

How sex affects transcriptional associations with schizophrenia across dorsolateral prefrontal cortex, hippocampus, and caudate nucleus

Authors: Kynon JM Benjamin^{1-3*+}, Ria Arora^{1,4+}, Arthur S Feltrin¹, Geo Perteau¹, Hunter H Giles^{1,5}, Joshua M Stolz¹, Laura D'Ignazio^{1,3}, Leonardo Collado-Torres^{1,6}, Joo Heon Shin¹, William S Ulrich¹, Thomas M Hyde¹⁻³, Joel E Kleinman^{1,2}, Daniel R Weinberger^{1-3,5,7}, Apuã CM Paquola^{1,3*}, Jennifer A Erwin^{1-3,7*}

¹Lieber Institute for Brain Development, Baltimore, MD, USA

²Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA

³Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

⁴Department of Biology, Johns Hopkins University Krieger School of Arts & Sciences, Baltimore, MD, USA

⁵Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

⁶Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA

⁷Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

*Corresponding authors

+Co-first authors

Abstract

Schizophrenia is a complex neuropsychiatric disorder with sexually dimorphic features, including differential symptomatology, drug responsiveness, and male incidence rate. To date, only the prefrontal cortex has been studied in large-scale transcriptome analyses for sex differences in schizophrenia. Here, we examined the BrainSeq Consortium RNA-sequencing and genotypes for the caudate nucleus (n=399), dorsolateral prefrontal cortex (DLPFC; n=377), and hippocampus (n=394) to characterize sex differences in schizophrenia. We identified genomic features (genes, transcripts, exons, and exon-exon junctions) associated with sex, sex-specific expression in schizophrenia, and sex-interacting expression quantitative trait loci (si-eQTL) associated with schizophrenia risk. We found 831 unique genes that exhibit sex differences across brain regions, enriched for immune-related pathways. X-chromosome dosage was significantly decreased in the hippocampus of male individuals with schizophrenia. Our sex interaction model revealed 148 junctions dysregulated for schizophrenia in a sex-specific manner. Sex-specific schizophrenia analysis identified dozens of expressed, sex-specific features also with enrichment in immune-related pathways. Finally, our si-eQTL analysis revealed 704 unique genes, nine of which are associated with schizophrenia risk. Overall, our results increased the number of annotated sex-biased features, identified sex-specific schizophrenia genes, and provided the first annotation of si-eQTL in the human DLPFC and hippocampus. Altogether, these results point to the importance of sex-informed analysis of sexually dimorphic traits, inform personalized therapeutic strategies in schizophrenia, and highlight the need for increased female samples for schizophrenia analyses.

Introduction

For more than a century, sex differences have been observed in schizophrenia – a complex, chronic neuropsychiatric disorder affecting approximately 1% of the adult population worldwide. These sex differences include differences in cognitive severity and age of onset; for example, female individuals appearing to be less vulnerable to altered verbal processing deficits¹, and male individuals having an earlier age of disease onset^{2,3}. Additionally, prenatal stress may significantly increase the risk of schizophrenia in male offspring as opposed to female offspring^{4,5}. To date, only two large-scale RNA-sequencing studies have examined sex differences in schizophrenia, and both focus exclusively on one brain region – the prefrontal cortex^{6,7}. Furthermore, the GTEx (Genotype-Tissue Expression) analysis of sex differences across 45 tissues found fewer than 100 differentially expressed genes (DEGs) in 13 of 14 brain regions. This small number of identified DEGs might be attributed to relatively limited sample size (114 to 209 individuals)⁸.

Leveraging schizophrenia genome-wide association studies (GWAS)^{9–11}, recent large-scale studies have used statistical association between genotype and expression (expression quantitative trait loci [eQTL]) to identify genomic features (gene, transcript, exon, and exon-exon junctions) underlying schizophrenia risk^{12–16}. However, these studies have not explored potential sex-interacting eQTL (si-eQTL). Furthermore, the GTEx study across 44 tissues found only four si-eQTL genes with a nominally significant false discovery rate (FDR) below 0.25 in only two of the 13 brain regions examined⁸. Other recent si-eQTL studies involving whole blood^{17,18} and lymphoblastoid cell lines¹⁹ have found fewer than 25 si-eQTL. As eQTL have tissue specificity²⁰, the field of neuropsychiatric genetics needs a sizable and comprehensive analysis of si-eQTL in the human brain.

Here, we leverage the BrainSeq Consortium RNA-sequencing and genotypes datasets to identify genes associated with sex, with sex-specific expression in schizophrenia, and with sex-interacting eQTL using a total of 1,170 samples across 504 individuals (**Table 1**) for the caudate nucleus (n=399), dorsolateral prefrontal cortex (DLPFC; n=377), and hippocampus (n=394). Our work increases the number of annotated sex-biased features, examines sex-chromosome dosage, identifies sex-specific schizophrenia features, provides the first annotation of si-eQTL in the human DLPFC and hippocampus, and increases si-eQTL annotations for the caudate nucleus. Altogether, these results provide novel insights into sex differences, highlighting the importance of sex-informed analysis of sexually dimorphic traits and informing personalized therapeutic strategies in schizophrenia.

Table 1. A sample breakdown of eQTL analysis for individuals (age > 13) postmortem caudate nucleus, DLPFC, and hippocampus from the BrainSeq Consortium, separated by sex. Abbreviations: Female (F), Male (M), Neurotypical Control (CTL), Schizophrenia (SZ), African American (AA), European American (EA), RNA integrity number (RIN).

Brain Region	Sex	Sample Size	Diagnosis	Ancestry	Age (mean \pm sd)	RIN (mean \pm sd)
Caudate Nucleus	F	126	76CTL/50SZ	79AA/47EA	50.2 \pm 16.9	7.8 \pm 0.9
	M	273	169CTL/104SZ	127AA/146EA	48.6 \pm 15.7	7.9 \pm 0.8
DLPFC	F	121	73CTL/48SZ	75AA/46EA	48.4 \pm 17.1	7.4 \pm 1.0
	M	256	156CTL/100SZ	129AA/127EA	44.6 \pm 16.1	7.8 \pm 0.9
Hippocampus	F	126	79CTL/47SZ	82AA/44EA	47.9 \pm 16.7	7.64 \pm 1.1
	M	268	182CTL/86SZ	131AA/137EA	44.4 \pm 16.2	7.7 \pm 1.0

Materials and Methods

The research described herein complies with all relevant ethical regulations. All specimens used in this study were obtained with informed consent from the next of kin under protocols No. 12-24 from the Department of Health and Mental Hygiene for the Office of the Chief Medical Examiner for the State of Maryland and No. 20111080 for the Western Institutional Review Board for the Offices of: 1) the Chief Medical Examiner for Kalamazoo Michigan, 2) University of North Dakota in Grand Forks North Dakota, and 3) Santa Clara County California. Details of case selection, curation, diagnosis, and anatomical localization and dissection can be found in previous publications from our research group^{12,13}.

BrainSeq Consortium RNA-sequencing data processing

We surveyed data from the BrainSeq Consortium^{12,13} for caudate nucleus, DLPFC, and hippocampus, specifically: phenotype information, FASTQ files, region-specific covariates, and single nucleotide polymorphism (SNP) array genotypes.

We re-mapped RNA-sequencing reads to the hg38/GRCh38 human reference genome (GENCODE release 41, GRCh38.p13) with the splice-aware aligner HISAT2²¹ (version 2.2.1). Following alignment, we collected quality control and alignment metrics for each sample using RNA-SeQC²² (version 2.4.2).

We performed quantification of major genomic features (genes, transcripts, exons, and exon-exon junctions) for each sample separately, as follows:

- We generated gene and exon read counts using featureCounts²³ (version 2.0.3) with default parameters for paired-end, reverse-stranded read counting.
- We estimated transcript expression (i.e., counts and transcripts per million [TPM]) with kallisto²⁴ (version 0.46.2) with default parameters for reverse-stranded reads.
- We extracted exon-exon junction coverage data for all spliced alignments found in the alignment files produced by HISAT2 using RegTools²⁵ (version 0.5.3). Using this reference-free method, we were able to include novel splicing patterns detected from the aligned files.

Following this quality control and quantitation, we packaged these data (i.e., counts, gene annotation, and quality control metrics) into RangedSummarizedExperiment R objects²⁶ using R code adapted from the SPEAQeasy RNA-seq processing pipeline²⁷.

BrainSeq Consortium imputation and genotype processing

We imputed genotypes as previously described¹². Briefly, we first converted genotype positions from hg19 to hg38 with liftOver²⁸. The Trans-Omics for Precision Medicine (TOPMed) imputation server^{29–31} was used for imputation of genotypes filtered for high quality (removing

low-quality and rare variants) using the genotype data phased with the Haplotype Reference Consortium (HRC) reference panels (https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html). Genotypes were phased per chromosome using Eagle³² (version 2.4). We performed quality control with the McCarthy Tools (<https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.3.0.zip>): specifically, we removed variants and samples with minor allele frequency (MAF) less than 0.01, missing call frequencies greater than 0.1, and Hardy-Weinberg equilibrium below a p-value of 1e-10 using PLINK 2.0^{33–35} (version 2.00a3LM). This resulted in 11,474,007 common variants.

For population stratification of samples, we performed multidimensional scaling (MDS) with PLINK version 1.9^{33–35} on linkage disequilibrium (LD)-independent variants. The first component separated samples by ancestry as reported by the medical examiner's offices.

