A functional role for non-coding variation in schizophrenia genome-wide significant loci

Roussos .. Sklar

Table of Contents

EXTENDED EXPERIMENTAL PROCEDURES

GWAS data sets

A large published SCZ GWAS data set (Ripke et al., 2013) (SCZ) in the form of summary statistic *P* values was obtained from public access website (https://pgc.unc.edu/). The SCZ meta-analysis included n=13,833 cases with SCZ and n=18,310 controls and identified nine LD independent regions that met genome-wide significance ($P < 5x10^{-8}$). In this dataset there were 3,538 out of 9,898,078 (9,815,700 after removing the major histocompatibility complex (MHC) locus) imputed SNPs that were genome-wide significant. A total of 22 LD independent regions were identified by replicating the top SNPs in an independent cohort. A previously published GWAS data set in rheumatoid arthritis (RA) (Stahl et al., 2010) was obtained through collaboration with investigators in the form of summary statistic *P* values. The RA GWAS data included a total of 5,485 seropositive individuals with RA and 22,609 controls of European descent. In the RA GWAS there were in total 13,798 out of 8,099,406 SNPs that reached genome-wide significance. For all the analyses presented here, the MHC locus (chr6: 25-35Mb) was excluded from both GWAS datasets.

eSNP data sets

Cohorts: Brain eSNPs were generated using the gene expression and genotyping data of Caucasian samples, included in the Braincloud (Colantuoni et al., 2011) (GEO accession number: GSE30272), NIA/NIH (Gibbs et al., 2010) (GEO accession number: GSE15745) and Harvard Brain Tissue Resource Center (HBTRC) (Zhang, 2013) (GEO accession number: GSE44772) datasets (Table S1). The following non-brain eSNP datasets were downloaded from public access websites: LCL (Xia et al., 2012) (http://www.bios.unc.edu/research/genomic_software/seeQTL/), liver (Innocenti et al., 2011) (http://www.scandb.org/), peripheral blood mononuclear cells (PBMC) (Westra et al., 2013) (http://genenetwork.nl/bloodeqtlbrowser), skin (Grundberg et al., 2012) and adipose tissue (Grundberg et al., 2012) (http://www.muther.ac.uk/Data.html).

DNA Genotyping QC and Imputation of Brain eSNP data sets: The preprocessing of Braincloud, NIA/NIH and HBTRC SNP data was done using plink (v1.07) (Purcell et al., 2007). More specifically, individuals were removed if they were outliers with respect to estimated heterozygosity or had missing SNPs > 10%. SNPs were removed if: missing genotype rate > 5%; Hardy–Weinberg equilibrium *P* value $<10^{-4}$; minor allele frequency <5%. Multidimensional scaling (MDS) was applied to detect population stratification and samples demonstrating non-Caucasian ancestry were eliminated. Genotyping data were imputed using the same reference panel that SCZ were imputed (1000 Genomes reference panel - phase I; March 2012). Prephasing and imputation of genotyping data was done with SHAPEIT (v2) (Delaneau et al., 2013) and IMPUTE2 (v2.3.0) (Howie et al., 2009), respectively. After imputation, we retained only SNPs that have been imputed with high certainty (Info score > 0.8) and SNPs were removed as described above

Microarray preprocessing: Microarray probes that did not map to human genome or mapped to more than a locus were removed using Blat (Kent, 2002) from UCSC

browser (http://genome.ucsc.edu). Probes containing SNPs with MAF > 0.01 according to 1000 Genomes European panel or in the current population were removed from the analysis. For preprocessing of the Braincloud and HBTRC data (two-color custom-spotted microarrays), we performed background correction on the linear scale, $\log^2$ transformation and loess normalization. For the NIA/NIH dataset (Illumina HumanRef-8 Expression BeadChips), we applied variance-stabilizing transformation and Robust Spline Normalization (Schmid et al., 2010). Probes that were not detected reliably in our gene expression data in >50% of samples were completely removed. Missing data for probes that were detected in >50% of samples were imputed to allow downstream processing. Known (age, sex, post-mortem interval (PMI), pH, RNA integrity number (RIN) and batch) and hidden confouders were removed using a Bayesian framework (Stegle et al., 2010) for joint modeling of diverse sources of phenotypic variability.

eSNP analyses: Human brain regional eSNP analysis was performed in eight gene expression and genotype datasets from 3 independent studies (Table S1). eSNP analysis was conducted based on linear regression models, adjusted for the first five MDS components. We define a significant *cis* interaction as any SNP that lies within 1Mb upstream or downstream from a gene. For multiple testing corrections we applied a false discovery rate (FDR) at 10%. We found that removal of confounds from gene expression datasets using a Bayesian framework, which accounts for both known and hidden confounds, compared to lineal models, resulted on average in ~2.2-fold greater number of discovered eSNPs (Table S2).

cis regulatory element (CRE) annotation description

Multiple CRE annotations were used in the current study (Table S2, S3).

ENCODE and REMC data: ChIP-seq and DHS data generated as part of the ENCODE (Maurano et al., 2012) and REMC (Zhu et al., 2013) projects for human brain/neuron (Table S2), T - helper cells, liver, skin and adipose tissue were downloaded from the NCBI repository (http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/). Additional data for the dorsolateral prefrontal cortex after fluorescence-activated cell sorting (FACS) were generated as described previously (Cheung et al., 2010; Shulha et al., 2013; Shulha et al., 2012) (Table S3). The non-brain ENCODE and REMC data that were used in the current study are the following:

1. T-helper: GSM621447; GSM665812; GSM665839; GSM701489; GSM701491; GSM701539; GSM772790; GSM772835; GSM772836; GSM772849; GSM772851; GSM772852; GSM772855; GSM772859; GSM772860; GSM772862; GSM772867; GSM772868; GSM772869; GSM772876; GSM772881; GSM772884; GSM772899; GSM772902; GSM772904; GSM772905; GSM772906; GSM772911; GSM772914; GSM772915; GSM772916; GSM772919; GSM772920; GSM772924; GSM772925; GSM772928; GSM772930; GSM772934; GSM772944; GSM772946; GSM772947;

GSM772948; GSM772953; GSM772954; GSM772955; GSM772963; GSM772973; GSM772985; GSM772986; GSM772987; GSM772988; GSM772997; GSM772998; GSM772999; GSM773004; GSM817166; GSM916024; GSM916026; GSM916027; GSM916059; GSM916071.

2. Liver: GSM1059451; GSM1059458; GSM1112808; GSM1112809; GSM1112814; GSM537697; GSM537698; GSM537705; GSM537706; GSM537707; GSM537709; GSM621629; GSM621630; GSM621654; GSM621675; GSM669910; GSM669972; GSM670008

3. Adipose: GSM621401; GSM621418; GSM621420; GSM621425; GSM621435; GSM621443; GSM621458; GSM669904; GSM669908; GSM669922; GSM669925; GSM669930; GSM669934; GSM669938; GSM669940; GSM669975; GSM669984; GSM669988; GSM669998; GSM670017; GSM670020; GSM670027; GSM670035; GSM670041; GSM670043; GSM670045; GSM772739; GSM772740; GSM772741; GSM772744; GSM772745; GSM772746; GSM772747; GSM772748; GSM772757; GSM772760; GSM772761; GSM772762; GSM772764; GSM772765; GSM772771; GSM772802; GSM772812; GSM772816; GSM772817; GSM772819; GSM772821; GSM772822; GSM906394; GSM906416; GSM916055; GSM916066

4. Skin: GSM1127060; GSM1127076; GSM669589; GSM669591; GSM669592; GSM774240; GSM774241; GSM817169; GSM817170; GSM817194; GSM817195; GSM817196; GSM817197; GSM817234; GSM817235; GSM817237; GSM817240; GSM817242; GSM817246; GSM817247; GSM878623; GSM878624; GSM878633; GSM878634; GSM878635; GSM878636; GSM878640; GSM878641; GSM878642; GSM878643; GSM878644; GSM878645; GSM878646; GSM878647; GSM878648; GSM878649; GSM941717; GSM941718; GSM941735; GSM941736; GSM941738; GSM941742; GSM958154; GSM958155; GSM958156; GSM958158; GSM958161; GSM958162; GSM958163; GSM958164; GSM958167; GSM958168

FACS H3K4me3 chromatin mark data: Previously published FACS H3K4me3 chromatin mark data (Cheung et al., 2010; Shulha et al., 2013; Shulha et al., 2012) with the addition of 3 samples were used in the current study (Table S3). Freshly frozen (never fixed) tissues from N=38 samples across lifespan, was provided by four independent brain banks. Specimens were obtained from the dorsolateral portion of the prefrontal cortex, primarily from cytoarchitectonic (Brodmann) Area 10 (BA10) and regions that border on BA10, including portions of BA9 and BA46. Tissue aliquots (200–500 mg/subject) were extracted in hypotonic lysis buffer, purified by ultracentrifugation and resuspended in 1x PBS, immunotagged with anti-neuronal nucleus (anti-NeuN, Millipore) antibody and sorted into NeuN(+) and NeuN(-) fractions. Mononucleosomal

preparations from at least 1×10^6 sorted nuclei were prepared for subsequent chromatin immunoprecipitation with anti-H3K4me3 antibody (Upstate/Millipore), followed by library preparation and sequencing on an Illumina Genome Analyzer II platform, as described previously (Cheung et al., 2010; Shulha et al., 2013; Shulha et al., 2012).

