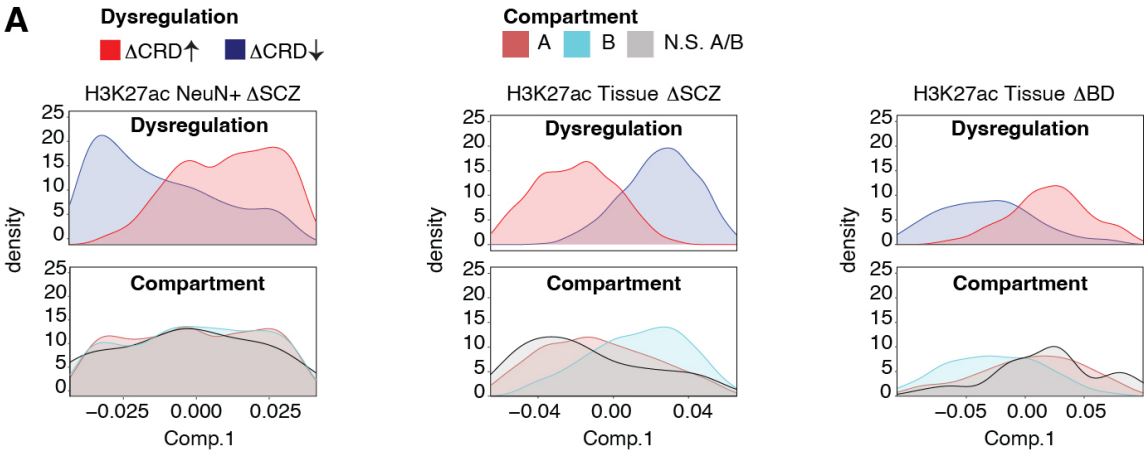

Figure S16 | Demonstration of nuclear organization in dysregulated CRDs. Density of dysregulated CRDs across principal component 1 (PC1) of ΔSCZ specific CRDs contact matrices from H3K27ac NeuN+ (left) and tissue (middle) and ΔBD specific CRDs contact matrices from H3K27ac tissue (right). The density of CRDs is stratified by "**Dysregulation**": hyperacetylated CRDs in red and hypoacetylated CRDs in blue and nuclear organization "**Compartment**": A(B) compartment CRDs in cyan(salmon) color. CRDs that showed no significant enrichment for A and B compartments are shown in gray and labeled as N.S. A/B.

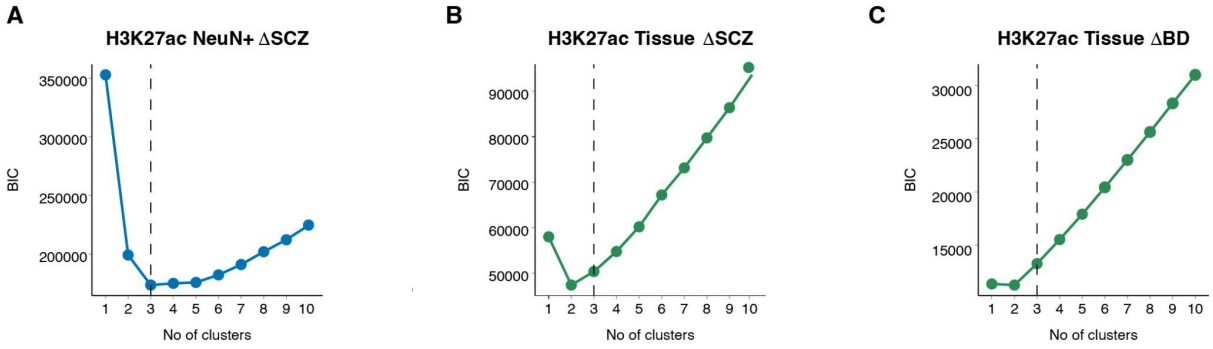

Figure S17 | Evaluation of optimal number of clusters in CRD contact matrices. Bayesian Information Criterion (BIC) using equation 2 as a function of number of Kmeans clusters on dysregulated CRD contact matrix of **(A)** H3K27ac NeuN+, **(B)** H3K27ac Tissue and **(C)** H3K27ac Tissue