Sample selection

We selected samples from the caudate nucleus, DLPFC, and hippocampus based on four inclusion criteria: 1) used RiboZero RNA-sequencing library preparation, 2) features an age greater than 13 years, 3) has a self-reported ethnicity of either African American or White American, and 4) has TOPMed imputed genotypes available. This resulted in a total of 1,170 samples from 504 unique individuals across the three brain regions for eQTL analysis. For expression-based analysis, we excluded individuals with age less than 17 years, resulting in a total of 1,127 samples from 480 unique individuals across the three brain regions.

Subject details

Of all 1,170 samples used in the eQTL portion of this study, 399 were from the caudate nucleus, 377 from the DLPFC, and 394 from the hippocampus. Out of the 1,170 samples, 126, 121, and 126 were female and 273, 256, 268 were male from the caudate nucleus, DLPFC, and hippocampus, respectively (**Table 1**). For the 1,127 samples used in the expression analyses of this study, 393, 359, and 375 samples were located in the caudate nucleus, DLPFC, and hippocampus, respectively (**Table 2**). More information can be found in **Tables 1 and 2**.

Table 2. A sample breakdown of expression analysis for adult (age > 17) postmortem caudate nucleus, DLPFC, and hippocampus from the BrainSeq Consortium. Abbreviations: Neurotypical Control (CTL), Schizophrenia (SZ), Female (F), Male (M), African American (AA), European American (EA), RNA integrity number (RIN).

Brain Region	Sample Size	Diagnosis	Sex	Race	Age (mean ± sd)	RIN (mean ± sd)
--------------	-------------	-----------	-----	------	-----------------	-----------------

Caudate Nucleus	393	239CTL/154SZ	121F/272M	205AA/188EA	49.6 ± 15.6	7.9 ± 0.9
DLPFC	359	211CTL/148SZ	114F/245M	200AA/159EA	47.4 ± 15.4	7.7 ± 0.9
Hippocampus	375	242CTL/133SZ	121F/254M	207AA/168EA	47.0 ± 15.3	7.6 ± 1.0

Match gender phenotype to sex chromosomes

To match gender phenotype with sex chromosomes, we applied the sex imputation function (*--check-sex*) from PLINK. This compares sex assignments in the input dataset with those imputed from X-chromosome inbreeding coefficients. We used a Jupyter Notebook (version 6.0.2) with the R kernel to compare reported gender with genotype-imputed sex (F estimates). Here, we found all gender phenotypes matched sex chromosomes with F estimates for females below 0.22 and males above 0.9 (**Fig. S1**).

Quality control and covariate exploration for sex

Observed expression measurements can be affected by biological and technical factors. To evaluate potential confounders for expression or sex, we first correlated technical and RNA quality variables (i.e., RIN, mitochondria mapping rate, overall mapping rate, total gene assignment, mean 3' bias, etc.) and removed highly correlated variables (Pearson, $r > 0.95$; **Fig. S2**) present in at least one brain region. To examine potential confounders, we next correlated these remaining variables and biological variables (i.e., diagnosis, age at death, global genetic similarity, status of antipsychotics at time of death, etc.) with gene expression as a function of sex (**Fig. S3A**). For model covariates, we used variables that had a significant correlation (Bonferroni corrected p-value < 0.05) with gene expression for either sex in at least one brain region. To account for possible hidden effects on gene expression not captured by the above covariates, we also applied surrogate variable analysis^{36,37}. When we regressed out biological, technical, and hidden effects, we found this reduced all spurious correlations (**Fig. S3B**).

Expression normalization

For expression normalization, we constructed edgeR^{38,39} objects in R (version 4.2) for each brain region by using raw counts and sample phenotype information. Next, we filtered out low expression counts using *filterByExpr* from edgeR (version 3.40.2), which keeps features above a minimum of 10 count-per-million (CPM) in 70% of the smallest group sample size (i.e., female individuals). Following this, we normalized library size using trimmed mean of M-values (TMM) before applying voom normalization⁴⁰ using limma⁴¹ (version 3.54.1) on four different

linear models that examine: 1) sex (**Equation 1**), 2) interaction of brain region and sex (**Equation 2**), 3) interaction of sex and diagnosis (**Equation 3**), and 4) diagnosis subset by sex (**Equation 4**). Example covariates for these linear models are diagnosis, age, brain region, genetic similarity (SNP PCs [principle components] 1-3), RNA quality (RIN, mitochondria mapping rate, gene assignment rate, genome mapping rate, rRNA mapping rate, and mean 3' bias). For sex, interaction of sex and diagnosis, and diagnosis by sex analyses, we also corrected for any hidden variance via surrogate variable analysis.

$$E(Y) = \beta_0 + \beta_1 Sex + \beta_2 Diagnosis + \beta_3 Age + \beta_4 MitoRate + \beta_5 rRNArate + \beta_6 TotalAssignedGenes + \beta_7 RIN + \beta_8 OverallMappingRate + \beta_9 Mean3Bias + \sum_{i=1}^3 \eta_i snpPC_i + \sum_{j=1}^k \gamma_j SV_j$$

Equation 1

$$E(Y) = \beta_0 + \beta_1 Sex * \beta_2 Region + \beta_3 Diagnosis + \beta_4 Age + \beta_5 MitoRate + \beta_6 rRNArate + \beta_7 TotalAssignedGenes + \beta_8 RIN + \beta_9 OverallMappingRate + \beta_9 Mean3Bias + \sum_{i=1}^3 \eta_i snpPC_i$$

Equation 2

$$E(Y) = \beta_0 + \beta_1 Sex * \beta_2 Diagnosis + \beta_3 Age + \beta_4 MitoRate + \beta_5 rRNArate + \beta_6 TotalAssignedGenes + \beta_7 RIN + \beta_8 OverallMappingRate + \beta_9 Mean3Bias + \sum_{i=1}^3 \eta_i snpPC_i + \sum_{j=1}^k \gamma_j SV_j$$

Equation 3

$$E(Y) = \beta_0 + \beta_1 Diagnosis + \beta_2 Age + \beta_3 MitoRate + \beta_4 rRNArate + \beta_5 TotalAssignedGenes + \beta_6 RIN + \beta_7 OverallMappingRate + \beta_8 Mean3Bias + \sum_{i=1}^3 \eta_i snpPC_i + \sum_{j=1}^k \gamma_j SV_j$$

Equation 4

As there was significant overlap of individuals among the three brain regions examined, we used *dream* (differential expression for repeated measures) from variancePartition⁴² (version 1.28.7) to correct for the random effect of duplicate individuals across brain regions to assess potential significant interactions between brain region and sex (**Equation 2**). As such, we applied voom via dream framework with *voomWithDreamWeights*.

Expression residualization

For residualized expression, we used voom-normalized expression and null models to regress out covariates as previously described¹². After regressing out covariates, we applied a z-score transformation. Null models were created without variable(s) of interest to examine: 1) sex (**Equation 4**), 2) interaction of brain region and sex (**Equation 5**), 3) interaction of sex and diagnosis (**Equation 6**), and 4) diagnosis (**Equation 6**).

$$E(Y) = \beta_0 + \beta_1 \text{Diagnosis} + \beta_2 \text{Age} + \beta_3 \text{MitoRate} + \beta_4 \text{rRNArate} + \beta_5 \text{TotalAssignedGenes} + \beta_6 \text{RIN} + \beta_7 \text{OverallMappingRate} + \beta_8 \text{Mean3Bias} + \sum_{i=1}^3 \eta_i \text{snpPC}_i$$

Equation 5

$$E(Y) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{MitoRate} + \beta_3 \text{rRNArate} + \beta_4 \text{TotalAssignedGenes} + \beta_5 \text{RIN} + \beta_6 \text{OverallMappingRate} + \beta_7 \text{Mean3Bias} + \sum_{i=1}^3 \eta_i \text{snpPC}_i + \sum_{j=1}^k \gamma_j \text{SV}_j$$

Equation 6

Differential expression analysis

Following voom normalization, we fit four linear models (**Equations 1-4**) to examine: 1) sex, 2) interaction of sex and brain region, 3) interaction of sex and diagnosis, and 4) diagnosis subset by sex. With our fitted model, we identified differentially expressed features using the *eBayes*⁴³ function from limma. Dream enabled us, in one step, to complete linear model fitting and differential expression calculation for interaction of sex and brain region.

Weighted correlation network analysis (WGCNA) analysis

We performed a signed network WGCNA⁴⁴ (version 1.72) analysis using residualized expression (**Equation 4**) to generate the co-expression network using all genes in a single block by brain region. First, we filtered genes and outlier individuals with the WGCNA function *goodSamplesGenes*. To remove any outlier individuals whose expression substantially deviated from the norm, we also filtered individuals with Z-normalized expression greater than 2.5. After evaluating power and network connectivity for each brain region, we selected a soft-thresholding power of eight for network constructions. We constructed networks using *bicor* correlation and set *deepSplit* to two for the caudate nucleus and hippocampus and three for the DLPFC. Additionally, we set *mergeCutHeight* to 0.15 and *minModuleSize* set to 50 for all brain regions and gene networks. We made the co-expression networks using Pearson correlation values with 381, 349, 364 samples and 23,488, 23,039, and 22,990 genes for the caudate nucleus, DLPFC,

and hippocampus, respectively. Significant associations with sex were determined using a linear model and Pearson correlation between binary sex and module eigengenes.