Processing of chromatin mark data: The majority of our FACS sequencing libraries has been included in previous publications (Cheung et al., 2010; Shulha et al., 2013; Shulha et al., 2012) (see also Table S3). All libraries contained single-end 36-bp reads and were mapped to the human genome with Bowtie2 (version 2.1.0) (Langmead and Salzberg, 2012). We allowed up to one mismatch and mapped all sequences to the gender appropriate genome hg19. Reads that mapped to multiple locations were discarded. As previously reported, H3K4me3 levels did not show correlations with postmortem interval and tissue pH (Cheung et al., 2010; Shulha et al., 2013; Shulha et al., 2012). We analyzed genome-matching reads with the MACS software (Zhang et al., 2008) (version 2.0.10) to identify significant peaks ($P < 1 \times 10^{-5}$) with bw = 230 bp, as defined experimentally by PCR, tSize = 36 bp, and other parameters set at default.

Identical analysis described above was applied for the ENCODE and REMC hg19-mapped sequence reads data. Because control data were available for each ChIP-seq peak data, paired ChIP-seq peak and control data were analyzed jointly using MACS to increase specificity of peak detection.

Definition of functional datasets

All SNP coordinates described here are relative to UCSC hg19. The functional datasets used in the current study were divided into 3 groups: eSNP, CRE and creSNP (eSNP in a *cis* regulatory element).

1. *eSNP:* The brain eSNP dataset was generated by including all significant *cis* eSNPs derived from each eSNP analyses. For the non-brain eSNPs (LCL, liver, skin, PBMC and adipose tissue), we used the list of *cis* eSNPs generated as described previously (Grundberg et al., 2012; Innocenti et al., 2011; Westra et al., 2013; Xia et al., 2012) at FDR 10%.
2. *CRE:* The ChIP-seq and DHS significant peaks were clustered into subgroups based on assay and origin of tissue (Figure S4). For definition of high confidence CRE intervals we divided the genome into 50 bp bins and included only intervals with > 50% overlap among the peaks of each track within the samples in that subgroup. The > 50% threshold is an arbitrary cutoff that was selected based on the following criteria: 1.) the count of identified CRE intervals, 2.) the total and

average CRE interval, 3.) the proportion of the human genome covered, and 4.) the overlap with curated data generated by the ENCODE and REMC. Overall, the > 50% threshold retains high confidence CREs without reducing significantly the size and number of included intervals. Finally, we integrated multiple CRE subgroup annotations for generating functional annotations defining active promoters (overlap of H3K4me3 with H3K9ac or H3K27ac), poised promoters (overlap of H3K4me3 with H3K27me3), active enhancers (overlap of H3K4me3 with H3K9ac or H3K27ac), repressed enhancers (overlap of H3K4me1 with H3K27me3) and open chromatin (DHS).

3. *creSNP:* The creSNP functional dataset was defined as the eSNPs that lie within different CRE subcategories.

GWAS Positional Annotation

LD calculation: Variant call format (VCF) genotype files for the European reference sample provided by the November 2010 release of Phase 1 of the 1000 Genomes Project (1000G) were obtained from the NCBI repository. Additional quality control based on standard GWAS procedures was performed on 1000G data using Plink version 1.07 (Purcell et al., 2007). More specifically, individuals were removed if they had missing SNPs > 10%. SNPs were removed if: missing genotype rate > 5%; Hardy–Weinberg equilibrium *P* value $<10^{-6}$; minor allele frequency <1%. The identity by state and identity by descent analysis implemented at Plink were used for estimation of relatedness. The individual with the higher missingness from each related pair was removed. For each 1000G SNP, the r^2 pairwise LD was calculated within 1,000,000 base pairs (1Mb) on either side of the SNP threshold of $r^2 \geq 0.8$.

We used a mixed approach for assigning the GWAS SNPs into functional categories. For the eSNP functional category, we leveraged the eSNP dataset in the densely mapped 1000G data to identify the GWAS studied SNP that was tagged, as a result of linkage disequilibrium (LD). For generating the CRE or creSNP functional categories we used a positional approach (ignoring the annotation categories of SNPs in LD with the tag SNP). This mixed model allows us to capture all possible SNPs that affect gene expression (tag or SNPs in LD), followed by positional selection of SNPs that lie within putative regulatory DNA regions. GWAS SNPs that did not fit into any of the above functional categories were assigned as the FUV category. The Table S7 describes the count of SCZ for each functional brain category.

Quantification of Enrichment

Stratified Q-Q plots were generated for assessment of the similarity or differences between the empirical cumulative distribution function of the functional and FUV datasets. Each empirical null distribution was corrected by applying a control method leveraging only the FUV SNPs that are less enriched for true associations. A similar approach was used by Schork *et al.* and allows correcting for global variance inflation

due to effects of cryptic relatedness and population substructure minimizing the deflation due to over-correction of test statistics for polygenic traits by standard genomic control methods (Schork et al., 2013). The inflation factor, λ_{GC} was computed as the median z-score squared divided by the expected median of a chi-square distribution with one degree of freedom. All test statistics described here were divided by λ_{GC}. In the stratified Q-Q plots, the enrichment of SCZ SNPs for a specific functional category is observed as a horizontal deflection from the FUV category.

The quantification of enrichment for each functional category was done using the categorical enrichment score (CES), as described previously (Schork et al., 2013). The CES provides a summary score of category-specific enrichment where the mean is taken over all SNP z-scores in the given category. It is estimated based on as the mean (z^2 - 1) and it is justified as a measure of enrichment based on a simple Bayesian mixture model framework. The CES is a conservative estimate of the variance attributable to non-null SNPs, given a standard normal null distribution and a non-null distribution symmetric around zero.

The statistical significance of CES was evaluated by permutation on a combined 10,000 randomized set of CREs and eSNPs. For CREs, we generated 100 random sets matched for distance from TSS, size and number of intervals. We generated 100 random sets of eSNPs matched for MAF, distance from TSS and local LD (based on sum r^2) with the most significant eSNP from each LD interval (using the clump function implemented in PLINK with r2 < 0.1 within 100-kb window). For each functional category, we estimated the null distribution of CES by combining the 100 random sets of CREs and eSNPs (10,000 total permutations).

Control for potential confounds: Linkage disequilibrium and minor allele frequency

We examined whether differences in the LD (estimated based on the sum r^2) and MAF among functional categories (Table S6), could lead to spurious enrichment. For each SNP, the SCZ GWAS summary value (log of z^2 after FUV inflation control) was included as the dependent variable and the genetic variance (estimated as MAF*(1-MAF)), total LD (sum r^2) and membership for functional annotation categories were included as the independent variables in multiple regression analysis (Table S7). The heteroscedasticity and correlated responses from cluster samples was corrected using the Huber-White method.

Differences in the LD (estimated based on the sum r^2) were observed among the functional categories (Table S6). We examined for enrichment of each functional category compared to the FUV SNPs after retaining for each annotation category the most significant, LD-independent SCZ SNP using the clumping function implemented at Plink (with parameters of $r^2 > 0.1$ and distance 100 kb). The significance values for the curves for each functional annotation were calculated using a two-sample Kolmogorov-Smirnov Test (Table S6).

Recombination hotspot interval calculation

We divided the genome into recombination hotspot intervals based on inferred recombination rates and genetic map positions for 1000 Genomes data (March 2012 all-populations map data distributed with IMPUTE2 software(Howie et al., 2009)). We used a sliding window analysis to identify hotspots defined as regions with greater than 20 cM/Mb average rate and cumulative distance ('width' in cM) greater than 0.01 cM, then conservatively filtered out hotspots less than 100 Kb or 0.1 cM from a longer genetic-width hotspot, and defined intervals as regions between consecutive hotspot midpoints. This resulted in 4,258 hotspots with mean (standard deviation) rate: 33 cM/Mb (11.6) and widths: 4.1 Kb (6.4) and 0.1 cM (0.098), and 4,280 hotspot intervals.

Regulatory Trait Concordance (RTC) analysis

The likelihood of a shared functional effect between a SCZ genome-wide significant SNP and an eSNP was assessed by the RTC approach (Grundberg et al., 2012; Nica et al., 2010). Each SCZ index SNP in Ripke et al. was mapped to recombination hotspot intervals. For each recombination interval we identified all genes with a significant eSNP. We then repeat the eSNP analysis based on linear regression models after conditioning on the SCZ index SNP and estimate the change in the eSNP statistical significance. If the GWAS index SNP and the eSNP tag the same functional variant, then by removing the genetic effect of the GWAS SNP, the eSNP statistical association will be reduced or completely lost. To take in account the LD structure within each recombination hotspot interval, we ranked the impact of SCZ index SNP on eSNP significance ($Rank_{GWAS}$) by correcting for all other SNPs within the same interval (N_{SNPs}). The rank estimates the number of SNPs which when used to correct the expression data, have a higher impact on the eSNP statistical significance (less significant adjusted *P*) than the GWAS SNP. The RTC is estimated based on the following formula:

$$RTC = \frac{N_{SNPs} - Rank_{GWAS}}{N_{SNPs}}$$

If the GWAS SNP is the same with the eSNP then the eSNP adjusted *P* will be 1, the $Rank_{GWAS}$ = 0 and RTC = 1. RTC score ranges from 0 to 1, with values ≥ 0.9 indicating likely causal regulatory effects, as demonstrated previously (Grundberg et al., 2012; Nica et al., 2010).

