

## Supplementary Materials for

### Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation

Suhn K. Rhie, Shannon Schreiner, Heather Witt, Chris Armoskus, Fides D. Lay, Adrian Camarena, Valeria N. Spitsyna, Yu Guo, Benjamin P. Berman, Oleg V. Evgrafov, James A. Knowles, Peggy J. Farnham\*

\*Corresponding author. Email: [peggy.farnham@med.usc.edu](mailto:peggy.farnham@med.usc.edu)

Published 13 December 2018, *Sci. Adv.* **4**, eaav8550 (2018)  
DOI: 10.1126/sciadv.aav8550

#### The PDF file includes:

Supplementary Materials and Methods  
Fig. S1. CNON Hi-C datasets.  
Fig. S2. Chromatin interactions with different  $q$  value cutoffs.  
Fig. S3. H3K27ac ChIP-seq clustering results.  
Fig. S4. NOME-seq depth and number of NDRs.  
Legends for tables S1 to S11  
References (55–67)

#### Other Supplementary Material for this manuscript includes the following:

(available at [advances.sciencemag.org/cgi/content/full/4/12/eaav8550/DC1](https://advances.sciencemag.org/cgi/content/full/4/12/eaav8550/DC1))

Table S1 (Microsoft Excel format). Information about the CNON ChIP-seq, NOME-seq, and Hi-C datasets.  
Table S2 (Microsoft Excel format). CNON TADs identified by Hi-C.  
Table S3 (Microsoft Excel format). CNON ChIP-seq peaks classified as promoters, enhancers, insulators, repressed, or heterochromatin regions.  
Table S4 (Microsoft Excel format). CNON chromatin interactions identified by Hi-C.  
Table S5 (Microsoft Excel format). Epigenomic classification of CNON Hi-C chromatin interactions.  
Table S6 (Microsoft Excel format). Predicted target genes of enhancers using CNON Hi-C.  
Table S7 (Microsoft Excel format). Coordinates of promoter H3K4me3, non-promoter H3K27ac, and CTCF peaks found in many individuals of CNON datasets.  
Table S8 (Microsoft Excel format). Identification and classification of NDRs identified by CNON NOME-seq.

Table S9 (Microsoft Excel format). Enhancers linked to genotype, gender, smoking status, and schizophrenia diagnosis.

Table S10 (Microsoft Excel format). Classification and characterization of SCZ risk-associated variants.

Table S11 (Microsoft Excel format). Predicted target genes of enhancers linked to individual variations.

## **Supplementary Materials and Methods**

### **Study Design**

Research objectives: As part of the PsychENCODE Consortium (5U01MH103346; [psychencode.org](http://psychencode.org)), the objectives of this study were to develop a 3D epigenomic map of primary Cultured Neuronal cells derived from Olfactory Neuroepithelium (CNON) and to investigate individual differences in epigenetic patterns at regulatory elements.

Research subjects: Nasal biopsy tissue samples were collected by recruitment by a previous NIH-funded GPC study (1R01MH085548) and recruitment through Los Angeles county/University of Southern California outpatient psychiatric clinics. In total, the previous study collected samples from 144 patients with DSM-IV criteria for schizophrenia (SCZ) and 111 controls (no psychiatric disorders and no family history of SCZ) after obtaining informed consent. This study was institutional-review board (IRB) approved by the University of Southern California (USC). Tissue samples were collected from the most superior region of the nasal spectrum mucosa. A detailed information on nasal biopsy process can be found from a previous study (1). In our study, we used CNON cells that were successfully established from biopsies from 63 living individuals.

Experimental Design: The overall design of the project was to perform Hi-C to detect chromatin interactions, ChIP-seq to detect regulatory elements on a genome-wide scale, and NOMe-seq to identify transcription factor binding sites within nucleosome-depleted regions contained within larger regulatory elements. For comparison to the data collected in this study, we used transcriptome and genomic sequencing for collected samples from a previous study (2). See table S1 for information concerning accessibility of new and previously collected data. The programs

and analytical methods used to identify and classify chromatin interactions, ChIP-seq, and NOME-seq peaks are described in detail below. For analysis of individual variation of the active state of regulatory elements, ChIP-seq peak strength was compared to genotype, smoking status, gender, and disease diagnosis; the statistical methods used for each of these comparisons are described below.

**Cell Culture:** CNON cultures were previously derived from nasal biopsy tissue samples (1, 55).

We used CNON cultures from a total of 63 individuals for this study. Metadata for each individual can be found in psychencode.org or table S1. CNON were cultured at 37°C with 5% CO<sub>2</sub> in plates coated with 1:10 Matrigel matrix (Corning, Inc. Cat# 354234) and 4506 media, as previously described (1). A detailed cell culture protocol for CNON can be found at <http://farnhamlab.com/protocols.htm> (3).