For module preservation analysis⁴⁵ between sex (female versus male), we constructed networks across brain regions for all samples, control-only samples, schizophrenia-only samples using the following parameters: 1) soft-thresholding power of 15 and 2) *nPermutations* set to 100. Similar to our signed network construction, we first filtered genes and outlier individuals with *goodSamplesGenes*. As the DLPFC failed to achieve a scale free topology across networks (i.e., all samples, control-only samples, or schizophrenia-only samples), we also removed outlier individuals after visual inspection of the sample dendrogram generated with *flashClust* (**Fig. S5**). This resulted in the removal of one female neurotypical control (17 years; DLPFC) individual and two individuals with schizophrenia: one female (80 years; DLPFC) and one male (66 years; DLPFC). For each brain region, we used the male-generated networks as the reference group. For the neurotypical control analysis, we generated networks with 240 (71 female; 169 male), 210 (66 female; 146 male), and 243 (74 female; 169 male) individuals for the caudate nucleus, DLPFC, and hippocampus, respectively. For the schizophrenia analysis, we generated networks with 153 (50 female; 103 male), 146 (48 female; 98 male), and 132 (47 female; 85 male) individuals for the caudate nucleus, DLPFC, and hippocampus, respectively. For the combined analysis (control and schizophrenia), we generated networks with 393 (121 female; 272 male), 356 (112 female; 244 male), and 375 (121 female; 254 male) individuals for the caudate nucleus, DLPFC, and hippocampus, respectively. We generated all networks with a total of 26,881, 26,627 and 26,727 genes for the caudate nucleus, DLPFC, and hippocampus, respectively. We considered a module to not be preserved if the Z-summary score was less than or equal to ten.

Random forest dynamic recursive feature elimination

For autosomal sex prediction, we used dRFETools⁴⁶ (version 0.1.17) in Python (version 3.7) to apply dynamic recursive feature elimination with random forest classification⁴⁷. We set the elimination rate to 10% and set 0.30 as the fraction of samples used for lowess smoothing. To reduce overfitting, we generated 10 sex-stratified folds for cross-validation with the *StratifiedKfold* function from scikit-learn⁴⁸. Model performance was measured using normalized mutual information, accuracy, and area under the receiver operating characteristic (ROC) curve with out-of-bag samples.

Multi-marker analysis of genomic annotation (MAGMA) enrichment analysis

For gene set enrichment analysis comparisons of DEGs and multiple GWAS summary statistics, we applied MAGMA⁴⁹ (version 1.10) on GWAS SNP p-values with European reference data downloaded from MAGMA (<https://ctg.cncr.nl/software/magma>). As the GWAS summary statistics were on hg19, we mapped our DEG gene locations from hg38 to hg19 using the GENCODE v41 GRCh37 lifted annotation file. Initially, we generated a SNP annotation file with the annotate flag (*--annotate*). Following this, we performed gene analysis on SNP p-values

using this SNP-level gene annotation file with the following: 1) PLINK input files (--bfile), 2) gene model set to SNP-wise mean, and 3) expanded gene boundaries (35 kb upstream and 10 kb downstream). We used default settings for all other parameters. Once completed, we performed gene-set enrichment analysis in MAGMA (--gene-results) using default parameters. We analyzed sex-specific DEGs by direction of effect (upregulated in females or males) across the caudate nucleus, DLPFC, and hippocampus with 12 traits (seven neuropsychiatric; **Table S1**). We executed this MAGMA pipeline using snakemake⁵⁰ (version 6.4.1).

X-chromosome inactivation (XCI) enrichment analysis

For XCI enrichment analysis, we downloaded the XCI status annotation from Tukiainen *et al.*⁵¹. We accessed the enrichment of sex bias for XCI status using Fisher's exact test with the known XCI categories, including 631 genes defined as escape (n=99), variable escape (n=101), or inactive (n=431). We corrected for multiple testing with the Bonferroni procedure.

Dosage compensation

Relative X expression (RXE) was determined as previously described⁵² with slight modifications; specifically, we used transcripts per million (TPM). We generated TPM using the mean of the read insert size for effective length (**Equation 7**). We extracted the average fragment length as estimated by kallisto per brain region. We dropped any genes with effective lengths less than or equal to one. Following TPM calculation, we performed a log2 transformation (**Equation 8**). Next, we filtered low-expressing genes present in at least 20% of samples. To compute RXE, we calculated the differences in the mean chromosome-wide log2 TPM expression with X-chromosome log2 TPM expression (**Equation 9**).

$$EffectiveLength = Length - [MeanFragmentLength] + 1$$

Equation 7

$$TPM = 1e6 * \frac{Count/EffectiveLength}{\sum (Count/EffectiveLength)}$$

Equation 8

$$RXE = \log_2(\text{mean TPM of X-chromosome genes}) \\ - \log_2(\text{mean TPM of all autosomal genes})$$

Equation 9

Sex-specific differential expression analysis for schizophrenia

To determine more stringent sex-specific differential expression features among the three brain regions using diagnosis subset by sex, we applied additional selection criteria following differential expression analysis. First, we removed any overlapping differentially expressed

features. Following removal, we tested features for significant differences in residualized expression (**Equation 6**) for the opposite sex using Mann-Whitney U and removed significant features (p-value < 0.05).

Subsampling male-only schizophrenia differential expression

For subsampling of the BrainSeq Consortium brain region analysis, we randomly sampled male individuals using the female sample sizes (121, 114, and 121 for the caudate nucleus, DLPFC, and hippocampus, respectively) and performed differential expression analysis (**Equation 6**) for schizophrenia. We performed this 1,000 times.

Functional gene term enrichment analysis

We determined significant enrichment for gene sets using the gene set enrichment analysis (GSEA)^{53,54}. Specifically, we performed GSEA with *gseGO* (gene ontology [GO] gene set database) from the clusterProfiler package⁵⁵ (version 4.6.2) and *gseDGN* (DisGeNET gene set database⁵⁶) from the DOSE package⁵⁷ (version 3.24.2). We defined the gene set “universe” as all unique genes tested for differential expression. For *gseGO*, we set minimal gene set size (*minGSSize*) to 10, maximum gene set size (*maxGSSize*) to 500, and p-value cutoff to 0.05. For *gseDGN*, we set *minGSSize* to five and p-value cutoff to 0.05. We used the default settings for all other parameters.

For gene-term enrichment analysis for WGCNA modules, we used GOATOOLS Python package⁵⁸ (version 1.2.3) with the GO database and hypergeometric tests for enrichment and depletion as previously described¹². Specifically, we converted GENCODE IDs to entrez IDs using pybiomart (<https://github.com/jrderuiter/pybiomart>; version 0.2.0). With entrez IDs, we applied enrichment analysis for each module. We performed multiple testing corrections using the Benjamini-Hochberg FDR method.

To measure GO term elements semantic similarity across brain regions, we used R package *GOSemSim*⁵⁹ (version 2.24.0) with the Wang method⁶⁰ and best-match average strategy.

Sex interacting eQTL analysis in *cis* and region specificity

To identify sex-interacting *cis*-eQTL (si-eQTL) across the caudate nucleus, DLPFC, and hippocampus, we first separated out female and male individuals and, using PLINK 2.0, excluded variants with MAF less than 0.05 and variants with less than one allele by sex. To generate a common list, we overlapped these filtered variants, resulting in a total of 6,816,103 SNPs. Following SNP filtering, we performed eQTL analysis using tensorQTL^{61,62} (version 1.0.7) for a sex-interaction model. We filtered low expression using the GTEx Python script, *eqtl_prepare_expression.py*, with modification for processing transcripts, exons, and junctions. This retained features with expression estimates greater than 0.1 TPM in at least 20% of samples

and aligned read count of six or more per brain region. Following low expression filtering, we performed TMM normalization on filtered counts using the GTEx Python script, *rnaseqnorm.py* (<https://github.com/broadinstitute/gtex-pipeline/tree/master/qtl/src/rnaseqnorm.py>). Following normalization, we implemented tensorQTL using an interaction linear regression model. To do this, we performed three major steps: 1) we adjusted expression for covariates (i.e., diagnosis, population stratification [SNP PCs 1-3], and expression PCs specific to brain region and feature); 2) selected *cis*-SNP using a mapping window of 0.5 Mb within the transcriptional start site (TSS) of each feature; and 3) filtered SNPs based on an interaction MAF greater than or equal to 0.05 and the minor allele present in at least 10 samples.

To assess sharing across brain regions and to increase our power to detect sex-interacting eQTL effects, we used multivariate adaptive shrinkage in R (*mashr*⁶³; version 0.2.57) as previously described¹². *mashr* uses an empirical Bayes approach to learn patterns of similarity among conditions (e.g., brain regions) and then leverage these prior patterns to improve accuracy of effect size estimates. We obtained effect sizes and standard errors for these effect sizes from the tensorQTL interaction model results. To account for correlations among measurements across brain regions (i.e., overlapping sample donors), we used the *estimate_null_correlation_simple* function to specify a correlation structure prior to fitting the *mash* model. The *mash* model included both the canonical covariance matrices and data-driven covariance matrices learned from our data. We defined the data-driven covariance matrices as the top three PCs from the principal components analysis (PCA) performed on the significant signals (i.e., most significant nominal p-values by brain region). To learn the mixture weights and scaling for the si-eQTL effects, we initially fit the *mash* model with a random set (i.e., unbiased representation of the results) of the tensorQTL interaction model results (i.e., 5% for gene-SNP pairs and 1% for transcript-, exon-, and junction-SNP pairs). We next fitted these mixture weights and scaling to all of the si-eQTL results in chunks. We extracted posterior summaries and measures of significance (i.e., local false sign rate [lfsr]). We considered si-eQTL significant if the lfsr < 0.05.