Chromosome Conformation Capture (3C) analysis

Postmortem prefrontal cortex brain tissue for cases with SCZ and controls was obtained from the University of Maryland and pair-matched for age, sex, PMI, and pH (*N*=3/group) (Table S8). PFC tissue (200mg) was dissected, homogenized, and

crosslinked for 10 minutes at 25°C in 1% formaldehyde, 1X protease inhibitor (Sigma), and 2 mL of lysis buffer (10 mmol/L Tris hydrogen chloride pH 8.0 / 10 mmol/L sodium chloride / 0.2% IPEGAL CA-630 (Sigma Aldrich, St. Louis, Missouri)). Crosslinking was stopped by the addition of glycine to a final concentration of 0.125 mol/L for 10 minutes at 4°C and the homogenate was incubated at 4°C for an additional 25 minutes. Cell were lysed by pipetting >50 times and spun at 5000 rpm, the supernatant was removed, and the pellet was washed twice with 1X New England Buffer 4 (NEB4) (New England Biolabs, Boston, Massachusetts). Samples were resuspended in 200 µl of 1X NEB4 and divided into four 50 µl aliquots. An additional 312 µl of 1X NEB4 was added to each aliquot. Samples were incubated at 65°C for 10 minutes in 1X NEB4 buffer and 38 µl of 1% SDS for more efficient digestion and to remove protein not associated with DNA. To quench SDS 10% of Triton X-100 was added to each sample and the samples were digested with *HindIII*-HF or *NcoI*-HF (NEB) at 37°C overnight with gentle shaking.

HindIII-HF or *NcoI*-HF was inactivated by the addition 86 µl of 10% SDS incubated for 30 minutes at 65°C. Ligation mixture (7.61 ml) was added to each sample. The ligation mixture consisted of 745 µl of 10% Triton X-100, 745 µl of 10X ligation buffer (1 M Tris HCL, pH 7.5, 1 M $MgCl_2$, 1 M DTT dithiothreitol (Bio-Rad)), 80 µl of 10 mg/ml bovine serum albumin (NEB), 80 µl of 100 mM ATP (Sigma) and 5960 µl of autoclaved water. 50 µl of T4 DNA ligase (1 U/µl, Invitrogen) was added to three aliquots and one sample was used as a no ligase control. Ligation proceeded for five hours at 16°C and samples were reverse cross-linked at 65°C overnight with 50 uL of 10 mg/ml of proteinase K (Sigma). For improved ligated DNA recovery another 50 µl of proteinase K was added and incubated at 65°C for two hours. DNA was extracted with phenol (pH 8.0, Fisher), and phenol-chloroform (1:1) (pH 8, Fisher). DNA was precipitated using 1/10 the volume of 3M sodium acetate (pH. 5.4) and 2.5 the volume of ice-cold ethanol overnight. The samples were centrifuged at 8000 RPM for 30 minutes and washed with 70% ethanol. The final DNA pellet was dissolved in 1X TE buffer (pH 8.0). Phenol and phenol-chloroform extraction and ethanol precipitation was repeated. The final 3C library was washed five times with 70% ethanol. Ligase and no ligase reactions were dissolved in 100 µl and 33 µl of TE buffer (pH 8.0) respectively (Mitchell et al., 2013). 3C libraries with and without the critical ligase step were run on a 2% agarose gel to visualize ligation efficiency. Samples ran at a higher molecular weight after ligation, indicated by an upward shift on the gels (Figure S5A) (Dekker, Rippe et al. 2002, Dekker 2006).

Physical looping interactions were quantified with PCR. Primers were designed less than 120 bp from the *HindIII* or *NcoI* restriction site (Table S9). The PCR products were resolved on a 2% agarose gel and the level of interaction between two primers was measured semiquantitatively using band intensities normalized with the background (raw 3C interaction) with ImageJ (Schneider et al., 2012). Library input was adjusted for each library according the interaction between two neighboring primers (<5000 bp apart) and two distant primers (<30,000 apart) control primer 1 (CCTGGATCATCAGACAGAACTAAAGCTCTT) located at chr13:99113854 and control primer 2 (CTTCAACTGAAAACACACGAACAGGAAGAA) located at chr13:99109553 (Figure S5B). A bacterial artificial control (BAC) 3C library was created for the *CACNA1C* region to normalize for primer efficiency starting with equimolar concentrations of RP11-465I2 and RP11-698B24 (Figure S5C) (Dekker, Rippe et al.

2002, Dekker 2006). For each library (*HindIII* and *NcoI*) in the human postmortem brain tissue studies, we transformed the raw 3C interaction to *Z* scores, followed by scaling (0 to 1) (Scaled 3C interaction). All 3C PCR products were sequence verified and the interactions were not present in the no ligase and water controls.

For primers #2, #4 and #5, physical looping interactions were further quantified with qPCR using an ABI Prism 7900 (Applied Biosystems). The thermal cycling program consists of 5 min at 95°C, followed by 40 cycles of 30 sec at 95°C, 30 sec at 60°C and 30 sec at 72°C. Only one DNA was amplified in each PCR (monoplex). The reactions were run in triplicate for each sample and DNA PCR product was measured through SYBR Green I (Life Technologies). For more accurate quantification of amplified DNA, we used the Relative Standard Curve Method[123]. To account for differences in the amount of input material between the samples, the SYBR Green I signal from each 3C interaction region was normalized to the geometric mean of the SYBR Green I signal of two endogenous references [large ribosomal protein (*RPLP0*) and Glucuronidase, Beta (*GUSB*)].

Human induced pluripotent stem cells differentiation into neurons

Human induced pluripotent stem cells (hiPSCs) were derived from fibroblasts of a control sample (GM03651) as described previously (Brennand et al., 2011). For neural differentiations, neural progenitor cells (NPCs) were dissociated with Accutase and plated in neural differentiation media (DMEM/F12, 1x N2, 1X B27-RA, 20 ng/ml BDNF (Peprotech), 20 ng/ml GDNF (Peprotech), 1 mM dibutyrl-cyclicAMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/Laminin-coated plates. Density is critical: for 6-well plates, 200,000 NPCs were plated per well. hiPSC derived-neurons were differentiated for ~6 weeks in neural differentiation media. The majority of forebrain hiPSC neurons are VGLUT1-positive, and so are presumably glutamatergic, although approximately 30% of neurons are GAD67-positive (GABAergic) (Brennand et al., 2011). Gene expression comparisons of 6-week-old neurons to the Allen BrainSpan Atlas indicate that our hiPSC neural cells resemble fetal brain tissue {Brennand, 2014 #2341}.

Transient Transfection and Luciferase Assays

We constructed luciferase reporter plasmids by cloning the regulatory sequence containing rs2159100 into the pGL4.24 vector (Promega) upstream of the minP. The regulatory sequence was introduced at the 5' and 3' by using *KpnI* and *XhoI*. We sequenced the inserted portions of the constructs to verify the nucleic acid sequences and the location of the SNP. Human embryonic kidney 293 (HEK-293) cells or Neuro-2a cells (40-60% confluent) were transfected with each reporter vector (450ng) or construct (500 ng to ensure the same copy number as the empty vector) and the Renilla luciferase expression vector pRL-TK (200 ng, Promega) using Lipofectamine 2000 (3:1, µL Lipofectamine: µgDNA) (Invitrogen) in 200 µL Opti-MEM (Invitrogen) in 12 well

plates. HEK 293 cells or Neuro-2a cells were grown in 1 ml DMEM supplemented with 10% FBS, and the media was not changed after the addition of the transfection reagents. Twenty-four hours after transfection, cells were lysed by the addition of 250 µL of Passive Lysis Buffer (Promega). The luciferase activity in the cell lysates was determined using the Dual Luciferase Reporter System (Promega) in quadruplicates. All experiments for HEK-293 and Neuro-2a were performed in duplicates. Firefly luciferase activities were normalized to that of Renilla luciferase and expression relative to the activity of the rs2159100 C allele was noted.

The regulatory sequence containing the rs2159100 C and T alleles ± 250 bp upstream/downstream (hg19; chr12:2,346,143-2,346,643) is provided below. Six nucleotides (in bold/underlined fonts) were added at the 5' and 3' ends as restriction sites for the *KpnI* and *XhoI,* respectively.

GGTACCACAATGCCTTGTGATACTCTTGTTCTTCTGGTTTGAGTTTTGGTAGATAAG
CACATCTGAGTCTTGCTGTGTTAATGTGTCTGTATTTTGGTGTATCTGCTTGCTTGT
CGTGTGGGGCATATGCCAAGTCCAGTAGTGGATGGGCTGGGGAAGACCAGACCT
TATCACATGGTGCCCTTGGGGGGAAATCTTAATTCCAATGTGTGAAACCAGTGAAA
GTATGATTTTCTGGGTCAATTTTAAAAATATA**[C/T]**GTTCAAGCAAAAAGCAACCTGT
TATCTCTTCTCTTTCTGCCTCTGCACACAGCAGCCTCCATTGCCTAGGGTATGATA
GTGTGGGTTCACTTTGTCCATCTCATTTGGATGACATCAGCGAAGATGCATTCTGT
ATCTCTCCACTGAGGCCTGTGACAGGACCTAATGTTTTGTGGAGCTGAGAGAAAAT
AAACCAAAATGACCCAATGGAATATAAATGCTGATTTCTGTTCCTGTTGTTTGACAG
AAA**CTCGAG**

SUPPLEMENTAL REFERENCES

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D.*, et al.* (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221-225.