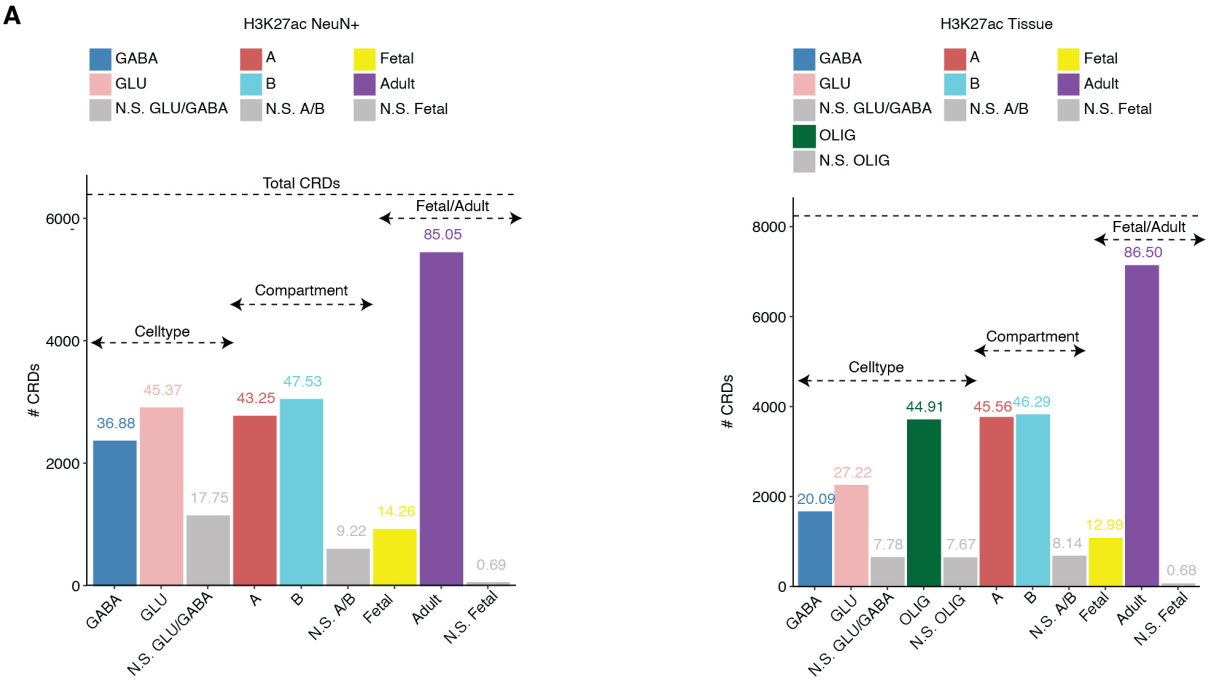

Figure S18 | Annotated CRDs. **(A)** H3K27ac NeuN+ (left) and H3K27ac Tissue (right) annotated CRDs are grouped into three categories: "**Celltype":** GLU (glutamatergic), GABA (gabaergic), OLIG (oligodendrocytes), "**Compartment**": A and B and "**Fetal/Adult"**: Fetal and Adults CRDs. CRDs with no significant enrichments for GLU/GABA, OLIG, A/B, Fetal are labeled as N.S. GLU/GABA, N.S. OLIG, N.S. A/B and N.S. Fetal respectively. The percentages above the bar plot are % of total CRDs annotated to a specific category. The dotted line shows the total number of CRDs.