Schizophrenia risk GWAS association

We downloaded the latest schizophrenia GWAS summary statistics with index and high-quality imputation SNPs as determined by Psychiatric Genomics Consortium (PGC version 3 [PGC3])¹¹. Following download, we selected and converted PGC3 GWAS SNPs associated with BrainSeq Consortium SNPs as previously described¹²; specifically, we converted GWAS SNPs from hg19 to hg38 using PyLiftover, merged them with BrainSeq Consortium SNPs on hg38 coordinates, and matched alleles.

Fine mapping and colocalization

We performed fine mapping and colocalization with gene level si-eQTL for the caudate nucleus, DLPFC, and hippocampus as previously described^{12,64} with slight modification for priors.

Briefly, we estimated priors from the tensorQTL nominal results with torus⁶⁵. Following estimation of priors, we implemented DAP-G^{66,67} (version 1.0.0) to generate posterior inclusion probabilities (PIP) that provide an estimate of the probability of a variant being causal for downstream colocalization with fastENLOC^{68,69} (version 1.0). We applied fastENLOC with schizophrenia GWAS (PGC3)¹¹.

We visualized colocalization results using P-P plots and eQTL results from sex-only analysis. Specifically, we used tensorQTL as described above, to apply gene-level *cis*-eQTL analysis to female and male individuals separately without modification for sex interaction. We used a gene body window of 0.5Mb, MAF greater than or equal to 0.01, and confounders generated from the *Sex-interacting eQTL analysis in cis and region specificity*. We determined significance for the most highly associated variant per gene using empirical p-values based on beta-distribution fitted with an adaptive permutation (1,000 to 10,000). These p-values were corrected for multiple testing across genes using Storey's q-value. For each brain region, we generated P-P plots using sex-specific nominal and permutation results (**Data S1**) for each significant colocalized gene identified (regional colocalization probability [RCP] > 0.5).

General replication analysis

Data download

We downloaded differential expression results for sex differences from the supplemental materials for Trabzuni et al., Mayne et al., and Gershoni and Pietrokovski⁷⁰⁻⁷². For sex differences in schizophrenia replication, we downloaded Qin et al. results⁷. For sex-interacting eQTL, we downloaded results from Trabzuni et al., Yao et al., Kukurba et al., and Shen et al.^{17-19,72}.

For CommonMind Consortium replication of differential expression analysis, we downloaded differential expression results for sex differences from Hoffman et al.⁶, as well as normalized expression from Synapse (syn18103849). These results included two cohorts: NIMH HBCC (National Institute of Mental Health's Human Brain Collection Core) and MSSM-Penn-Pitt (MSSM: Mount Sinai NIH Brain Bank and Tissue Repository, Penn: University of Pennsylvania Brain Bank of Psychiatric illnesses and Alzheimer's Disease Core Center, and Pitt: University of Pittsburgh NIH NeuroBioBank Brain and Tissue Repository) cohorts⁷³.

Dosage compensation replication

For dosage compensation replication, we calculated TPM using a mean insert size of 200. We computed the RXE as described above (**Dosage compensation**). We downloaded gene TPM from the GTEx v8 portal (<https://www.gtexportal.org/home/datasets>), as well as sample phenotype information. We computed RXE as described above (**Dosage compensation**).

π_1 replication analysis

For π_1 analysis, we initially selected all significant genes or eQTL (nominally significant, p-value < 0.05) from our results and compared them with results from an external dataset (i.e., CommonMind Consortium [CMC]). Using the p-values from the external dataset, we calculated π_0 with *qvalue* function from *qvalue*⁷⁴ (version 2.30.0). We calculated π_1 with **Equation 10**. The π_1 statistic represents the fraction of effects shared between the two datasets.

$$\pi_1 = 1 - \pi_0 \quad \text{Equation 10}$$

Graphics

We generated venn diagrams with *matplotlib_venn* (version 0.11.5) Python (version 3.8) package. We generated UpSet plots in R using *ComplexHeatmap*⁷⁵ (version 2.6.2). Unless otherwise stated, we generated box plots and scatterplots in R using *ggpubr* (version 0.4.0). We generated enrichment dot plots, enrichment heatmaps, and gene term enrichment plots using *ggplot2*⁷⁶. To generate circos plots, we used *circlize*⁷⁷ (version 0.4.11) and *ComplexHeatmap* utilities in R. We used *plotnine* (version 0.12.1), a Python implementation of *ggplot2*, to generate enrichment heatmaps comparing public datasets with BrainSeq Consortium analysis, RXE scatterplots, and sex-interacting eQTL box plots. To generate rank-rank hypergeometric overlap (RRHO), we used the *RRHO2*^{78,79} (version 1.0) and *lattice* packages in R.

Code availability

All code and Jupyter Notebooks are available through GitHub at https://github.com/LieberInstitute/sex_differences_sz with more detail⁸⁰.

Results

Sex-specific expression across the caudate nucleus, DLPFC, and hippocampus

We first explored sex differences in the brain of the 480 unique individuals (caudate nucleus [n=393], DLPFC [n=359], and hippocampus [n=375]) by performing differential expression of sex after adjusting for diagnosis, age, ancestry (SNP PCs 1-3), RNA quality, and hidden variances (**Equation 1** and **Table 2**) using the BrainSeq Consortium dataset^{12,13}. We observed 831 unique, DEGs (FDR < 0.05; **Fig. 1A**) between the sexes across the caudate nucleus (n=689 DEGs [279 upregulated in females; 410 upregulated in males]), DLPFC (n=256 [99 upregulated in females; 157 upregulated in males]), and hippocampus (n=147 [64 upregulated in females; 83 upregulated in males]). Of these 831 unique DEGs, the sex chromosomes showed the most significant sex-biased expression (**Data S2**). Interestingly, most sex-associated DEGs for the caudate nucleus and DLPFC were autosomal (**Table S2**). When we expanded our analysis to the isoform-level, we identified an additional 859 unique genes associated with a differentially expressed transcript, exon, or exon-exon junction (**Fig. S4**). Furthermore, we observed a similar pattern of majority autosomal genes; however, the most significant differentially expressed features in this context were located on sex chromosomes (**Table S2** and **Data S2**).

To evaluate the function and association of the sexually dimorphic gene expression to heritable complex traits, we performed gene set enrichment analysis and MAGMA enrichment analysis on DEGs separately per brain region. For MAGMA gene set enrichment, we did not observe any significant associations (**Data S3**). This may be partly explained by the underrepresentation of allosomes in GWAS studies⁸¹. With gene set enrichment analysis, we observed significant enrichment (GSEA, q-value < 0.05) of several ontology and disease terms for each brain region (**Fig. 1B** and **Data S4**). As expected, we found enrichment for dosage compensation associated with female-biased (upregulated in females) DEGs across all brain regions. Additionally, we found significant enrichment of fibroid related disease terms (i.e., uterine fibroids, tubulointerstitial fibrosis, and fibroid tumor) associated with female-biased DEGs across all brain regions. The vast majority of enrichment terms (1,477 of 1,515 ontology and disease terms) were associated with female-bias DEGs across brain regions. The small fraction of enrichment terms associated with male-biased DEGs included histone modification, androgen receptor signaling, and autistic behavior (**Data S4**). In contrast, female-biased DEGs were primarily enriched for immune-related pathways (i.e., neutrophil migration, macrophage activation, humoral immune response, complement activation, and astrocyte development) across the brain (**Data S4**). We also found these enrichment patterns replicated in co-expression network modules⁴⁴ associated with sex (**Data S5**). Even so, female- and male-constructed network modules were well preserved ($Z > 10$; **Data S6**), similar to reports by others⁴⁵. Altogether, these results imply that networks shared between female and male individuals are well preserved, while significant sex-specific modules are highly enriched for immune-related pathways.

Autosomes influence sex differences in the brain

With so many autosomal sex-biased DEGs, we asked to what degree sex-biased autosomal genes contribute to sex differences in the brain. We separated allosomal and autosomal DEGs and performed principal component analysis to assess explained variance of these DEGs across the brain. While the first principal component of all allosome DEGs explained approximately 97% variance across brain regions, the autosomal DEGs also showed significant association with sex (**Fig. S6**). Interestingly, we found that explained variance drastically increased with only the 10 most significant autosomal DEGs across brain regions (90%, 87%, and 88% for the caudate nucleus, DLPFC, and hippocampus; **Fig. 1C** and **Fig. S6**). As it appeared that a small subset of these autosomal DEGs explained a large proportion of expression variances between the sexes, we formally tested this using dynamic recursive feature elimination (dRFE)⁴⁶. To this end, we applied random forest classification using 10-fold, sex-stratified, cross-validation with dRFE and found a median of 110 genes (77, 32.5, and 221 for caudate nucleus, DLPFC, and hippocampus respectively) with perfect test score accuracy for sex classification (**Table S3** and **Fig. S7**). Interestingly, the prediction accuracy seemed to be driven by one pseudogene (*RPS10P3* [ribosomal protein S10 pseudogene 3]) shared across brain regions (**Fig. S8**). Additionally, *RPS10P3* – located on chromosome 9 – has previously been reported to be associated with five different traits^{82–86}, including sex-interacting cleft lip⁸⁶ and lateral ventricle temporal horn volume in psychosis⁸⁵. These results indicate that a small subset of autosomal genes significantly contributes to sexually dimorphic gene expression in the brain.