**ChIP-seq:** CNON ChIP assays were performed using an antibody to H3K4me3 (Cat# 9751S Lot# 7 or Lot#8, Cell Signaling Technology, Inc.), H3K4me1 (Cat# C15410037 Lot# A1657D, Diagenode, Inc.; Cat# 39297 Lot# 1714002, Active Motif), H3K27ac (Cat# 39133 Lot# 1613007, Active Motif), H3K27me3 (Cat# 9733 Lot# 8), H3K9me3 (Cat# 13969 Lot# 1), or CTCF (Cat# 23913002 Lot# 34614003, Active Motif, or Cat# 3418 Lot#1, Cell Signaling and Technology, Inc.), as previously described (5, 27); a detailed ChIP-seq protocol can be found at <http://farnhamlab.com/protocols.htm>. All antibodies used were validated according to ENCODE standards; the antibody validation documents are available at <https://www.synapse.org/#!/Synapse:syn3326332> as part of PsychENCODE. Each ChIP-seq library was sequenced on either Illumina Hiseq 2000 or Nextseq 500 machines (SE 100bp reads) to produce fastq files. All ChIP-seq data were mapped to hg19 and peaks were called using MACS2 (56) against the pooled inputs from 12 libraries after preprocessing data with the

ENCODE3 ChIP-seq pipeline (<https://www.encodeproject.org/chip-seq/>). The quality of each ChIP-seq library was assessed and all datasets passed ENCODE3 standards (<https://www.encodeproject.org/data-standards/>). To call reproducible peaks from two different individuals (SEP044 and SEP045), the IDR tool (<https://github.com/nboleyle/idr>) for CTCF datasets or the naïve overlap tool for histone mark datasets was used, as suggested in the ENCODE3 ChIP-seq standards document (<https://www.encodeproject.org/pages/pipelines/>). All ChIP-seq datasets with complete metadata information were deposited in Synapse as part of the PsychENCODE consortium ([www.psychENCODE.org](http://www.psychENCODE.org)). ChIP-seq datasets used in this study is detailed in table S1. Coordinates of H3K4me3, H3K27ac, and CTCF peaks found in CNON from many individuals (n=56, 47, 33 for H3K4me3, H3K27ac, and CTCF, respectively) can be found in table S7.

**Chromatin states classification:** After identifying regions enriched for each of the histone modifications and/or CTCF using the set of reproducible peaks from two different individuals, a sequential classification scheme was used to sort the elements into different categories: 1) H3K4me3 peaks within +/- 2kb from a known TSS (obtained from Gencode version 19, [https://www.gencodegenes.org/human/release\\_19.html](https://www.gencodegenes.org/human/release_19.html)) were classified as active promoters (n=17,427); 2) H3K27ac peaks not contained within the set of promoters were classified as active enhancers (n=42,407); 3) CTCF peaks not overlapping with the sets of promoters or enhancers were classified as insulator elements (n=27,588); 4) regions bound by H3K27me3 and not identified as an active promoter, enhancer, or insulator were classified as repressed elements (n=343,072); and 5) regions bound by H3K9me3 and not identified as active promoter, enhancer, insulator or repressed element were classified as heterochromatin (n=165,883); see table S3.

**Hi-C:** In situ Hi-C experiments were performed in CNON following the original protocol by Rao et al (17) with minor modifications listed here. Hi-C was performed in duplicate and  $3 \times 10^6$  cells were used for each experiment. 100U of MboI restriction enzyme (NEB, R0147) was used to digest the chromatin. For ligation, 2000U of T4 DNA Ligase (NEB, M0202) was added and samples were incubated at room temperature for 4 hours with slow rotation. Hi-C material was sheared to a size of 300-500bp using a Covaris instrument (Covaris S2, Woburn, MA). Biotin-tagged DNA was pulled down using Dynabeads MyOne Streptavidin C1 beads (Life technologies, 65002) with 2X Binding Buffer (2X BB: 10mM Tris-HCl (pH 7.5), 1nM EDTA, 2M NaCl). Hi-C libraries were amplified with 14 cycles of PCR using Illumina primers. Each library was sequenced (100bp paired-end) to produce ~900M read pairs using an Illumina HiSeq 2000 (table S1). The quality of each Hi-C library was checked with FastQC (57), HICUP version 0.5.8 (58), and HiC-Pro version 2.8.0 (51) (fig. S1A-B). Raw fastq files were processed through the HiC-Pro version 2.8.0 (51) to make the raw contact count matrices for multiple resolutions (e.g. 5kb, 10kb, 20kb, 40kb, or 100kb). The matrices were normalized using the iterative correction method (iced python library).