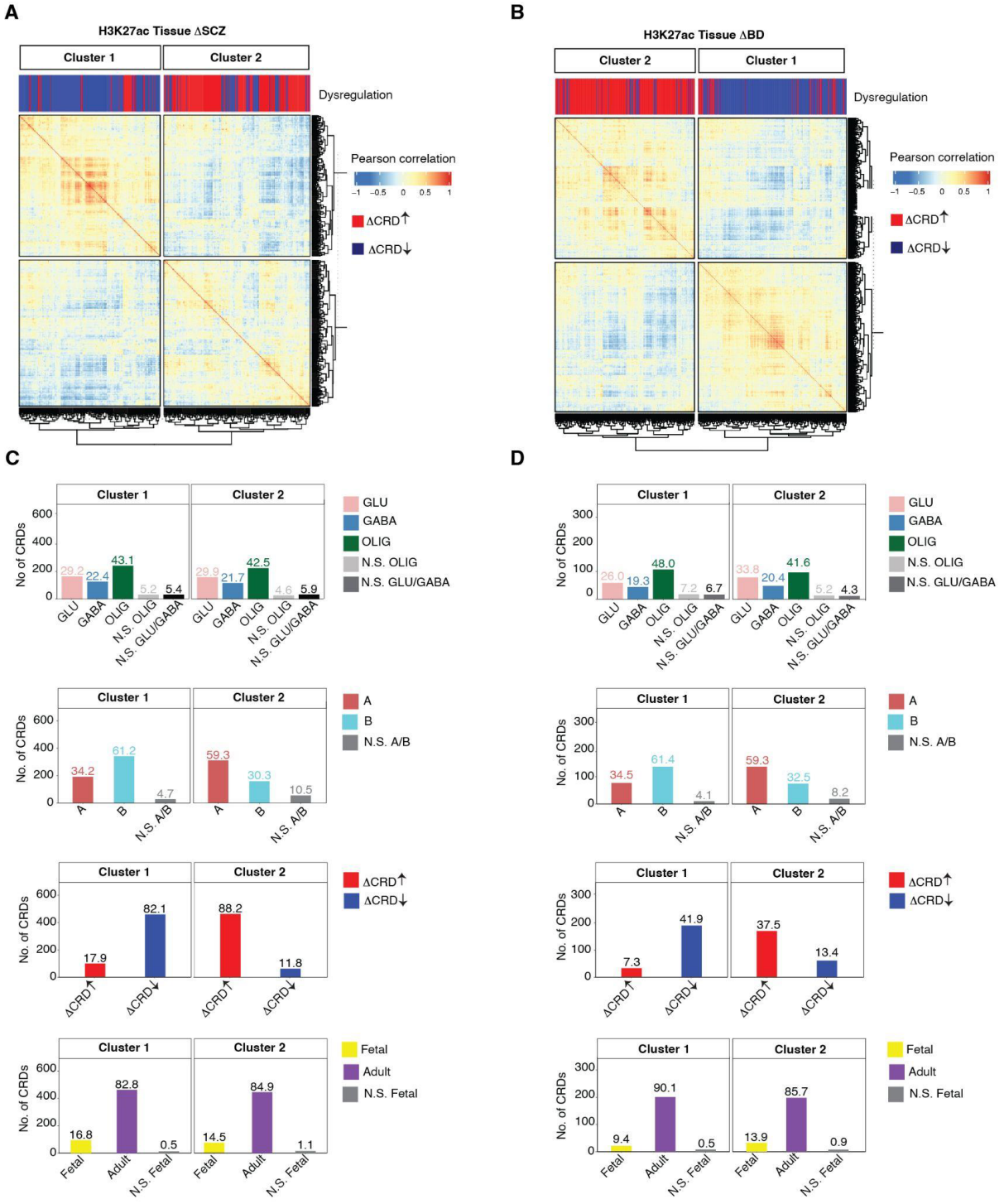

Figure S19 | Fingerprinting disease-sensitive CRDs in H3K27ac Tissue. (**A**) CRD contact matrices of SCZ and **(B)** BD sensitive in H3K27ac Tissue clustered into two large clusters; notice striking separation of cluster 1 and cluster 2 representing hypo in navy and hyperacetylated CRDs in red. Composition of **(C)** SCZ and **(D)** BD sensitive annotated CRDs in H3K27ac Tissue by cell type (GABA in light blue , GLU in pink), compartments (A in cyan, B in indianred.), dysregulation (hypo vs hyper-acetylation 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.