Cheung, I., Shulha, H.P., Jiang, Y., Matevossian, A., Wang, J., Weng, Z., and Akbarian, S. (2010). Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America *107*, 8824-8829.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R.*, et al.* (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature *478*, 519-523.

Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nature methods *10*, 5-6.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J.*, et al.* (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet *6*, e1000952.

Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A.*, et al.* (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet *44*, 1084-1089.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet *5*, e1000529.

Innocenti, F., Cooper, G.M., Stanaway, I.B., Gamazon, E.R., Smith, J.D., Mirkov, S., Ramirez, J., Liu, W., Lin, Y.S., Moloney, C.*, et al.* (2011). Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet *7*, e1002078.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-1195.

Mitchell, A.C., Bharadwaj, R., Whittle, C., Krueger, W., Mirnics, K., Hurd, Y., Rasmussen, T., and Akbarian, S. (2013). The Genome in Three Dimensions: A New Frontier in Human Brain Research. Biological psychiatry.

Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet *6*, e1000895.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics *81*, 559-575.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M.*, et al.* (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet.
Roussos, P., Katsel, P., Davis, K.L., Bitsios, P., Giakoumaki, S.G., Jogia, J., Rozsnyai, K., Collier, D., Frangou, S., Siever, L.J.*, et al.* (2012a). Molecular and genetic evidence for abnormalities in the nodes of Ranvier in schizophrenia. Archives of general psychiatry *69*, 7-15.
Roussos, P., Katsel, P., Davis, K.L., Siever, L.J., and Haroutunian, V. (2012b). A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. Archives of general psychiatry *69*, 1205-1213.
Schmid, R., Baum, P., Ittrich, C., Fundel-Clemens, K., Huber, W., Brors, B., Eils, R., Weith, A., Mennerich, D., and Quast, K. (2010). Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. BMC Genomics *11*, 349.
Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods *9*, 671-675.
Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan, P.F., Kelsoe, J.R., O'Donovan, M.C., Furberg, H., Schork, N.J.*, et al.* (2013). All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLoS Genet *9*, e1003449.
Shulha, H.P., Cheung, I., Guo, Y., Akbarian, S., and Weng, Z. (2013). Coordinated cell type-specific epigenetic remodeling in prefrontal cortex begins before birth and continues into early adulthood. PLoS Genet *9*, e1003433.
Shulha, H.P., Cheung, I., Whittle, C., Wang, J., Virgil, D., Lin, C.L., Guo, Y., Lessard, A., Akbarian, S., and Weng, Z. (2012). Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. Archives of general psychiatry *69*, 314-324.
Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li, Y., Kurreeman, F.A., Zhernakova, A., Hinks, A.*, et al.* (2010). Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet *42*, 508-514.
Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS computational biology *6*, e1000770.
Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research *38*, e164.
Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E.*, et al.* (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet *45*, 1238-1243.
Xia, K., Shabalin, A.A., Huang, S., Madar, V., Zhou, Y.H., Wang, W., Zou, F., Sun, W., Sullivan, P.F., and Wright, F.A. (2012). seeQTL: a searchable database for human eQTLs. Bioinformatics *28*, 451-452.
Zhang, B.G., G.; Bodea, L.G.; Wang, Z.; McElwee, J.; Podtelezhnikov, A.A.; Zhang, C.; Xie, T.; Tran, L.; Dobrin, R.; Fluder, E.; Clurman, B.; Melquist, S.; Narayanan, M.; Bennett, D.A.; Suver, C.; Shah, H.; Mahajan, M.; Lamb, J.R.; Molony, C.; Stone, D.J.;

Gudnason, V.; Myers, A.J.; Schadt, E.E.; Neumann, H.; Zhu, J.; Emilsson, V. (2013). Tracing Multi-System Failure in Alzheimer Disease to Causal Genes. Cell *In Press*.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L.*, et al.* (2013). Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues. Cell *152*, 642-654.

Supplementary Tables

Table S1, related to "Methods: eSNP data sets". Characteristics of the brain gene expression and genotype cohorts

Dataset	Subjects	Tissue-RNA	RNA assay	Probes	Tissue-DNA	DNA assay	SNPs	Imputed SNPs	GEO accession number
BrainCloud (BC)	108	DLPFC	Illumina custom	26,040	CRBLM	Infinium II 650K or HD Gemini 1M Duo	649,903	5,754,042	GSE30272
NIH	145	DLPFC	Human Ref-8	11,624	CRBLM	HumanHap550	553,424	5,178,203	GSE15745
	144	STG							
	144	CRBLM							
	143	Pons							
Harvard Brain Tissue Resource Center (HBTRC)	146	DLPFC	Agilent custom	28,592	CRBLM	HumanHap650Y	548,978	5,371,864	GSE44772
	110	PVC							
	117	CRBLM							

Table S2, related to "Methods: *cis* regulatory element annotations". Brain related epigenomic data generated by the ENCODE and REMC projects that were used in the current study

Type	Assay	N	Region	ChIP-seq	Control
Adult brain-homogenate	H3K27ac	14	Anterior caudate	GSM1112811	GSM669978
				GSM772832	GSM772826
			Angular gyrus	GSM773016	GSM772931
			Cingulate gyrus	GSM1112813	GSM670003
				GSM773011	GSM772990
			DLPFC	GSM1112810	GSM669960
				GSM773015	GSM773010
			Hippocampus	GSM1112791	GSM669971
				GSM773020	GSM773019
				GSM916035	GSM916037
			Substantia nigra	GSM1112778	GSM772864
				GSM997258	GSM772864
			Temporal gyrus	GSM1112812	GSM670010
				GSM772995	GSM772991
	H3K4me1	14	Anterior caudate	GSM669970	GSM669978
				GSM772830	GSM772826
			Angular gyrus	GSM772962	GSM772931
			Cingulate gyrus	GSM670033	GSM670003
				GSM773007	GSM772990
			DLPFC	GSM670015	GSM669960
				GSM773014	GSM773010
			Hippocampus	GSM669962	GSM669971
				GSM773021	GSM773019
				GSM916039	GSM916037
			Substantia nigra	GSM669941	GSM670042
				GSM772898	GSM772864
			Temporal gyrus	GSM670036	GSM670010
				GSM772992	GSM772991
	H3K4me3	14	Anterior caudate	GSM670031	GSM669978
				GSM772829	GSM772826
			Angular gyrus	GSM772959	GSM772931
			Cingulate gyrus	GSM669905	GSM670003
				GSM773008	GSM772990
			DLPFC	GSM670016	GSM669960
				GSM773012	GSM773010
			Hippocampus	GSM670022	GSM669971
				GSM773022	GSM773019
				GSM916040	GSM916037
			Substantia nigra	GSM670038	GSM670042
				GSM772901	GSM772864
			Temporal gyrus	GSM669992	GSM670010
				GSM772996	GSM772991
	H3K9ac	6	Anterior caudate	GSM669950	GSM669978
			Cingulate gyrus	GSM670032	GSM670003
			DLPFC	GSM670021	GSM669960
			Hippocampus	GSM669906	GSM669971
			Substantia nigra	GSM669977	GSM670042
			Temporal gyrus	GSM669918	GSM670010
	H3K27me3	12	Anterior caudate	GSM669915	GSM669978
				GSM772827	GSM772826
			Angular gyrus	GSM772983	GSM772931
			Cingulate gyrus	GSM772989	GSM772990
			DLPFC	GSM772833	GSM773010
			Hippocampus	GSM1112800	GSM773019
				GSM669913	GSM669971
				GSM916038	GSM916037
			Substantia nigra	GSM669953	GSM670042

				GSM772937	GSM772864
			Temporal gyrus	GSM772772	GSM670010
				GSM772993	GSM772991
Fetal brain-homogenate	DHS	13	Fetal brain	GSM1027328	NA
				GSM530651	NA
				GSM595913	NA
				GSM595920	NA
				GSM595922	NA
				GSM595923	NA
				GSM595926	NA
				GSM595928	NA
				GSM665804	NA
				GSM665819	NA
				GSM878650	NA
				GSM878651	NA
				GSM878652	NA
	H3K4me1	4	Fetal brain	GSM706850	GSM706851
				GSM806934	GSM817243
				GSM806942	GSM806948
			Germinal matrix	GSM806939	GSM806947
	H3K4me3	3	Fetal brain	GSM806935	GSM817243
				GSM806943	GSM806948
			Germinal matrix	GSM806940	GSM806947
	H3K27me3	3	fBrain	GSM806937	GSM817243
				GSM806945	GSM806948
				GSM916061	GSM706851
Primary Cell Culture/iPS	H3K4me1	7	Neurosphere ganglionic eminence	GSM1127066	GSM1127091
				GSM707008	GSM707014
				GSM817232	GSM817245
			Neurosphere cortical cells	GSM707003	GSM707013
				GSM817230	GSM817244
			H1 derived neuronal progenitor cells	GSM772808	GSM772805
			H9 derived neuronal cells	GSM772785	GSM772737
	H3K4me3	5	Neurosphere ganglionic eminence	GSM1127062	GSM1127091
				GSM707009	GSM707014
			Neurosphere cortical cells	GSM707004	GSM707013
			H9 derived neuronal progenitor cells	GSM772736	GSM772805
			H9 derived neuronal cells	GSM772776	GSM772737
	H3K27me3	7	Neurosphere ganglionic eminence	GSM1127078	GSM1127091
				GSM707011	GSM707014
				GSM941715	GSM817245
			Neurosphere cortical cells	GSM707006	GSM707013
				GSM941713	GSM817244
			H1 derived neuronal progenitor cells	GSM772801	GSM772805
			H9 derived neuronal cells	GSM772787	GSM772737