**TAD characterization:** For the identification of topologically associating domains (TADs), the normalized Hi-C contact matrices, binned with 40kb resolution, were used and the domains were called using the TopDom program (9); see table S2. SEP044 Hi-C and SEP045 Hi-C datasets were compared after down-sampling the SEP044 Hi-C library to have the same number of reads as the SEP045 Hi-C library. Normalized contact count matrices at 100kb resolution were used to calculate Pearson correlation (fig. S1C). The size and distribution of TADs across the genome were measured and compared (fig. S1D-E). The overlap analysis of TADs found from each library was performed using the following rules: if the boundaries of the domains are within

80kb windows (2 blocks) or the domains are overlapped  $\geq 80\%$  between libraries, we considered that these TADs overlapped (fig. S1F). The Vennerable R package ([http://download.r-forge.r-project.org/src/contrib/Vennerable\\_3.0.tar.gz](http://download.r-forge.r-project.org/src/contrib/Vennerable_3.0.tar.gz)) was used to generate weighted Venn diagrams. To identify heterochromatic TADs, we defined H3K9me3-marked TADs having more than 50% of their genomic region marked by H3K9me3 peaks. To compare gene expression in the 830 H3K9me3-marked SEP044 TADs to the 5,970 other TADs identified in SEP044 (Fig. 2C), RNA-seq data for the SEP044 CNON cells was obtained from Evgrafov et al (2). To plot ChIP-seq marks at the TAD boundaries, CTCF, H3K4me3, H3K9me3, and H3K27me3 ChIP-seq tags were normalized using the MakeTagDirectory script from HOMER (<http://homer.ucsd.edu/homer/>) (59). Normalized ChIP-seq tags at the TAD boundaries were used to make average plots using the annotatePeaks.pl script (hist parameter 10kb, size parameter 500kb) from HOMER.

**Chromatin interactions:** Intra-chromosomal significant interactions (50kb-10Mb range) from each of the Hi-C datasets were called after adjusting local background using Fit-Hi-C (16); see table S4 (10kb resolution,  $q$ -value  $< 1e-12$ ). Hi-C datasets were then pooled to increase the coverage and allow a higher resolution (5kb resolution), enabling the identification of more promoter-enhancer chromatin interactions. Intra-chromosomal significant interactions (50kb-10Mb range) called with different  $q$ -value cutoffs using Fit-Hi-C (16) were classified (fig. S2). The anchors of the interactions were intersected with the chromatin states identified as described above and the top 25 most frequent chromatin interaction categories are shown in Fig. 3B and fig. S2. Hi-C chromatin interaction heatmaps were visualized using the HiTC R package (60) with 10kb resolution for Fig. 2A-B. RNA-seq data (2) for the SEP044 CNON cells was used to

plot normalized gene expression levels for the potential target genes found from different categories of promoter-enhancer interactions (Fig. 3E).

**Comparison of CNON ChIP-seq datasets:** Using the DiffBind R package (52), ChIP-seq signals were normalized across individual samples (setting minOverlaps to 2 in the dba.count function) at identified ChIP-seq peak regions. A correlation coefficient was calculated using the dba.plotHeatmap function across the ChIP-seq libraries. In addition to the ENCODE3 quality control measurements described above in the ChIP-seq section, a correlation coefficient/clustering analysis was also performed to identify outlier datasets (fig. S3). High-quality ChIP-seq datasets (n=56, 47, 4, 22, and 11 for H3K4me3, H3K27ac, H3K4me1, H3K27me3, and H3K9me3, respectively) were used for this study (table S1). To determine the genomic distribution of ChIP-seq peaks, the set of all TSS for known genes was obtained from Gencode (version 19, [https://www.gencodegenes.org/human/release\\_19.html](https://www.gencodegenes.org/human/release_19.html)); promoters were defined as  $\pm 2$ kb windows from each TSS and non-promoters were defined as  $> 2$ kb from each TSS. Bedtools (<https://github.com/arq5x/bedtools2>) was used to intersect genomic regions and identify ChIP-seq peaks shared among individuals; sets of 72,937 H3K4me3 peaks at promoters ( $\pm 2$ kb from each TSS), 254,705 H3K27ac peaks at non-promoters, and 754,059 CTCF peaks (at either promoter or non-promoter regions) were identified to be present in at least one of the individual CNON samples. Sets of 16,711 H3K4me3 peaks at promoters ( $\pm 2$ kb from each TSS), 61,419 H3K27ac peaks at non-promoters, and 62,433 CTCF peaks (at either promoter or non-promoter regions) were identified to be present in at least 20% of the individual CNON samples. Normalized ChIP-seq tags at non-promoter H3K27ac peaks found in the SEP044 sample for all individuals (100%), at least 20% individuals, less than 20% individuals, and only in SEP044



sample were used to make average plots in Fig. 4C using the `annotatePeaks.pl` script (hist parameter 10bp, size parameter 4kb) from HOMER (<http://homer.ucsd.edu/homer/>) (59).