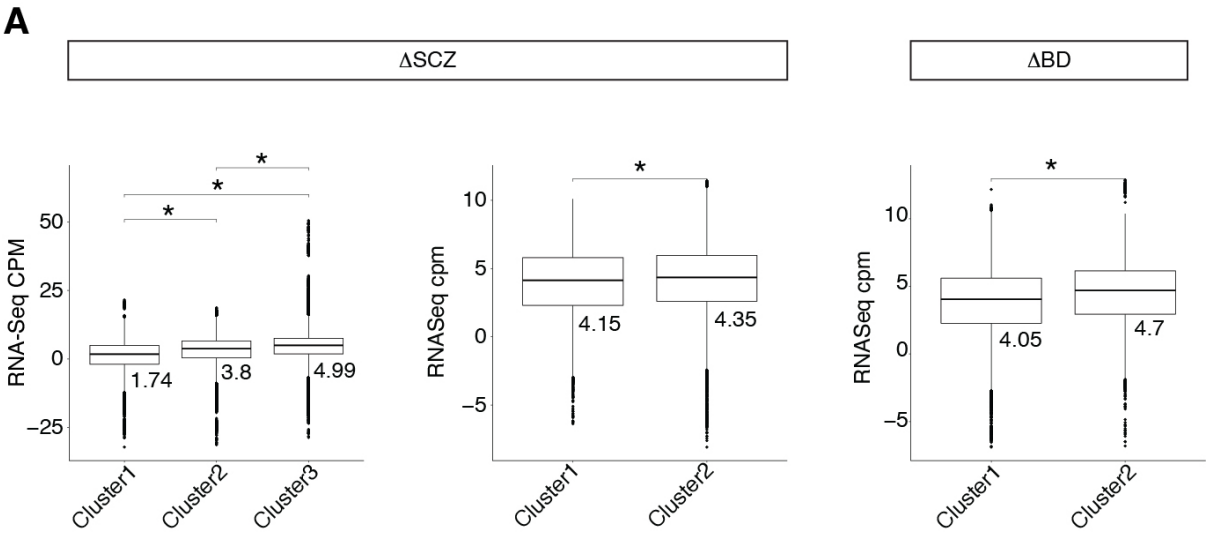

Figure S20 | RNAseq expression of disease-specific CRDs. (A) Box plot of RNA-Seq expression of genes that are within ΔSCZ H3K27ac NeuN+ (cluster1=1930, cluster2=1779, cluster3=3414 genes) and Tissue CRDs (cluster1=1430, cluster2=2078 genes) and ΔBD H3K27ac Tissue CRDs (cluster1=557, cluster2=765 genes) stratified by identified clusters of CRD contact matrices. The center shows the median, the box shows the interquartile range, whiskers indicate the highest/lowest values within 1.5x the interquartile range, and potential outliers from this are shown as dots. * represents p value < .05 while p values are estimated using Wilcoxon test.