Table S3, related to "Methods: *cis* regulatory element annotations". Brain fluorescence-activated cell sorting samples used in this study and sequencing statistics

Age group	Sample ID	Brain bank	Gender	Age	PMI	pH	Cause of death	Cell type	Total reads (M)	Unique map (%)	Previously published
NeuN(+): Fetal/Infancy (Prenatal - 1 yr)	S1	UM-BTB	Female	34 gw	9	6.6	Multiple congenital anomalies	Neuron	12.3	82	PLoS
	S2	UM-BTB	Female	39 gw	2	N/A	Unknown	Neuron	14.5	84	PLoS
	S3	UM-BTB	Female	40 gw	11	6.2	Respiratory insufficiency	Neuron	13.9	87	PLoS
	S4	UM-BTB	Male	0.5 yr	N/A	6	Complications of prematurity	Neuron	6.7	85	PNAS
	S5	UM-BTB	Male	0.6 yr	8	6.6	Positional asphyxia	Neuron	2.1	73	PNAS
	S6	UM-BTB	Male	0.6 yr	20	6.6	Unknown	Neuron	25.3	83	-
	S7	UM-BTB	Female	0.8 yr	8	6	Respiratory insufficiency	Neuron	4.7	80	PNAS
NeuN(+): Childhood and Adolescence (1 - 20 yr)	S8	UM-BTB	Female	2.8 yr	12	6.5	Drowning	Neuron	6.3	78	PNAS
	S9	HBTRC	Female	4 yr	N/A	6.2	Acute pneumonia	Neuron	11.9	74	PLoS
	S10	UM-BTB	Male	4.7 yr	17	6.5	Drowning	Neuron	3.8	79	PNAS
	S11	UM-BTB	Male	8.8 yr	5	6.7	Cardiac arrhythmia	Neuron	2.2	81	PNAS
	S12	UM-BTB	Male	14 yr	13	7	Accident	Neuron	3.2	74	PNAS
	S13	HBTRC	Male	17 yr	30.8	6.6	Hanging	Neuron	20.2	84	PLoS
	S14	UCI/UCD	Male	18 yr	22	7	Drowning	Neuron	12.8	84	PLoS
	S15	HBTRC	Male	19 yr	18.6	6	Pneumonia	Neuron	13.2	89	PLoS
NeuN(+): Early/Middle adulthood (20 - 40 yr)	S16	MPRC	Male	20 yr	15	6.4	Gunshot to chest; homicide	Neuron	10.7	86	AGP
	S17	MPRC	Male	23 yr	12	6.6	Unknown	Neuron	13.4	83	PLoS
	S18	HBTRC	Male	24 yr	21.3	6.2	Myocardial infarction	Neuron	18.4	80	PLoS
	S19	MPRC	Male	24 yr	22	6.5	Gunshot to chest; suicide	Neuron	4.1	82	AGP
	S20	HBTRC	Male	26 yr	32.8	6.2	Heart attack	Neuron	18.1	87	PLoS
	S21	MPRC	Female	28 yr	24	6.7	Congenital heart disease	Neuron	23.4	85	AGP
	S22	UCI/UCD	Male	38 yr	16	6.1	Cardiac	Neuron	7.6	66	AGP
	S23	HBTRC	Female	41 yr	14	6.6	Unknown	Neuron	12.7	81	PLoS
	S24	MPRC	Male	55 yr	17	6.9	Cardiac	Neuron	6	82	AGP
NeuN(+): Late adulthood (>60 yr)	S25	UCI/UCD	Male	63 yr	24.5	6.9	Cardiac	Neuron	8.4	90	PLoS
	S26	HBTRC	Male	64 yr	28.2	6.2	Unknown	Neuron	17.5	81	PLoS
	S27	UCI/UCD	Female	68 yr	7	6.1	Respiratory failure	Neuron	8.3	72	PNAS
	S28	UCI/UCD	Female	69 yr	7.3	6	Multiple injuries	Neuron	3.5	62	-
	S29	UCI/UCD	Female	69 yr	40	6.3	Cardiac	Neuron	4.6	83	PNAS
	S30	MPRC	Female	74 yr	18.5	7.2	Cardiac	Neuron	7.5	90	PLoS
	S31	MPRC	Male	74 yr	12	6.7	Cardiac	Neuron	3.1	83	PLoS
	S32	UCI/UCD	Male	81 yr	8	6.6	Cardiac	Neuron	10.3	77	PLoS
NeuN(-): Glia	S3	UM-BTB	Female	40 gw	11	6.2	Respiratory insufficiency	Glia	13	81	PLoS
	S6	UM-BTB	Male	0.6 yr	20	6.6	Unknown	Glia	22	72	PLoS
	S10	UM-BTB	Male	4.7 yr	17	6.5	Drowning	Glia	7.1	59	-
	S15	HBTRC	Male	19 yr	18.6	6	Pneumonia	Glia	15	76	AGP
	S18	HBTRC	Male	24 yr	21.3	6.2	Myocardial infarction	Glia	17.8	77	AGP

	S28	UCI/UCD	Female	69 yr	7.3	6	Multiple injuries	Glia	6.2	84	PNAS

Brain Banks: HBTRC – Harvard Brain Tissue Resource Center (Dr. Francine Benes); UM-BTB – University of Maryland Brain and Tissue Bank for Developmental Disorders (Dr. Ron Zielke); MPRC – Maryland Psychiatric Research Center (Dr. Rosalinda Roberts and Dr. Andree Lessard); UCI/UCD – University of Irvine/Davis (Dr. Ted Jones and Dr. William E. Bunney Jr.).
PLoS = Shulha et al., PLoS Genet 9(4):e1003433 (2013); AGP = Shulha et al., Archives of General Psychiatry 69:314-324 (2012); PNAS = Cheung et al., Proceedings National Acad Sci USA 107:8824-8829 (2010). yr = years; gw = gestational week.

Table S4, related to Figure 1 and Table 1. Counts of SNPs per annotation category

Type	Functional category	Mixed
eSNP	eSNP	1,479,508
CRE	Adult brain - H3K27ac	469,551
	Adult brain - H3K4me1	515,206
	Adult brain - H3K4me3	230,791
	Adult brain - H3K9ac	308,869
	Adult brain - H3K27me3	302,550
	NeuN(-) - H3K4me3	112,615
	NeuN(+): Fetal/Infant - H3K4me3	141,445
	NeuN(+): Childhood/Adolescence - H3K4me3	130,728
	NeuN(+): Early/Middle Adulthood - H3K4me3	150,203
	NeuN(+): Late Adulthood - H3K4me3	134,636
	Fetal brain - DHS	276,468
	Fetal brain - H3K4me1	202,424
	Fetal brain - H3K4me3	191,884
	Fetal brain - H3K27me3	437,175
	Primary Cell Culture/iPS - H3K4me1	768,697
	Primary Cell Culture/iPS - H3K4me3	166,256
	Primary Cell Culture/iPS - H3K27me3	813,572
	Active promoter	209,349
	Active enhancer	419,981
	Poised promoter	81,260
	Repressed enhancer	165,117
creSNP	Adult brain - H3K27ac	107,640
	Adult brain - H3K4me1	110,442
	Adult brain - H3K4me3	55,651
	Adult brain - H3K9ac	74,973
	Adult brain - H3K27me3	40,766
	NeuN(-) - H3K4me3	28,854
	NeuN(+): Fetal/Infant - H3K4me3	34,784
	NeuN(+): Childhood/Adolescence - H3K4me3	32,438
	NeuN(+): Early/Middle Adulthood - H3K4me3	36,882
	NeuN(+): Late Adulthood - H3K4me3	33,826
	Fetal brain - DHS	46,770
	Fetal brain - H3K4me1	38,386
	Fetal brain - H3K4me3	46,720
	Fetal brain - H3K27me3	76,217
	Primary Cell Culture/iPS - H3K4me1	146,275

	Primary Cell Culture/iPS - H3K4me3	41,853
	Primary Cell Culture/iPS - H3K27me3	136,248
	Active promoter	51,728
	Active enhancer	95,701
	Poised promoter	13,743
	Repressed enhancer	25,547
FUV	Functionally unannotated variants (FUV)	6,599,035