Brain region interaction with sex

To understand the regional specificity of sex DEGs, we compared DEGs from each brain region. We observed a significant enrichment of shared DEGs across the three brain regions (Fisher's exact test, p-value < 0.05; **Fig. 1D**) with the majority (60 DEGs, 62.5%) on sex chromosomes. For replication analysis, we compared the DEGs with previous sex differences analysis in the brain^{6,70–72} and found greater than 62% of DEGs were significantly differentially expressed in all brain regions except for the GTEx cerebellum and anterior cingulate cortex (**Fig. S9**) with a concordant direction of effect between BrainSeq Consortium and GTEx brain regions (Fisher's exact test, p-value < 0.01). For a more in-depth comparison, we examined the sex differences found using the CMC DLPFC⁶. We also discovered a large number of DEGs on sex chromosomes (39 of 51 [76.5%] and 41 of 54 [75.9%] for the NIMH HBCC and MSSM-Penn-Pitt cohorts, respectively) similar to our BrainSeq Consortium analysis. Additionally, we observed significant pairwise enrichment of these CMC DEGs with our BrainSeq Consortium DEGs across brain regions (Fisher's exact test, p-value < 0.01; **Fig. S10**). Altogether, this suggests that X- and Y-linked genes drive brain-wide sex expression differences and autosomal genes drive brain region-specific differences. Additionally, autosomal DEGs were less likely to replicate in different datasets.

Interestingly, we found that all genes, regardless of significant association with sex, showed a significant positive correlation for the direction of effect between pairwise comparisons of the three brain regions (Spearman, $\rho > 0.69$, $p\text{-value} < 1.4\text{e-}108$; **Fig. 1E**). At significant levels (DEGs, adjusted $p\text{-value} < 0.05$), these pairwise correlations dramatically increased (Spearman, $\rho > 0.99$, $p\text{-value} < 4\text{e-}104$; **Fig. S11A**) with significant concordant direction of effect (RRHO; **Fig. S12**). Expanded analysis of transcripts, exons, and exon-exon junctions displayed a similar pattern with all shared, differentially expressed (DE) features (adjusted $p\text{-value} < 0.05$) having significant concordant direction and significant positive correlation among brain regions (Spearman, $\rho > 0.97$, $p\text{-value} \cong 0$; **Fig. S11B-D**). Moreover, at significant levels (adjusted $p\text{-value} < 0.05$), all directions agreed between the CMC DLPFC and the BrainSeq Consortium brain regions with a significant positive correlation (Spearman; $\rho > 0.97$ for all pairwise comparisons; $p\text{-value} < 1.1\text{e-}44$; **Fig. S13**). In summary, the direction of change for sexually dimorphic genes is generally shared across multiple brain regions and independent datasets.

We next evaluated the degree of sex bias among brain regions formally with an interaction model for sex and brain region. Here, we found extensive interactions (**Fig. 1F**), particularly in the caudate nucleus as compared to the DLPFC (adjusted $p\text{-value} < 0.05$, DEGs=528) and the caudate nucleus as compared to the hippocampus (adjusted $p\text{-value} < 0.05$, DEGs=71). In contrast, we found only five genes (*ZNF736P9Y*, *ENSG00000285756*, *TUBBP1*, *TBL1Y*, and *ENSG00000285679*) with region-specific expression for sex between the DLPFC and hippocampus. When we expanded our analysis to the isoform level, we identified more than double unique DEGs (1303 [775 isoform only], 198 [127 isoform only], and 23 [18 isoform only], for caudate nucleus vs DLPFC, caudate nucleus vs hippocampus, and DLPFC vs hippocampus, respectively; **Fig. S14**, **Table S4** and **Data S7**).

To understand the functional significance of these brain region-specific transcriptional changes, we applied GSEA and MAGMA enrichment for each pairwise comparison. While we did not find enrichment for any brain or non-brain traits (MAGMA; **Data S3**), we did observe significant enrichment (GSEA, $q\text{-value} < 0.05$) of several ontology terms for all pairwise comparisons (**Fig. S15** and **Data S8**). Interestingly, we observed terms associated with myelination (i.e., myelination, axon ensheathment, and ensheathment of neurons) for comparisons between the DLPFC and the other two brain regions. For brain region-specific transcriptional sex differences between the caudate nucleus and hippocampus, we also observed enrichment for cognition, neurotransmission, and regulation of synaptic plasticity. For transcriptional sex differences between the caudate nucleus and DLPFC, we observed additional terms associated with gene silencing. For transcriptional sex differences between the DLPFC and hippocampus, we observed additional terms associated with receptor signaling. Altogether, this analysis highlights the importance of brain region-specific transcriptional sex differences, which are significantly enriched for neurotransmission and myelination.

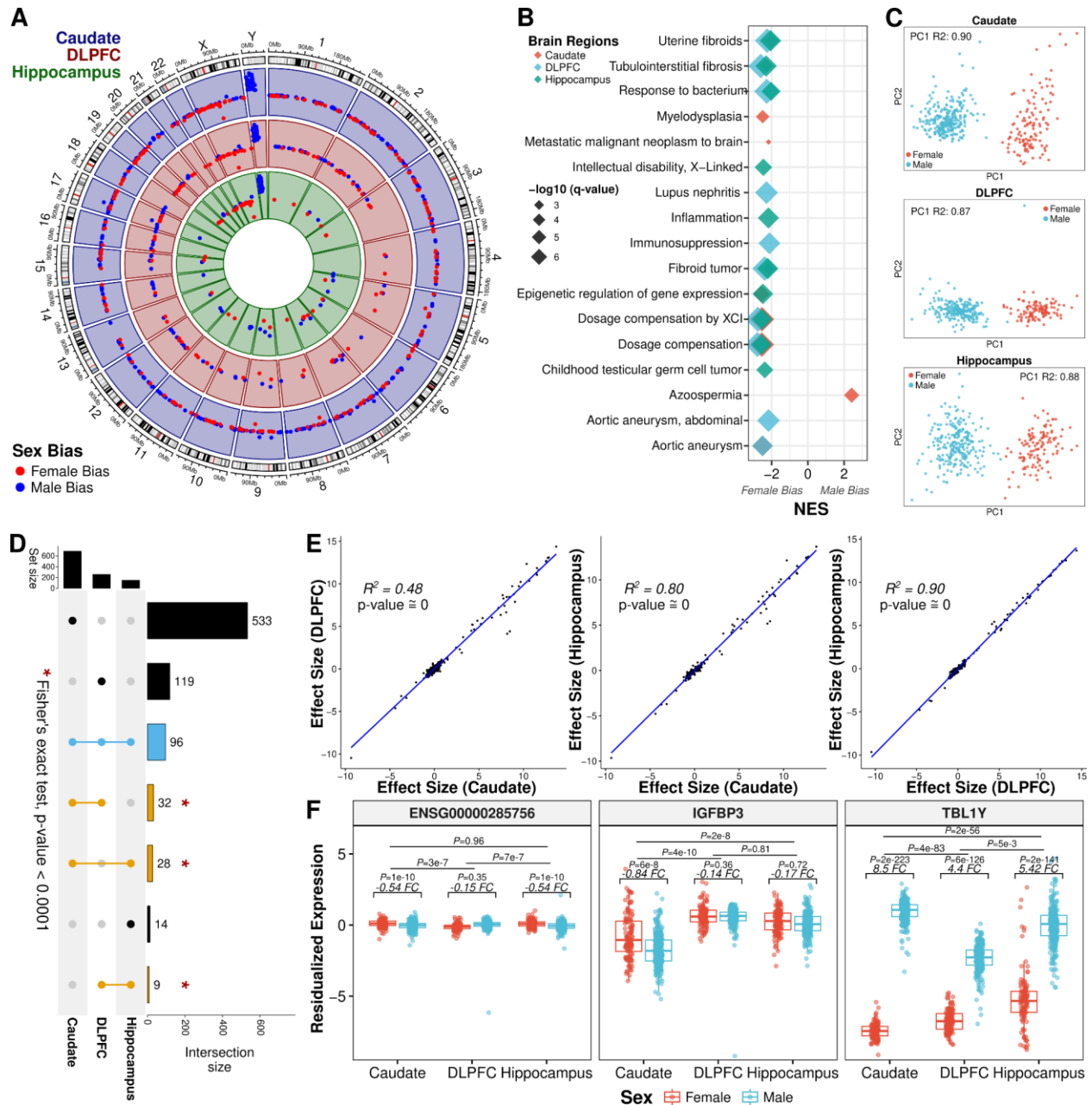


Fig. 1. Sex-biased expression across the caudate nucleus, DLPFC, and hippocampus. A. Circos plot showing significant differentially expressed genes (DEGs) for the caudate nucleus (blue; n=393; 121 female and 272 male), DLPFC (red; n=359; 114 female and 245 male), and hippocampus (green; n=375; 121 female and 254 male) across all chromosomes. Female bias (upregulated in female individuals) in red, and male bias (upregulated in male individuals) in blue. **B.** Gene set enrichment analysis (GSEA) of sex differential expression analysis across brain regions, highlighting the top ten most significant terms upregulated in females (female bias) or males (male bias). NES: normalized enrichment score. XCI: X-chromosome inactivation. **C.** Scatterplots of the estimated proportion of expression variance explained by sex within the 100 most significant autosomal DEGs (i.e., adjusted p-value) for the caudate nucleus