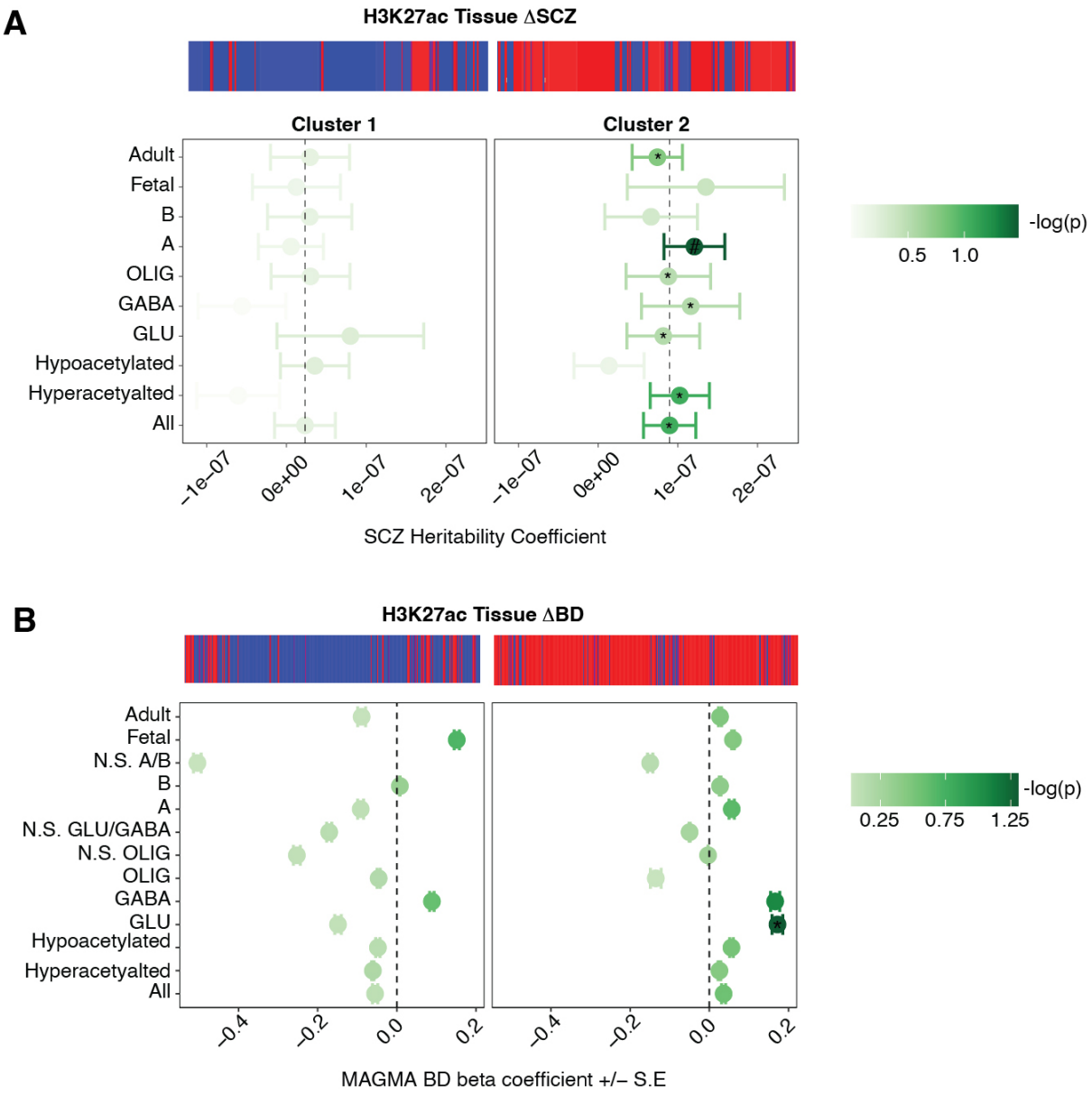

Figure S21 | Cluster 2 is enriched for SCZ and BD risk variants. (A) SCZ heritability coefficients of genetic variants overlapping histone peaks from SCZ sensitive CRDs in H3K27ac Tissue stratified by cluster 1 and 2. **(B)** Coefficients of enrichment of common BD risk variants using MAGMA in genes within BD sensitive CRDs. in H3K27ac Tissue stratified by cluster 1 and 2. The overlap of peaks and genes with genetic variants was assessed using LD score regression in **(A)** and MAGMA in **(B)**. "#": Significant for enrichment in LD score regression **(A)** and MAGMA **(B)** after FDR correction of multiple testing across all tests in the plot (Benjamini & Hochberg); "·": Nominally significant for enrichment