**NOMe-seq**: To perform Nucleosome Occupancy and Methylome Sequencing (NOMe-seq), chromatin was treated with the M.CviPI methyltransferase, which methylates GpC dinucleotides that are not protected by nucleosomes or other proteins that are tightly bound to the chromatin (GpC<sup>m</sup> does not occur in the human genome and therefore there is no endogenous background of GpC<sup>m</sup>). Following bisulfite treatment of the M.CviPI-methylated chromatin (which converts unmethylated Cs to Ts and thus allows the distinction of GpC from GpC<sup>m</sup>) and subsequent genomic sequencing, Nucleosome-Depleted Regions (NDRs) were ascertained on a genome-wide scale. The bisulfite treatment also allows the distinction of CpG from C<sup>m</sup>pG and thus the endogenous methylation status of the genome was also obtained in the same sequencing reaction. CNON NOMe-seq data from 5 different individuals was generated as previously described and sequenced between 700 million and 1.1 billion reads (paired end, 100bp) using an Illumina HiSeq 2000 to produce fastq files (3). Each fastq file was aligned to a bisulfite-converted genome (hg19) using BSMAP (53) and processed as previously described (3, 5). To identify the methylation status of CpG sites (in all HCG trinucleotides) and GpC sites (in all GCH trinucleotides) from the bam file, the Bis-SNP (61) program was used. For identification of NDRs (table S8), the `findNDRs` function in the `aaRon` R package (<https://github.com/astatham/aaRon>) was used with the  $p$ -value cutoff  $10^{-15}$ . To determine the relationship between sequencing depth and the number of NDRs, we deeply sequenced the SEP030 NOMe-seq library, which has the least number of duplicate reads. We found that the sequencing depth is highly correlated with the number of identified NDRs (especially for non-promoter NDRs) (fig. S4). Therefore, the union set of the NDRs was identified using NDRs from

the most deeply sequenced library (SEP030; sequenced at 1.1 billion read pairs) and the NDRs found in at least two of the other 4 libraries, identifying a set of 166,240 NDRs. NDRs were compared with H3K4me3, H3K27ac, and CTCF ChIP-seq peaks that were found in at least 20% of the individuals (see above Comparison of CNON ChIP-seq datasets section). We identified 45,793 promoter NDRs, 58,059 insulator NDRs, 74,323 enhancer NDRs, and 28,657 NDRs that did not overlap with the 3 classes of active regulatory elements (which we called NDRs without features) (table S8). Bis-tools (<https://github.com/dnaase/Bis-tools>) (62) was used to visualize DNA methylation and accessibility signals in average plots and heatmaps with clustering. A detailed NOME-seq protocol and analysis pipeline can be found in <http://farnhamlab.com/protocols.htm> (3).

**Visualization of ChIP-seq and NOME-seq data:** All of the generated ChIP-seq and NOME-seq libraries were inspected by visualizing ChIP signal tracks in the genome browsers. To visualize ChIP-seq and NOME-seq data and generate the genome browser screenshots, the Integrative Genomics Viewer (IGV) (63) and UCSC Genome browser (<https://genome.ucsc.edu>) were used.

**Motif Analysis:** To discover motifs, we scanned known motifs at promoter NDRs, insulator NDRs, enhancer NDRs, and NDRs without features using the findMotifsGenome.pl script from HOMER (<http://homer.ucsd.edu/homer/>) (59). The top 50 enriched motifs in CNON enhancer NDRs (n=74,323) were identified by ranking motifs with *q*-values after filtering ones found in less than 5% of the NDRs. The presence of the top 50 enriched motifs in CNON enhancer NDRs was annotated using a binary matrix. We found 8 enhancer NDR clusters using a binary method for distance matrix computation and Ward's method for hierarchical clustering with the binary matrix. A heatmap was created using the fastcluster (<http://danifold.net/fastcluster.html>) and heatmap.3 R packages (<https://cran.r-project.org/web/packages/heatmap3/index.html>).

### **Identification of differentially enriched enhancers linked to genotypic or phenotypic**

**classifications:** H3K27ac ChIP-seq signals for 40 samples that have matched genotype and RNA-seq data were normalized using the DiffBind R package (52). From these samples, 254,705 non-promoter H3K27ac sites were used to find differentially enriched enhancers linked to genotypic or phenotypic classifications. We identified 165,482 H3K27ac sites that had at least one variant underneath the ChIP-seq peak; 6,726,766 variants from the available genotype data (2) were used to fit a linear regression model between enhancer and genotype. In total, we identified 4,167 unique H3K27ac sites that showed a correlation of peak strength with 8,847 unique SNPs (multiple testing adjusted  $p$ -value  $< 0.05$ ). To identify differentially enriched H3K27ac sites linked to phenotypic classifications (gender, smoking, or schizophrenia), we used the `dba.contrast` function from the DiffBind R package (52). After identifying statistically significantly associated H3K27ac sites by gender, smoking, or schizophrenia (FDR  $< 0.05$ ), we measured empirical  $p$ -values after generating 100 datasets with randomly scrambled labels for each phenotypic classification. The list of H3K27ac peaks (using an FDR  $< 0.05$  and empirical  $p$ -value  $< 0.05$ ) linked to each classification can be found in table S9.