Table S5, related to Table 1. SNP enrichment for different GWAS P values

Type		Functional category	SNP enrichment for different GWAS *P* values*		
			$P < 10^{-3}$	$P < 10^{-5}$	$P < 5x10^{-8}$
eSNP		eSNP	3.68	8.07	14.13
CRE	integrated	Active promoter	2.13	2.88	4.57
		Active enhancer	2.30	3.16	7.08
		DHS	1.58	2.39	5.13
		Poised promoter	1.40	1.02	2.29
		Repressed enhancer	1.13	0.82	1.51
	non-integrated	Adult brain - H3K27ac	2.25	3.08	6.31
		Adult brain - H3K4me1	2.27	3.54	6.92
		Adult brain - H3K4me3	1.94	2.75	4.17
		Adult brain - H3K9ac	2.24	3.40	7.24
		Adult brain - H3K27me3	0.89	1.02	0.42
		NeuN(-) - H3K4me3	1.82	2.94	4.79
		NeuN(+): Fetal/Infant - H3K4me3	1.92	2.56	4.68
		NeuN(+): Childhood/Adolescence - H3K4me3	1.96	2.88	5.50
		NeuN(+): Early/Middle Adulthood - H3K4me3	1.91	2.94	5.13
		NeuN(+): Late Adulthood - H3K4me3	2.03	3.42	6.61
		Fetal brain - DHS	1.58	2.39	5.13
		Fetal brain - H3K4me1	2.13	3.38	7.59
		Fetal brain - H3K4me3	2.08	2.81	4.07
		Fetal brain - H3K27me3	0.88	0.44	NA
		Primary Cell Culture/iPS - H3K4me1	1.80	2.27	3.63
		Primary Cell Culture/iPS - H3K4me3	2.04	2.81	4.37
		Primary Cell Culture/iPS - H3K27me3	1.11	1.20	0.98
creSNP	integrated	Active promoter	5.17	9.31	17.38
		Active enhancer	5.66	11.46	29.51
		DHS	4.06	10.94	26.92
		Poised promoter	4.03	5.36	13.80

		Repressed enhancer	3.12	4.67	9.55
	non-integrated	**Adult brain - H3K27ac**	5.63	11.46	26.30
		Adult brain - H3K4me1	5.73	13.46	30.20
		Adult brain - H3K4me3	4.74	9.10	16.98
		Adult brain - H3K9ac	5.54	11.51	28.84
		Adult brain - H3K27me3	2.54	6.01	3.24
		NeuN(-) - H3K4me3	4.50	9.31	19.05
		NeuN(+): Fetal/Infant - H3K4me3	4.73	8.11	19.05
		NeuN(+): Childhood/Adolescence - H3K4me3	4.66	8.49	20.42
		NeuN(+): Early/Middle Adulthood - H3K4me3	4.67	9.10	17.78
		NeuN(+): Late Adulthood - H3K4me3	5.04	10.69	22.91
		Fetal brain - DHS	4.06	10.94	26.92
		Fetal brain - H3K4me1	5.30	14.09	38.02
		Fetal brain - H3K4me3	4.92	8.89	14.13
		Fetal brain - H3K27me3	2.29	1.69	NA
		Primary Cell Culture/iPS - H3K4me1	4.65	9.64	17.78
		Primary Cell Culture/iPS - H3K4me3	4.85	9.10	17.38
		Primary Cell Culture/iPS - H3K27me3	2.84	4.89	5.37

* SNP enrichment for different GWAS P values illustrates the estimated change in terms of the proportion of SNPs in different functional categories that reach different P values ($P < 10^{-3}$; $P < 10^{-5}$; $P < 5 \times 10^{-8}$) in comparison to functionally unannotated variants
NA indicates that no SNPs exceed the specified GWAS P value

Table S6, related to Figure 1 and 2 and Table 1. Per functional category average per SNP total tagged LD, MAF and P values of the enrichment for all functional categories compared to FUV estimated with the Kolmogorov-Smirnov statistic after removing SNPs with r^2 > 0.1 and those that were <100 kb from a more strongly associated variant in the SCZ study.

Dataset		average sum r^2	Average MAF	D	*P* value
FUV		72.2	0.135	-	-
eSNP		142.9	0.232	0.1021	<2.2E-16
CRE	**Adult brain - H3K27ac**	61.5	0.141	0.0247	<2.2E-16
	Adult brain - H3K4me1	63.4	0.142	0.0239	<2.2E-16
	Adult brain - H3K4me3	62.9	0.139	0.0393	<2.2E-16
	Adult brain - H3K9ac	61.3	0.140	0.0321	<2.2E-16
	Adult brain - H3K27me3	45.5	0.150	0.0258	<2.2E-16
	NeuN(-) - H3K4me3	65.5	0.137	0.0468	<2.2E-16
	NeuN(+): Fetal/Infant - H3K4me3	63.1	0.137	0.0433	<2.2E-16
	NeuN(+): Childhood/Adolescence - H3K4me3	64.1	0.137	0.044	<2.2E-16
	NeuN(+): Early/Middle Adulthood - H3K4me3	63.4	0.138	0.0442	<2.2E-16
	NeuN(+): Late Adulthood - H3K4me3	64.5	0.137	0.0445	<2.2E-16
	Fetal brain - DHS	60.4	0.141	0.0406	<2.2E-16
	Fetal brain - H3K4me1	53.3	0.140	0.0324	<2.2E-16
	Fetal brain - H3K4me3	63.1	0.137	0.0401	<2.2E-16
	Fetal brain - H3K27me3	48.8	0.148	0.0235	<2.2E-16
	Primary Cell Culture/iPS - H3K4me1	57.9	0.143	0.0228	<2.2E-16
	Primary Cell Culture/iPS - H3K4me3	65.5	0.137	0.0427	<2.2E-16
	Primary Cell Culture/iPS - H3K27me3	50.2	0.150	0.0162	<2.2E-16
	Active promoter	61.8	0.138	0.0386	<2.2E-16
	Active enhancer	60.8	0.141	0.0254	<2.2E-16
	Poised promoter	45.4	0.141	0.0472	<2.2E-16

	Repressed enhancer	45.1	0.144	0.036	<2.2E-16
creSNP	**Adult brain - H3K27ac**	112.5	0.233	0.072	<2.2E-16
	Adult brain - H3K4me1	114.0	0.232	0.0596	<2.2E-16
	Adult brain - H3K4me3	120.8	0.232	0.0642	<2.2E-16
	Adult brain - H3K9ac	113.0	0.233	0.0687	<2.2E-16
	Adult brain - H3K27me3	87.6	0.229	0.0291	2.10E-05
	NeuN(-) - H3K4me3	124.5	0.232	0.0627	<2.2E-16
	NeuN(+): Fetal/Infant - H3K4me3	121.1	0.232	0.0638	<2.2E-16
	NeuN(+): Childhood/Adolescence - H3K4me3	122.2	0.232	0.0619	<2.2E-16
	NeuN(+): Early/Middle Adulthood - H3K4me3	121.2	0.232	0.065	<2.2E-16
	NeuN(+): Late Adulthood - H3K4me3	122.4	0.232	0.063	<2.2E-16
	Fetal brain - DHS	115.9	0.231	0.0342	3.33E-16
	Fetal brain - H3K4me1	102.8	0.227	0.0541	<2.2E-16
	Fetal brain - H3K4me3	123.0	0.231	0.0701	<2.2E-16
	Fetal brain - H3K27me3	95.6	0.233	0.0439	<2.2E-16
	Primary Cell Culture/iPS - H3K4me1	112.4	0.232	0.0599	<2.2E-16
	Primary Cell Culture/iPS - H3K4me3	126.2	0.232	0.0657	<2.2E-16
	Primary Cell Culture/iPS - H3K27me3	92.0	0.234	0.0486	<2.2E-16
	Active promoter	117.9	0.232	0.068	<2.2E-16
	Active enhancer	111.4	0.233	0.0679	<2.2E-16
	Poised promoter	93.0	0.233	0.0286	0.001345
	Repressed enhancer	82.7	0.230	0.0324	1.87E-06

Table S7, related to Figure 1 and 2 and Table 1. Multiple regression analysis in the SCZ GWAS

Variables	coefficient	SE	95% CI	*P* value
Intercept	-1.4218	0.0012	(-1.424,-1.419)	$< 2 \times 10^{-16}$
LD (sum r^2)	0.0006	5.72E-06	(0.0006,0.00063)	$< 2 \times 10^{-16}$
Genetic variance (MAF*(1-MAF))	1.1652	0.0086	(1.1485,1.182)	$< 2 \times 10^{-16}$
eSNP	0.1260	0.0021	(0.1219,0.1301)	$< 2 \times 10^{-16}$
DHS	0.0112	0.0046	(0.0021,0.0203)	0.016
Active promoter	0.0266	0.0067	(0.0135,0.0398)	7.2×10^{-5}
Active enhancer	0.0395	0.0045	(0.0306,0.0483)	$< 2 \times 10^{-16}$
Poised promoter	-0.0241	0.0100	(-0.0437,-0.0046)	0.016
Repressed enhancer	-0.0128	0.0068	(-0.0262,0.0005)	0.060

Table S8, related to Figure 5. Demographics of postmortem cohort used for the 3C experiments