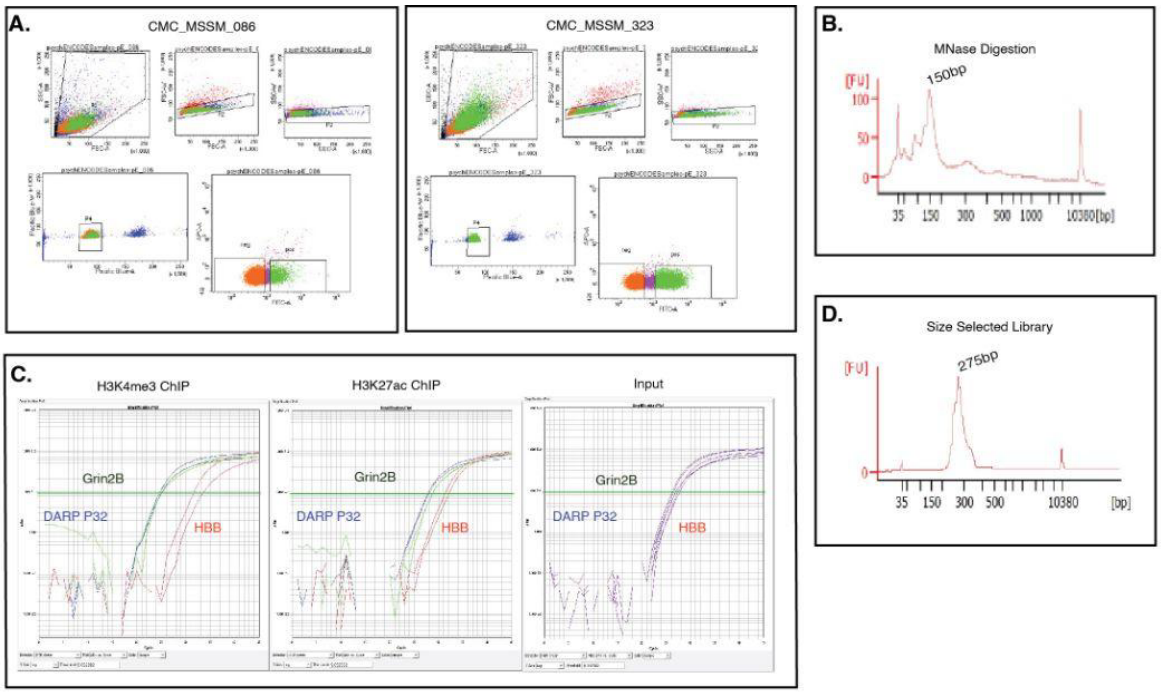

Figure S22 | Chromatin immunoprecipitation followed by deep sequencing quality controls. **(A)** Inspection and quantification of nuclei into neuronal and non-neuronal fractions with fluorescence activated nuclei cytometry (FANS) of brain samples CMC_MSSM_086 (left panel) and CMC_MSSM_323 (right panel). **(B)** Agilent Bioanalyzer 2100 track shows the digestion of native chromatin with MNase intro, primarily mononucleosomes (~150bp). **(C)** Quantitative Polymerase Chain Reaction (qPCR) data showed enrichment of histone marks H3K27ac and H3K4me3 with the neuronal genes GRIN2B(green curve) and DARP-P32 (blue curve), but not for negative control gene HBB(red curve). There was also equal enrichment of all three genes in the inputs (corresponding sample without histone antibodies). **(D)** Agilent Bioanalyzer 2100 track shows the final ChIP-Seq size selected library product with Sage Science Pippin HT with size 275 bp.

Supplementary Tables

Table S1: Metadata of samples in study-1 and study-2.

Table S2: Genomic coordinates of consensus peaks.

Table S3: Differential analysis of peaks across SCZ cases and controls and across BD cases and controls.

Table S4: GWAS enrichment of brain traits and non-brain related traits in peaks stratified by differentially upregulated and downregulated peaks.

Table S5: Genomic coordinates of identified CRDs.

Table S6: Annotation and differential analysis of CRDs.

Table S7: Pathway analysis of $\Delta CRD_{\Delta Peaks}$

Table S8: Genomic coordinates of H3K27ac Gabaergic, glutamatergic and oligodendrocytic specific peaks.