### **Post Genome-Wide Association Studies (GWAS) Analysis of schizophrenia risk-related**

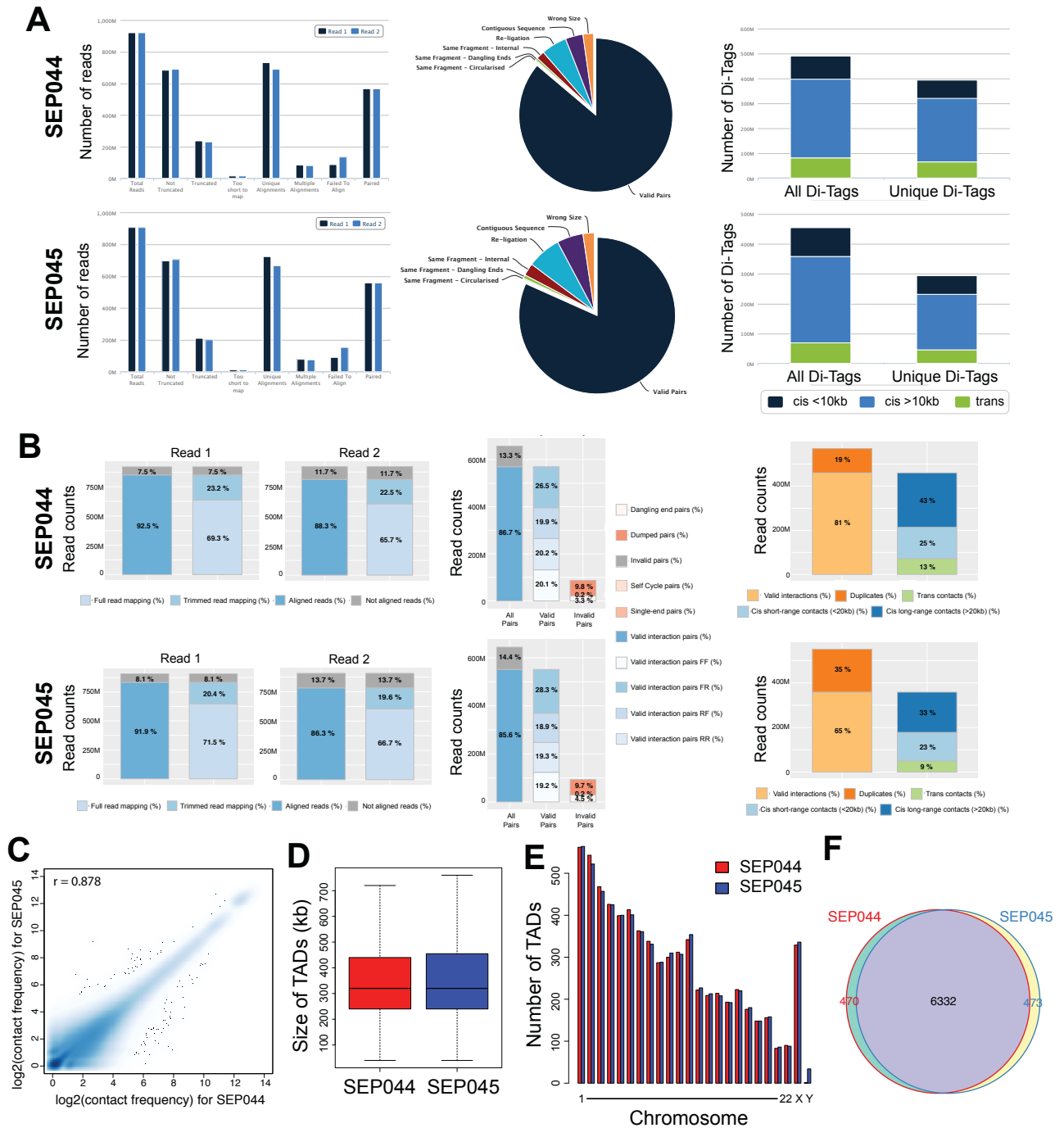
**SNPs:** To identify variants that are associated with increased risk for schizophrenia (SCZ) and are located in CNON regulatory elements, we first collected information concerning schizophrenia risk loci and index variant information from PGC (39) and Li et al (40). Using the FunciSNP R package (48), we identified 3,403 high LD variants ( $R^2 > 0.5$  with index variants, matching the ethnic group for GWAS) that are located within 1Mb of an index variant and within a CNON biofeature; 72,937 H3K4me3 peaks at promoters ( $\pm 2$ kb from each TSS), 254,705 H3K27ac peaks at non-promoters, 754,059 CTCF peaks (at either promoter or non-promoter

regions), and 166,240 NDRs. Among these 3,403 SNPs, 2,891 were within either the same TAD or CTCF-CTCF interaction loops as the index variants. We used the motifbreakR program (54) to identify transcription factor motifs that are overlapped by the 215 SCZ risk-related SNPs that fall within an NDR of a regulatory element using motif databases from ENCODE (64), Factorbook (65), Hocomoco (66) and HOMER (<http://homer.ucsd.edu/homer/motif/motifDatabase.html>) (59). The motifs for which TF binding is predicted to be significantly altered by the allele changes (using a score threshold of 0.9) are listed in table S10. We also identified TFs previously reported to be bound in the regulatory elements where the 215 variants reside (100bp windows of a variant), using ENCODE TF ChIP-seq datasets (67); see also (<http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeRegTfbsClustered/>) (table S10). SCZ risk-related SNPs that fall within H3K4me3, H3K27ac and CTCF peaks were used to fit a linear regression model between ChIP-seq signal and genotype, as described above (see above Identification of differentially enriched enhancers linked to genotypic or phenotypic classifications section).