Subject	Gender	Age	PMI (hr)	pH	Condition	Match
35	Male	52	12	6.8	Control	1
28	Male	31	15	6.6	Control	17
11	Male	46	19	6.7	Control	40
1	Male	53	11	6.8	Schizophrenia	35
17	Male	31	14	6.5	Schizophrenia	28
40	Male	44	14	6.6	Schizophrenia	11

Table S9, related to Figure 5. Primers used for the *CACNA1C* 3C experiments

Name	Sequence	Restriction enzyme	Position	Size	GC %	T_m
anchor_primer_*CACNA1C_HindIII*	GCTAAGCATTGCTGAAAATCACAAGGCTAG	*HindIII*	chr12:2155780-2155809	30	30	43.3
anchor_primer_*CACNA1C_NcoI*	CCTCGGAGGAGGGATTAATCCAGAC	*NcoI*	chr12:2162635-2162659	25	56	60.1
P1	CTCCTCACTCCCAGGAATCACTCAG	*NcoI*	chr12:2308557-2308589	33	56	60.1
P2	CCACAACAAAGAATCATCTGGCACAAAATG	*HindIII*	chr12:2309660-2309689	30	30	40
P3	GTTACCATTATATGCAGTTCTCCCAGCCA	*NcoI*	chr12:2345793-2345821	29	44.8	60
P4	GTTGGCCTGTTTTGAGACATAACCAGTTGG	*HindIII*	chr12:2346778-2346807	30	30	46.7
P5	GTATCTCCTGTGATGGAGTGACGAGGTAAC	*HindIII*	chr12:2347400-2347429	30	30	50
P6	TTGCTCCCCTTCTTGCTGGAGTTAA	*NcoI*	chr12:2349421-2349445	25	48	60.3
P7	AAGAACCTAAAGACTTGGGCCCTGG	*NcoI*	chr12:2351870-2351894	25	52	60.8
P8	CTAAGCTCTTACTCTTCCCCAGTTTTTACTTCC	*NcoI*	chr12:2361930-2361962	33	42.4	59.6
P9	TCTCACCTTCTGGCAGGGTAAGTACATT	*HindIII*	chr12:2362375-2362402	28	46.4	60.8
P10	GAGGGAGGCAGAGTTTCTTTGTACTACATC	*HindIII*	chr12:2367214-2367243	30	30	46.7
P11	GAAAAGGAGAGGAGCACAGATTGCTCA	*NcoI*	chr12:2369704-2369731	28	48.1	60.4
P12	CTTGATTAGACCTGCAGAATGTCTGCACTG	*HindIII*	chr12:2373539-2373568	30	30	46.7
P13	GGACCATCTTCCCTGCAAATGCAG	*NcoI*	chr12:2374762-2374786	25	54.2	60.5
P14	TCACAGAATTCTACTGAAAATGCAGGGTATGG	*NcoI*	chr12:2392522-2392553	32	40.6	59.9
P15	GCTAAATACTCGTCATTTATACCTGGCCGG	*HindIII*	chr12:2395289-2395318	30	30	46.7
P16	GAGAAAATACAGATTCCTAGGCCTCAT	*HindIII*	chr12:2408739-2408765	27	30	43.3
P17	CAGGAGCAGGACGAGAGGATGTGTAATTAA	*NcoI*	chr12:2409886-2409916	31	46.7	60.9

Supplementary Figures

Figure S1, related to Figure 1 and Table 1. Stratified Q-Q plots show enrichment for eSNP and CRE or eSNP and creSNP categories in the adult brain-homogenate **(a, b)**, fetal brain-homogenate **(c, d)**, brain-FACS **(e, f)** and Primary Cell Culture/iPS **(g, h)** tissue. All empirical null distributions were corrected for inflation by using the FUV inflation control. The major histocompatibility complex locus (chr6: 25-35Mb) was excluded from the SCZ dataset.

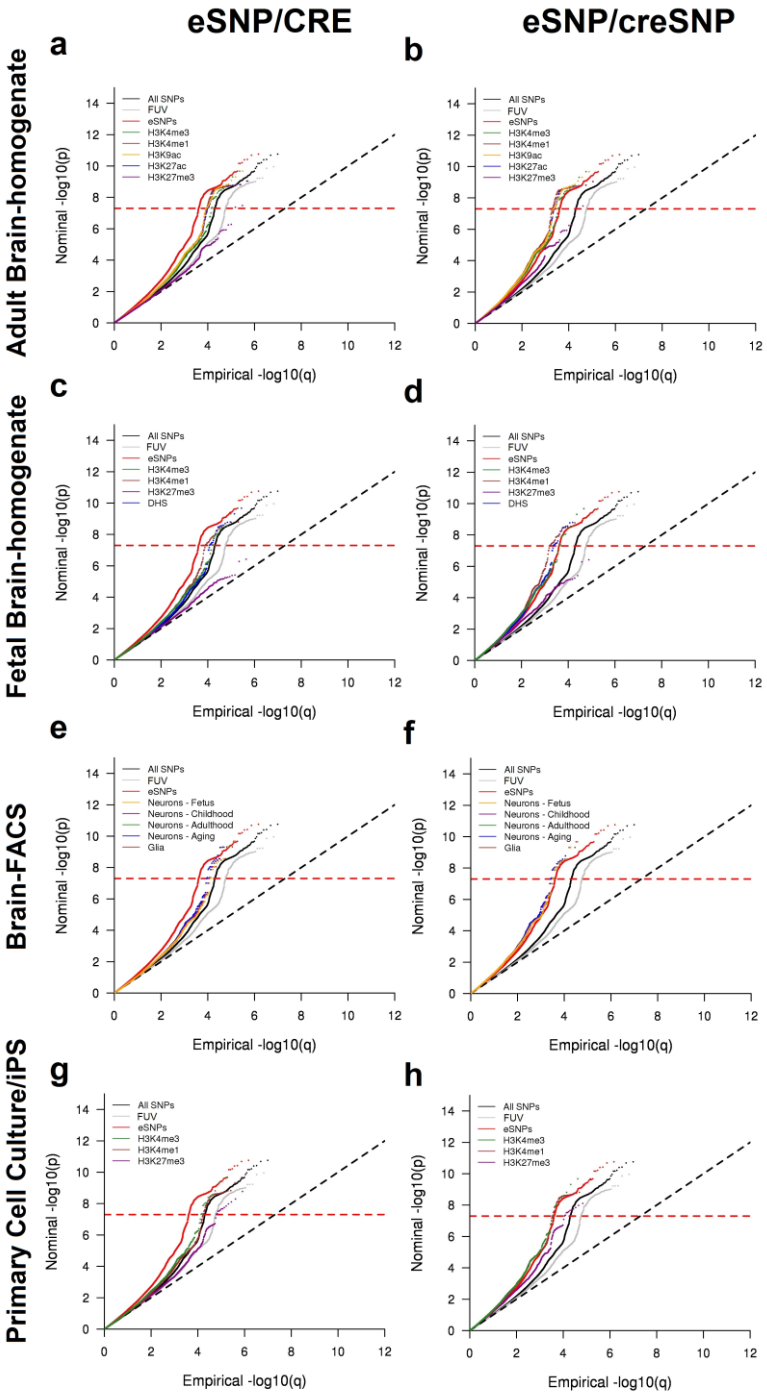

Figure S2, related to Figure 2. Categorical enrichment for the individual functional annotation as measured by the CES. The CES are scaled using the maximum value across functional categories (H3K4me1 in fetal brain tissue). For each functional category, we performed 10,000 permutations to calculate the null distribution of categorical enrichment for comparison to observed categorical enrichment. Functional categories with empirical *P* values ≤ 10^{-4} are indicated with asterisk (*). All summary statistics were corrected for inflation by using the functionally unannotated variant (FUV) inflation control. The major histocompatibility complex locus (chr6: 25-35Mb) was excluded from the SCZ dataset.