(DEGs, n=60), DLPFC (DEGs, n=25), and hippocampus (DEGs, n=42). **D.** UpSet plot showing overlap of DEGs across the caudate nucleus, DLPFC, and hippocampus. Blue is shared across the caudate nucleus, DLPFC, and hippocampus; orange, shared between two brain regions; and black, unique to a specific brain region. * Indicating p-value < 0.0001 for two-sided, Fisher's exact test. **E.** Scatterplot of effect size (logFC) for all genes tested showing concordant positive directionality with significant two-sided, Spearman correlation (R^2) of all genes. A fitted trend line is presented in blue as the mean values \pm standard deviation. **F.** Example box plots of genes showing an interaction between sex and brain region. FC= fold change log2 (male / female). Female individuals in red and male individuals in blue. Adjusted p-value (P) annotation using dream⁴² (default of Satterthwaite approximation) generated statistics annotation. Box plots show the median and first and third quartiles, and whiskers extend to 1.5 \times the interquartile range.

XCI and dosage compensation in the brain

As 70% of the brain-wide, sexually dimorphic genes are located on sex chromosomes, we next evaluated the dosage of X-linked genes compared to autosomes. In order to equalize the dosage of X-linked genes between XX females and XY males, female mammals epigenetically silence one X chromosome in a process called XCI (X-chromosome inactivation). XCI is a chromosome-wide process where the majority of X-linked genes are nearly completely silenced, and a minority of X-linked genes either escape X inactivation or show variable X inactivation. When we examined the DEGs by brain region for X-linked gene dosage, we found the majority of DEGs were enriched for XCI escape genes (**Fig. S16**, Fisher's exact test, Bonferroni < 0.01; **Fig. 2A**), reflecting dosage compensation for the majority of X-linked genes subject to XCI as seen in previous studies^{51,87}.

For all three brain regions, we found XCI escape genes were significantly enriched within the female-biased DEGs, whereas male-biased DEGs, surprisingly, were also significantly enriched for genes escaping XCI across brain regions (Fisher's exact test, Bonferroni < 0.01; **Fig. 2A**). These male-biased escaping XCI genes were all located on the PAR (pseudoautosomal regions) of both X and Y chromosomes (*AKAP17A*, *ASMTL*, *ASMTL-AS1*, *CD99*, *CD99P1*, *DHRXS*, *GTPBP6*, *IL3RA*, *LINC00106*, *PLCXD1*, *PPP2R3B*, and *ZBED1*; **Data S9**). Additionally, we found the most male-biased XCI-annotated genes in the caudate nucleus (n = 16, 9, and 10 DEGs for caudate nucleus, DLPFC, and hippocampus, respectively), which were mostly annotated as escape XCI genes (**Data S9**). In contrast, we only found enrichment of variable XCI genes in the caudate nucleus (Fisher's exact test, Bonferroni < 0.01). Altogether, XCI escaping genes demonstrated higher expression in female individuals across brain regions, suggesting sex differences shared across brain regions are associated with well-documented XCI escaping genes for females.

Next, we evaluated differences in chromosome-wide dosage by comparing the relative X chromosome expression (RXE) to autosomes (**Fig. 2B** and **Fig. S17**). Interestingly, we observed a significant decrease of RXE in male individuals only in the DLPFC (Mann-Whitney U, p-value

= 0.047), demonstrating region-specific dosage compensation. We also observed a similar trend of decreased RXE in the DLPFC from the CMC MSSM-Penn-Pitt cohort (Mann-Whitney U, p-value = 0.07; **Fig. S18A**) but not the GTEx frontal cortex (**Fig. S19**). Even so, the large RXE variation across the 13 GTEx brain regions demonstrated region-specific dosage compensation (**Fig. S19**).

As we found differences in the DLPFC between sexes, we next asked if this might be due to individuals with schizophrenia. Interestingly, we found decreased RXE in the hippocampus of male patients (Mann-Whitney U, p-value = 0.004; **Fig. 2C**) but not in the caudate nucleus, DLPFC, or CMC DLPFC (**Fig. S18B**). However, there was no significant interaction between sex and diagnosis status for any brain region for RXE. These results demonstrate slight differences between X-chromosome dosage in the hippocampus of individuals with schizophrenia.

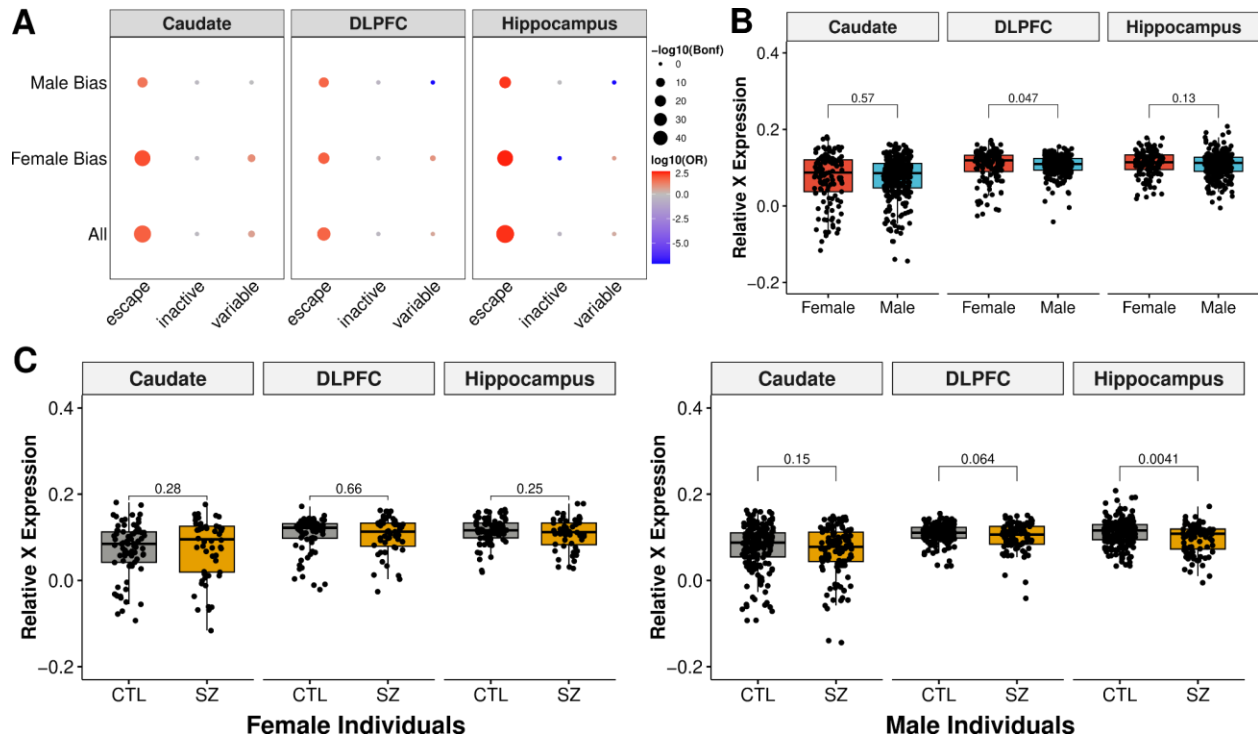


Fig. 2. X-linked gene expression and dosage compensation observed across the caudate nucleus, DLPFC, and hippocampus. **A.** Enrichment of significant sex-biased genes relative to genes known to escape X-chromosome inactivation (XCI) across the caudate nucleus (n=393; 121 female and 272 male), DLPFC (n=359; 114 female and 245 male), and hippocampus (n=375; 121 female and 254 male). Dot plot of enrichment (two-sided, Fisher's exact test) of differentially expressed genes (DEGs) for XCI genes by brain region separated by male bias (upregulated in male individuals), female bias (upregulated in female individuals), and all DEGs. Size of dots denotes $-\log_{10}$ of Bonferroni corrected p-values. Color relates to \log_{10} odds ratio (OR) with depletion in blue and enrichment in red. **B.** Box plot showing relative X-chromosome

expression (RXE) comparison between female (red) and male (blue) individuals with two-sided, Mann-Whitney U two-tailed, p-values annotated. **C.** Box plot showing significant differences between neurotypical control (CTL; gray) and schizophrenic (SZ; gold) individuals in the caudate nucleus, DLPFC, and hippocampus for female (left) and male (right) individuals with annotation of two-sided, Mann-Whitney U p-values. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range.