**Predicting target genes of enhancers:** By overlapping chromatin interaction anchors with the chromatin states identified as described above (using reproducible peaks found in SEP044 and SEP045), we identified 12,970 Hi-C promoter-enhancer interactions (5kb resolution) which included 5,623 genes. We then searched within CTCF-CTCF interaction loops for co-localizing H3K4me3-marked promoters and non-promoter H3K27ac pairs; from this analysis, we also predicted an additional 12,120 putative target genes from ~700k promoter-enhancer pairs. To predict target genes of enhancers found in at least one of the 47 individuals, we also searched within TADs for co-localizing H3K4me3-marked promoters (from a set of 72,937 H3K3me3

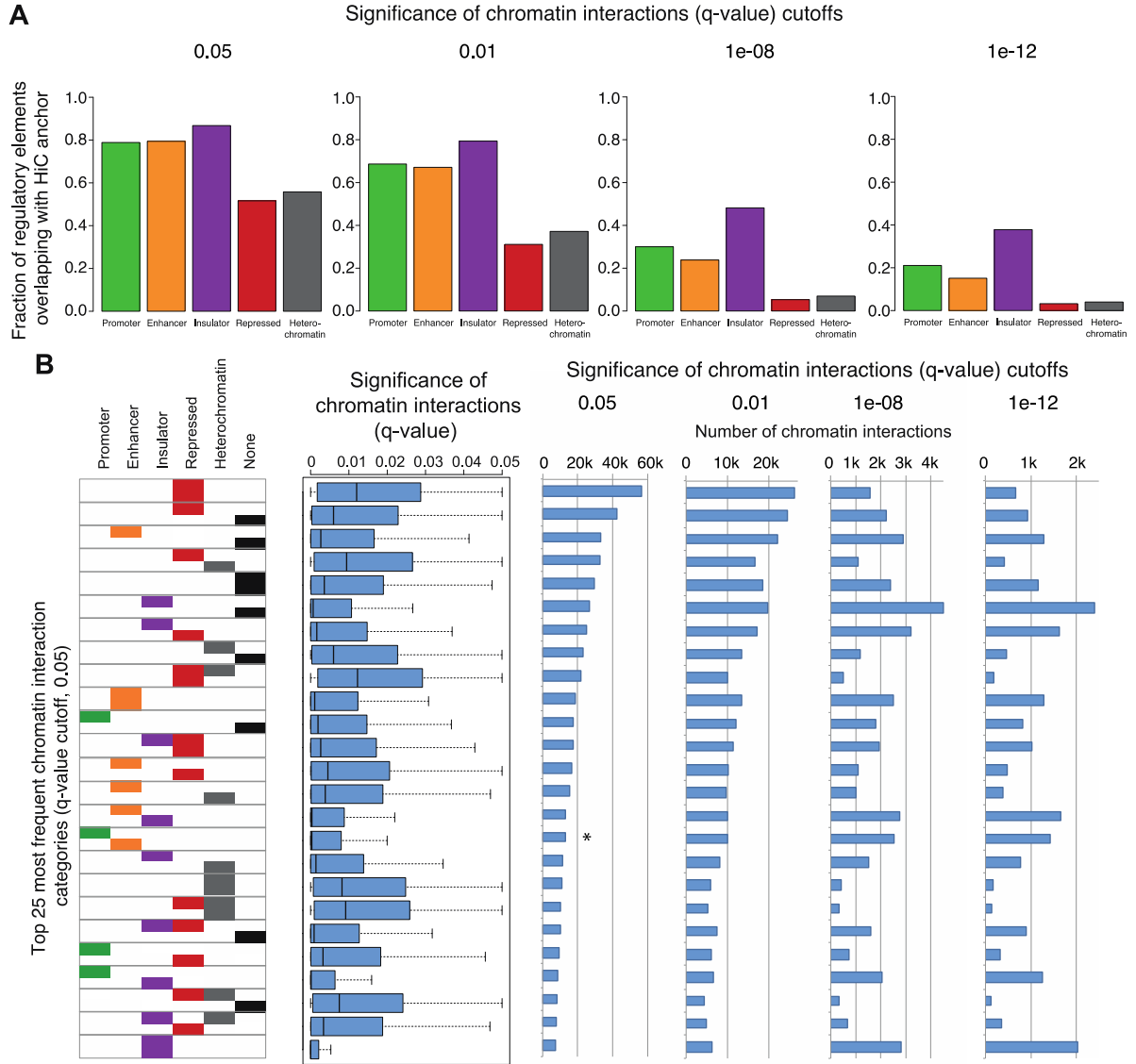
peaks +/- 2kb from a TSS) and 254,705 H3K27ac non-promoter H3K27ac peaks. From this, we predicted an additional 28,579 putative target genes linked to 216,840 enhancers from ~1.8M promoter-enhancer pairs. Using these promoter-enhancer pairs, we fitted a linear regression model between enhancer and gene expression data. We found that 5,115 unique gene expression levels showed a correlation with the peak strength of 10,635 unique H3K27ac peaks (multiple testing adjusted  $p$ -value  $< 0.05$ , 13,500 promoter-enhancer pairs) (table S11).