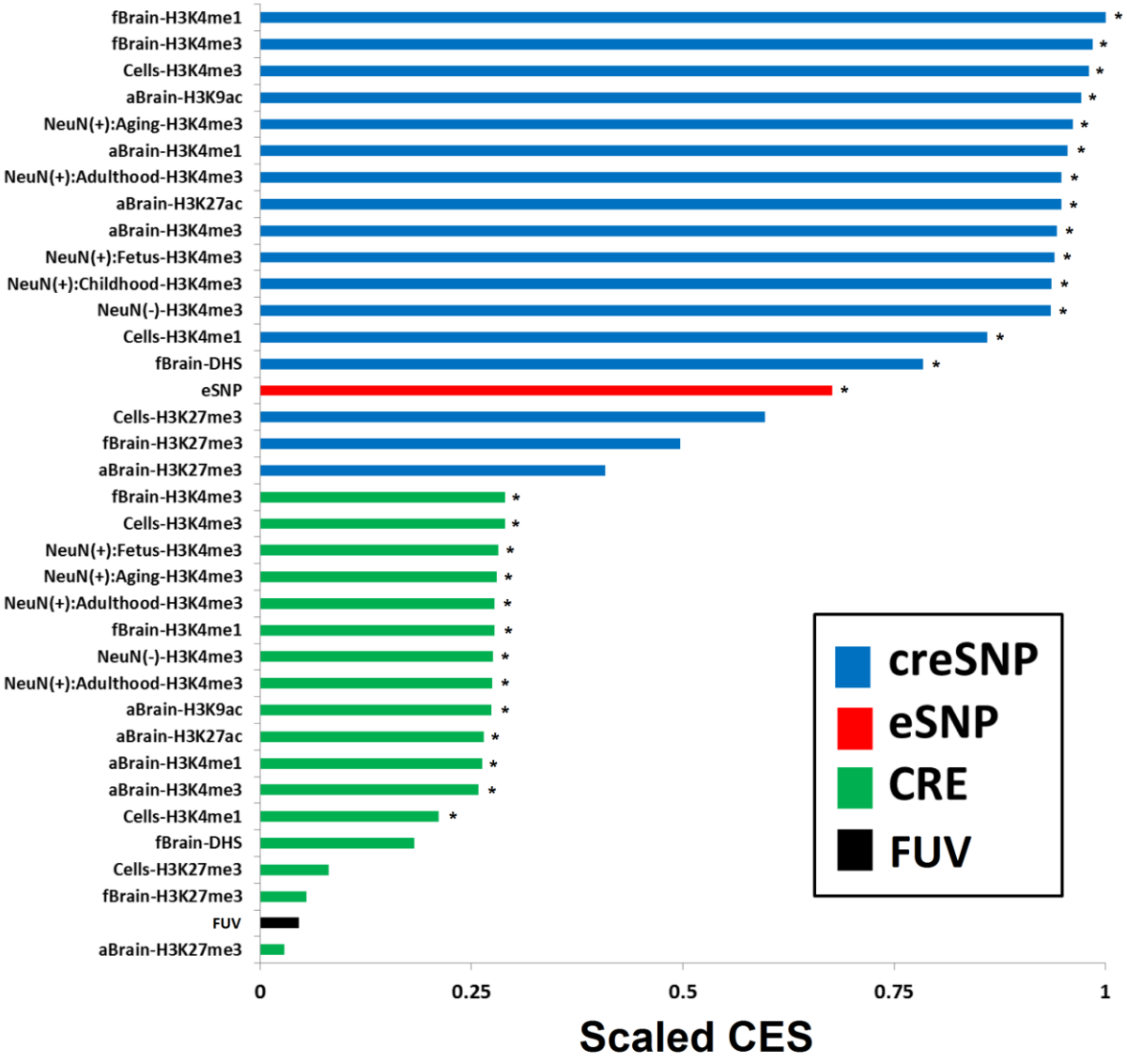

Figure S3, related to Figure 1 and Table 1. Stratified Q-Q plots show enrichment for combined functional annotation categories in SCZ. SNPs with $r^2 > 0.1$ and those that were <100 kb from a more strongly associated variant in the SCZ study were removed for each functional category. Higher enrichment was observed across the eSNP and **(a)** combined CRE or **(b)** creSNP functional annotation categories. We computed significance values for the curves for each functional annotation category relative to those for FUV SNPs, using a two-sample Kolmogorov-Smirnov Test. The enrichment was significant for all functional categories when compared with the FUV category. The major histocompatibility complex locus (chr6: 25-35Mb) was excluded from the SCZ dataset.

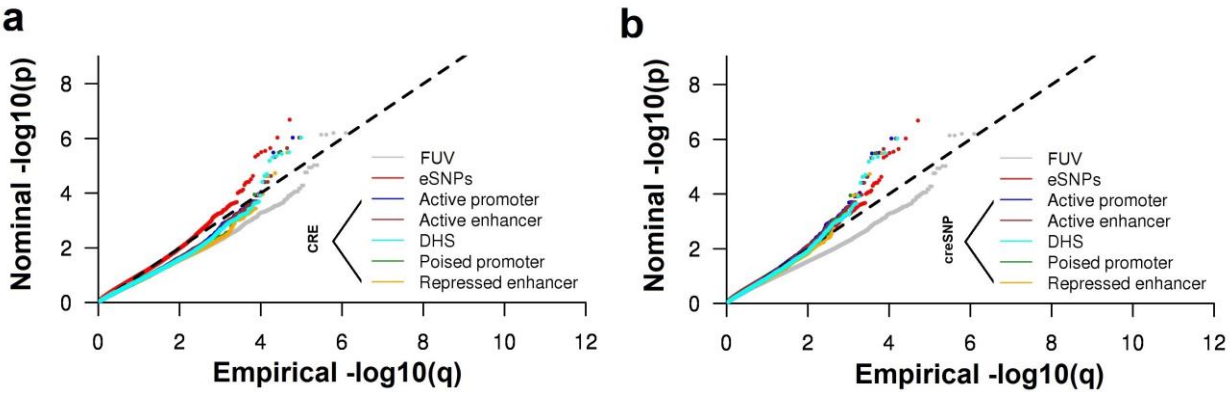

Figure S4, related to "Methods: *cis* regulatory element annotations". Cross-correlation map generated by clustering histone modification and DHS peaks based on pair-wise correlations. Heat indicates degree of positive (blue) or negative (yellow) correlation between data sets. **(a)** Brain related epigenomic data (n=140) that were used in the current study. Tissue: adult brain -homogenate (red); fetal brain - homogenate (black); brain - FACS: NeuN(-) (green); brain - FACS: NeuN(+) (cyan); primary cell culture/iPS (blue). **(b)** T-helper cell related epigenomic data (n=50). **(c)** Liver related epigenomic data (n=14). **(d)** Skin related epigenomic data (n=46). **(e)** Adipose tissue related epigenomic data (n=41) that were used in the current study. Tissue: adipose nuclei (black); adipose derived mesenchymal stem cells (red).

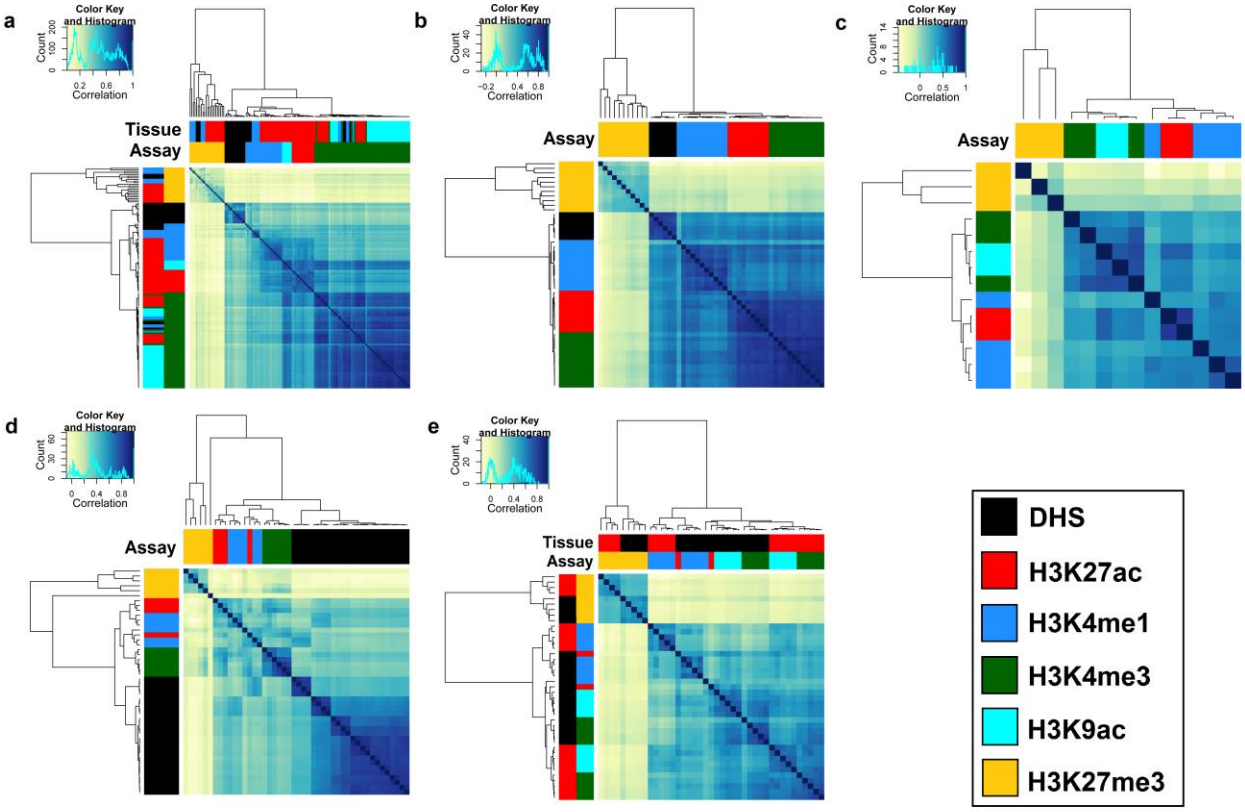

Figure S5, related to Figure 5. Chromosome conformation capture library controls. A) 3C libraries were made with (+ligase) and without (-ligase) ligase. Two independent aliquots of ligase libraries were created. B) 3C libraries were normalized using a neighboring primer (left) and quantified (right). C) A BAC control was generated for the *CACNA1C* region using equimolar concentrations of RP11-465I2 and RP11-698B24 (left) to normalize for primer efficiency.