Interaction of schizophrenia and sex in the brain

After investigating sex differences in the brain without consideration of diagnosis in 480 unique individuals (caudate nucleus [n=393], DLPFC [n=359], and hippocampus [n=375]), we next identified statistically significant differentially expressed features (adjusted p-value < 0.05) with respect to sex differences and diagnosis through an interaction model. No genes, transcripts, or exons were significant by this interaction model, similar to a previous study⁶. While overall replication with CMC DLPFC was limited, we found significant correlation of nominally significant (p-value < 0.05) transcriptional signatures between DLPFC and CMC DLPFC, NIMH HBCC cohort ($\pi_1 = 0.51$; Spearman, $\rho = 0.60$, p-value < 0.01; **Fig. S20** and **Table S5**). In contrast, on the junction level, 148 junctions demonstrated a significant (adjusted p-value < 0.05; **Fig. S21A** and **Data S10**) interaction between sex and diagnosis across the caudate nucleus (nine novel junctions; **Fig. S21B**), DLPFC (89 novel junctions; **Fig. S21B**), and hippocampus (47 novel junctions and three junctions associated with *DDX11L1* [DEAD/H-Box Helicase 11 Like 1]; **Fig. S21B**).

We also examined differential expression for schizophrenic female and male individuals separately across the caudate nucleus, DLPFC, and hippocampus using RRHO analysis to increase our power of detecting transcriptional changes. Here, we found schizophrenia-related transcriptional signatures, while concordant direction of effect varied dramatically depending on sex and brain region. Specifically, female schizophrenia transcriptional signatures showed the strongest pattern of sharing between the caudate nucleus and the DLPFC (**Fig. 3A**), while males showed the strongest pattern of sharing between DLPFC and hippocampus (**Fig. 3B**). These patterns are similar to those found in ours and others' previous schizophrenia analyses adjusted for sex^{12,13}. Altogether, sex-adjusted schizophrenia analysis largely reflects male transcriptional changes likely due to larger male sample sizes.

Next, we examined female and male transcriptional changes for schizophrenia within individual brain regions. Here, we found the strongest shared signature within the caudate nucleus with very little observable overlap for the DLPFC and hippocampus (**Fig. 3C**). Furthermore, we examined DEGs (adjusted p-value < 0.05) and found a large overlap for the caudate nucleus but little overlap between females and males for the DLPFC and hippocampus (**Fig. 3D**). For the DLPFC and hippocampus, the limited observable overlap is apparent in the opposing enrichment patterns between DLPFC and hippocampus in females (**Fig. 3A**) and males (**Fig. 3B**); while female transcriptional signatures associated with genes upregulated in schizophrenia, male

transcriptional signatures associated with genes downregulated. Similar trends occurred for transcripts, exons, and junctions (**Data S11** and **Fig. S22**).

Sex-specific schizophrenia expression in the brain

To further examine female- and male-specific schizophrenia DEGs, we applied a more stringent filter to exclude: 1) genes shared between female-only and male-only schizophrenia analyses and 2) genes detected as significantly different (Mann-Whitney U, p -value < 0.05) with residualized expression for opposite sex-only individuals. For female-specific schizophrenia analysis, the single DLPFC DEG and 169 DEGs from the caudate nucleus were filtered out because they showed significant residualized expression for male individuals with schizophrenia (Mann-Whitney U, p -value < 0.05) resulting in a total of 194 female-specific schizophrenia DEGs in the caudate nucleus (**Table S6** and **Data S11**). For male-specific schizophrenia analysis, we found 1130, 5, and 149 DEGs for the caudate nucleus, DLPFC, and hippocampus, respectively. Due to the small number of DLPFC-identified DEGs, we found no DEGs shared across brain regions for male-specific schizophrenia analysis (**Table S7**). We found two genes (*EDN3* and *PLD4*) shared between the caudate nucleus and DLPFC and 23 genes (**Table S7**) shared between the caudate nucleus and hippocampus.

We hypothesized that the smaller sample size for female individuals might explain why we identified few if any female-specific schizophrenia DEGs within the BrainSeq Consortium dataset. To test this hypothesis, we performed 1,000 random samplings of the male individuals at female sample sizes ($n = 121$, $n = 114$, and $n = 121$, for the caudate nucleus, DLPFC, and hippocampus, respectively) and calculated DEGs for each brain region and permutation. On average, we identified a drastically smaller number of schizophrenia DEGs (median male-only schizophrenia DEGs of 347, 0, and 1 for the caudate nucleus, DLPFC, and hippocampus, respectively; **Fig. S26**) in our subsampled male samples that showed no significant difference (permutation p -value = 0.30, 0.80, and 0.69 for the caudate nucleus, DLPFC, and hippocampus, respectively) between the number of schizophrenia DEGs identified from the female-only analysis. Altogether, the smaller female sample size, at least partially, explains the lack of identification of female-specific schizophrenia DEGs within the BrainSeq Consortium datasets.

Given the role of estrogen in the treatment of schizophrenia⁸⁸, we next examined sex hormone expression in the context of schizophrenia in the brain. Here, we did not observe any sex-specific or sex-specific schizophrenia DEGs for sex hormone expression (*AR* [androgen receptor], *ESR1* [estrogen receptor 1], *ESR2* [estrogen receptor 2], and *PGR* [progesterone receptor]) across the brain. As sex hormone levels vary with age, we next examined any potential interaction between sex hormones and age. For female and male individuals, we found no significant interactions of diagnosis status and age for the caudate nucleus and DLPFC. For the hippocampus, we found a nominally significant upregulation of *ESR2* (linear regression, p -value = 0.016 [FDR = 0.098]; **Fig. S23**) in female individuals with schizophrenia compared with neurotypical controls and *PGR* in male individuals with schizophrenia compared with neurotypical controls as a function

of age (linear regression, p -value = 0.013 [FDR = 0.16]; **Fig. S24**). These results suggest that sex hormones expression may potentially interact with age and diagnosis status in the hippocampus.

We next examined functional and MAGMA enrichment of these sex-specific, schizophrenia transcriptional changes. For MAGMA gene set enrichment, we found significant enrichment of schizophrenia and neutrophils – a white blood cell type – traits for male-specific DEGs upregulated in individuals with schizophrenia for the caudate nucleus and hippocampus (**Data S3**). We also found significant enrichment of the schizophrenia trait for female-specific DEGs upregulated in individuals with schizophrenia for the caudate nucleus (**Data S3**). In contrast, we found significant enrichment of basophil and eosinophils for male-specific DEGs downregulated in individuals with schizophrenia for the DLPFC (**Data S3**). With these immune-related cell type enrichment, we were not surprised to observed significant depletion (GSEA, q -value < 0.05) of immune-related pathways for both female- and male-only schizophrenia analysis (**Fig. S25A,B** and **Data S12**). Interestingly, we found the hippocampus showed the greatest degree of similarity between female- and male-only schizophrenia functional enrichment analysis (best-match average > 64%; **Fig. S25C**). In contrast, the caudate nucleus showed the lowest degree of similarity between sex-specific schizophrenia analysis with biological processes and molecule function showing 42% and 38% similarity in GO terms, respectively. This seemingly contradictory analysis highlights the potential impact of the larger number of sex-specific schizophrenia DEGs identified in the caudate nucleus (**Fig. 3D**). In addition to this functional analysis, we also investigated gene co-expression networks between female and male individuals by diagnosis. While we observed complete preservation of modules for neurotypical control individuals across brain regions, we found one module significantly not preserved ($Z < 10$) for the DLPFC (94 genes [46% protein coding]; **Data S6**). While we did not find any GO terms significantly enriched, this module included terms related to DNA binding (i.e., *HDHC3*, *MAZ*, *PCBP1*, *TOP3B*, and zinc finger proteins) and transposable and repetitive elements (i.e., *LRRC24* and *TIGD7*).

We next compared our results with the recent meta-analysis for sex-specific schizophrenia DEGs in the prefrontal cortex⁷. Of the 46 male-specific DEGs identified by Qin et al., we found a total of three overlapping genes: one gene overlapping (*PARD3*) with the caudate nucleus stringent female-specific DEGs and two overlapping genes (*USE1* and *ABCG2*) with the hippocampus stringent male-specific DEGs, which all shared direction of effect. When we compared the full set of female and male schizophrenia DEGs across brain regions, we found an additional three overlapping genes (*CD99*, *GABARAPL1*, and *LIN7B*) shared with the caudate nucleus. Of these three only *GABARAPL1* had a discordant direction of effect.