**Gene ontology analysis:** Target genes of enhancers linked to genotypic or phenotypic classifications as well as enhancers with SCZ risk-associated SNPs were predicted using the promoter-enhancer pairs identified above. Gene ontology analysis of potential target genes was performed using the GSEA tool, MSigDB (<http://www.broadinstitute.org/gsea>). Hypergeometric test was used to calculate  $p$ -value, and false discovery rate ( $q$ -value)  $< 0.05$  was used to select significantly enriched gene sets.



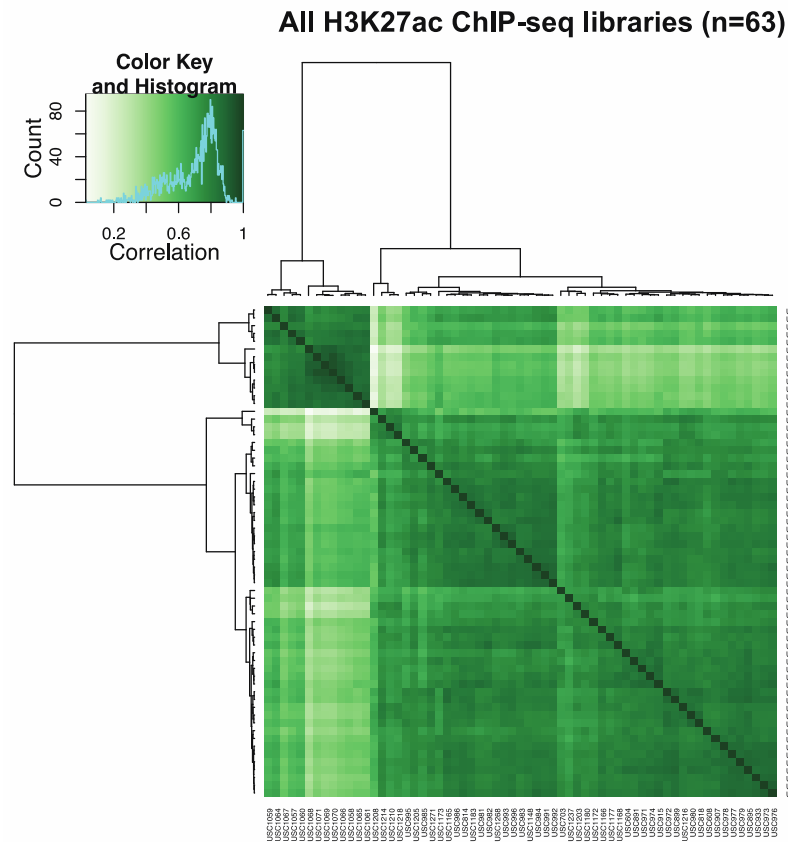
**Fig. S1. CNON Hi-C datasets.** (A) SEP044 and SEP045 Hi-C library QC analysis using the HICUP program (58). Left panel: the number of Hi-C reads found in each read category. Middle panel: the fraction of different types of read pairs identified in each dataset; the majority of the read pairs were valid pairs. Right panel: the number of DI-Tags that are cis short-range, cis long-

range and trans; most of the interactions are in cis (i.e. on the same chromosome). **(B)** SEP044 and SEP045 Hi-C library QC using the HiC-Pro program (51). Left panel: the number of Hi-C reads trimmed and aligned. Middle panel: the number of reads with pairs in each category. Right panel: the number of reads in valid interactions (cis short-range, cis long-range, trans) and the number of duplicates. **(C)** Pearson correlation between SEP044 and SEP045 at 100kb resolution. **(D)** A boxplot indicating the size of the TADs for the SEP044 (red) and SEP045 (blue) Hi-C datasets. **(E)** The number of TADs found in each chromosome for the SEP044 and SEP045 Hi-C datasets; SEP044 is from a female and therefore does not have ChrY TADs. **(F)** Overlap of TADs found in the SEP044 and SEP045 Hi-C datasets.

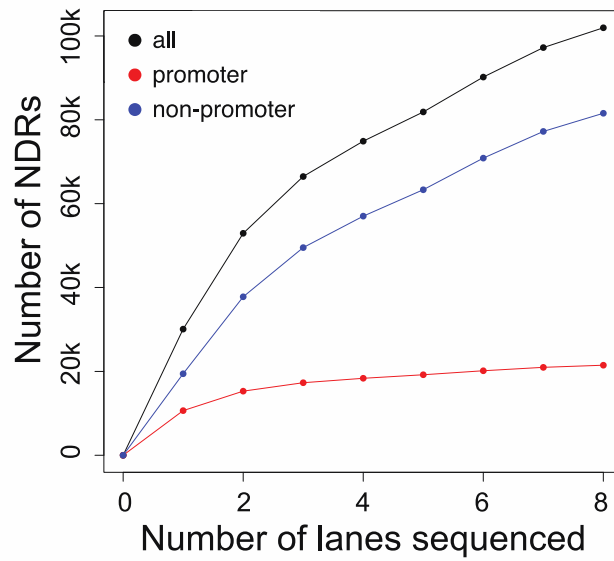


**Fig. S2. Chromatin interactions with different  $q$  value cutoffs.** (A) Fraction of regulatory elements overlapping with Hi-C anchor obtained using different  $q$ -value cutoffs for chromatin interactions ( $q$ -value cutoffs: 0.05, 0.01, 1e-08, 1e-12). (B) The 25 most frequent chromatin interaction categories found using a  $q$ -value cutoff of 0.05 (Left). The significance of the Hi-C interactions in each category (Middle). The number of interactions in each category, using  $q$ -value cutoffs of 0.05, 0.01, 1e-08, and 1-e12 (Right). The asterisk indicates the category representing interactions with a promoter on one end and an enhancer on the other end.





**Fig. S3. H3K27ac ChIP-seq clustering results.** After processing 63 H3K27ac ChIP-seq datasets, clustering analysis identified a subset of H3K27ac ChIP-seq samples that were very different from the rest of the datasets (shown by the small dark green box in the upper left of the clustering box). Further investigation revealed that these samples were all prepared from cells grown in the same 2-week period (USC1057 through USC1070 libraries). Because it was likely that these peak differences were due to a cell culture batch effect and because replicate datasets from some of these same individuals clustered with the majority of the samples, these outlier H3K27ac datasets were removed from all further analyses. The clustering analysis also identified poor quality datasets (e.g. datasets with low peak number or high background) and incorrectly labeled datasets (e.g. due to a sample swap or barcoding issue); these datasets were also removed, leaving a total of 47 H3K27ac datasets.



**Fig. S4. NOMe-seq depth and number of NDRs.** The SEP030 NOMe-seq library was sequenced in 8 lanes (each lane gave ~200M properly aligned read pairs). NDRs were identified using the processed data from each sequenced file. Shown is the number of cumulative unique NDRs from each additional sequence file.

**Table S1. Information about the CNON ChIP-seq, NOME-seq, and Hi-C datasets.**

**Table S2. CNON TADs identified by Hi-C.** (A) TADs found from SEP044 Hi-C dataset, (B) TADs found from SEP045 Hi-C dataset, (C) H3K9me3-marked TADs.

**Table S3. CNON ChIP-seq peaks classified as promoters, enhancers, insulators, repressed, or heterochromatin regions.**

**Table S4. CNON chromatin interactions identified by Hi-C.** (A) Chromatin interactions found from the SEP044 Hi-C dataset. (B) Chromatin interactions found from the SEP045 Hi-C dataset.

**Table S5. Epigenomic classification of CNON Hi-C chromatin interactions.** (A) All Hi-C chromatin interactions, (B) Hi-C anchors overlapping with promoter regions, (C) Hi-C anchors overlapping with enhancer regions, (D) Hi-C anchors overlapping with insulator regions, (E) Hi-C anchors overlapping with repressed regions, (F) Hi-C anchors overlapping with heterochromatin regions.

**Table S6. Predicted target genes of enhancers using CNON Hi-C.** Predicted target genes of enhancers identified by (A) direct promoter-enhancer chromatin interactions, (B) within CTCF-CTCF interactions loops, and (C) TADs.

**Table S7. Coordinates of promoter H3K4me3, non-promoter H3K27ac, and CTCF peaks found in many individuals of CNON datasets.** Coordinates of promoter H3K4me3 (A), non-promoter H3K27ac (B), and all CTCF (C) peaks found in many individuals of CNON datasets (n=56, 47, 33 for H3K4me3, H3K27ac, and CTCF, respectively).

**Table S8. Identification and classification of NDRs identified by CNON NOMe-seq.**

Identification and classification of (A) all NDRs, (B) promoter NDRs, (C) enhancer NDRs, (D) insulator NDRs, and (E) NDRs without features identified by CNON NOMe-seq.

**Table S9. Enhancers linked to genotype, gender, smoking status, and schizophrenia**

**diagnosis.** Enhancers linked to genotype (A), gender (B), smoking status (C), and schizophrenia diagnosis (D).

**Table S10. Classification and characterization of SCZ risk-associated variants.**

Classification and characterization of SCZ risk-associated variants (A) high LD SCZ variants that lie within CNON regulatory elements, (B) motifs and TFs related to the high LD SCZ variants that lie within CNON NDRs, (C) genes identified by a SCZ risk-related variant within a H3K4me3 promoter peak, (D) SCZ risk-associated regulatory variants correlated with ChIP-seq signals.

**Table S11. Predicted target genes of enhancers linked to individual variations.**

Predicted target genes of enhancers linked to genotype (A), gender (B), smoking status (C), schizophrenia diagnosis (D), schizophrenia-GWAS (E) and a list of genes that showed statistically significant correlation between RNA-seq and H3K27ac ChIP-seq data (F).