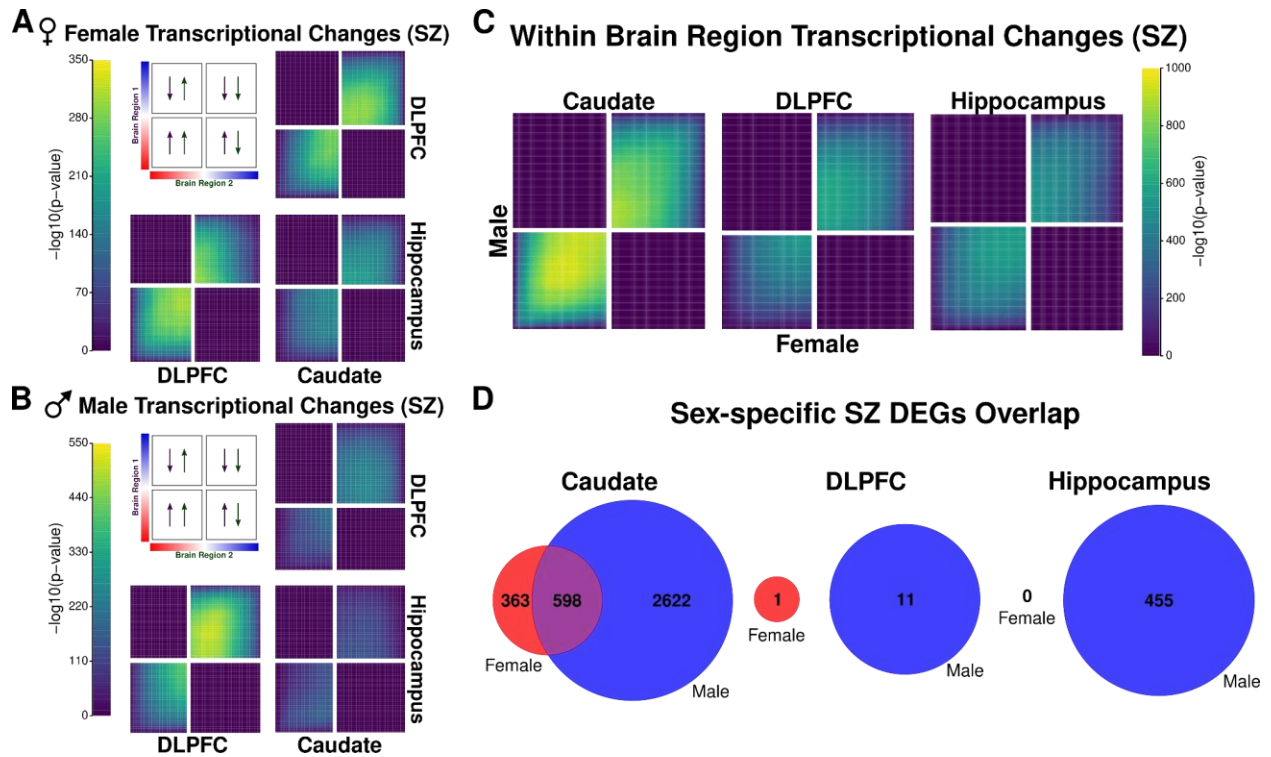


Fig. 3. Transcriptional changes for schizophrenia shared between sexes within brain regions. RRHO (rank-rank hypergeometric overlap) maps comparing schizophrenia transcriptional changes for all genes between brain region pairs stratified by direction of effect in **A.** females and **B.** males. The panel presents the overlapping relationship between two brain regions. The color bars represent the degree of significance [$-\log_{10}(\text{p-value})$] of overlap between two brain regions. Arrows show the direction of effect for schizophrenia (upregulated or downregulated in schizophrenia) by brain region. **C.** RRHO map comparing female and male schizophrenia transcriptional changes within brain regions for all genes. The color bar represents the degree of significance [$-\log_{10}(\text{p-value})$] of the overlap between the sexes. **D.** Venn diagram showing overlap within brain regions for sex-stratified schizophrenia differentially expressed genes (DEGs; female in red and male in blue; $\text{FDR} < 0.05$). Female-specific schizophrenia DEGs in red, male-specific schizophrenia DEGs in blue, and schizophrenia DEGs shared between female- and male-specific schizophrenia analyses in purple. SZ: schizophrenia.

Sex-dependent eQTL in the brain

We asked whether genetic regulation of expressed features would manifest differently in females compared to males for the caudate nucleus, DLPFC, and hippocampus. We tested for statistical interaction between genotype and sex in the brain by applying multivariate adaptive shrinkage (mash) modeling in the 504 individuals (age > 13) for the caudate nucleus ($n=399$), DLPFC ($n=377$), and hippocampus ($n=394$). We identified hundreds of sex-interacting variants (si-eQTL) across brain regions for gene-, transcript-, exon-, and junction-level analysis (**Table 3** and **Data S13**). For example, we found 703, 545, and 546 gene-level si-eQTL (local false sign rate

[lfsr] < 0.05) for the caudate nucleus, DLPFC, and hippocampus, respectively, accounting for 704 unique genes with si-eQTL (eGenes; **Table 3** and **Data S13**) driven by the caudate nucleus. Only 21 (3.0%) of these eGenes (si-eQTL associated with unique genes) were located on the X chromosome; the majority of eGenes were located on autosomes, similar to sex-specific expression analysis. We found this proportion aligns with the ratio of autosomes to allosomes and shows no significant shift in distribution between eGenes and genes tested (Kolmogorov-Smirnov test, p-value=0.89).

To understand the regional specificity of these si-eQTL, we examined the proportion of si-eQTL detected across brain regions. Here, we found the majority (544 [77%]) of eGenes were shared across brain regions (**Fig. 4A**), which was also observed on the isoform level (i.e., transcripts, exons, and junctions; **Fig. S27**). Remarkably, all of the shared si-eQTL showed concordant directionality. Furthermore, the DLPFC and hippocampus showed nearly identical si-eQTL effect sizes (**Fig. 4B**), which was confirmed with the high level of replication across brain regions ($\pi_1 > 0.996$; **Fig. S28**). The few brain region-specific si-eQTL showed small but significant sexual dimorphic genetic regulation of expression (**Fig. S29**). Unsurprisingly, the exon- and junction-level sharing showed a smaller proportion of shared si-eQTL across brain regions at an effect size within a factor of 0.99 (**Fig. S30**), suggesting alternative isoform usage drives differences in si-eQTL effect size across brain regions.

To evaluate the functional relevance underlying si-eQTL in the caudate nucleus, DLPFC, and hippocampus, we performed functional gene term enrichment analysis on the eGenes for each brain region. We observed significant enrichment (hypergeometric, adjusted p-value < 0.05) across brain regions (**Data S14**), including enrichment for neurogenesis and cellular localization (**Fig. 4C**). Notably, we found that these enriched GO terms showed high semantic similarity⁶⁰ (best-match average, 62-100%) across brain regions (**Fig. S31**). While we did not find these si-eQTL associated eGenes significant enrichment for sex-specific DEGs (Fisher's exact test, p-value > 0.10), we did find significant enriched (Fisher's exact test, FDR < 0.05) for neurological disorders including schizophrenia^{12-14,89}, autism spectrum disorder⁸⁹, and Alzheimer's disease⁹⁰ (**Fig. S32**).

When we compared our si-eQTL with previous work in whole blood and in lymphoblastoid cell lines, we found no overlap with the 19 si-eQTL identified in whole blood^{17,18} and two genes (*ATG4C* and *CA2*) of the 21 si-eQTL identified in lymphoblastoid cell lines¹⁹ and also present in the caudate nucleus si-eQTL. We next compared our results with the four si-eQTL (q-value < 0.25) identified in GTEx brain regions (amygdala and nucleus accumbens basal ganglia)⁹¹ and found no overlaps. When we expanded to the 369 si-eQTL (q-value < 0.25) from all 43 GTEx tissues⁹¹, we found two overlapping genes encoding noncoding RNAs (*ENSG00000270605* and *ENSG00000272977*) between the caudate nucleus and suprapubic skin and spleen GTEx tissues, respectively. Relatively low replication rate with GTEx brain regions can, in part, be attributed to low sample sizes in the GTEx dataset⁹².

We next set out to determine if any of these si-eQTL had causal associations with schizophrenia risk (PGC3 GWAS p-value $< 5e-8$)¹¹. To formally identify variants associated with schizophrenia risk, we performed colocalization analysis on fine-mapped gene level si-eQTL across brain regions (**Data S16**). We identified nine unique genes across the three brain regions with significant colocalization (regional colocalization probability [RCP] > 0.5 ; **Data S17**): one genes (*ENSG00000287222*; **Fig. S33**) in the caudate nucleus; seven genes in the DLPFC (*ACE*, *ELAC2*, *FURIN*, *MMD*, *ZSCAN29*, *LINC00320*, and *ENSG00000289128*; **Fig. S34**); and five genes in the hippocampus (*ACE*, *ELAC2*, *MMD*, *STRC*, and *ENSG00000287222*; **Fig. S35**). Interestingly, the brain regions with the largest detection of schizophrenia risk genes were the DLPFC and hippocampus; three of the nine colocalized genes were shared between these two brain regions. Additionally, the colocalized gene identified in the caudate nucleus (*ENSG00000287222*) was also significant in the hippocampus. Furthermore, we also identified this shared gene as a sex-specific schizophrenia DEG (male-specific, adjusted p-value = 0.026) in the caudate nucleus. While we did not observe any tissue-specific overlap between the other eight colocalized genes and sex-specific schizophrenia DEGs – potentially due to low number of identified DEGs – we did observe one additional colocalized gene (*FURIN*; male-specific, adjusted p-value = 0.023) as a sex-specific schizophrenia DEG in the caudate nucleus. Interestingly, we found *ACE* to be downregulated in schizophrenia for both sexes in the caudate nucleus. Altogether, this correlates with the high level of sharing of si-eQTL across brain regions, suggesting that while sex-interacting colocalized genes are highly recurrent across the brain, sex-specific dysregulation of these si-eQTL may be brain-region specific.

Table 3. Summary of sex-interacting eQTL (lfsr < 0.05) across brain regions for genes, transcripts, exons, and exon-exon junctions associated with all si-eQTL. eQTL: number of variant-feature pairs for each feature type: genes, transcripts, exons, and exon-exon junctions. eFeature: number of unique features that have eQTL associations. eGene: number of eQTL associations with unique genes. lfsr: local false sign rate⁹³.

Brain Regions		Caudate Nucleus	DLPFC	Hippocampus
Gene	eQTL	3,274	2,464	2,465
	eFeature	703	545	546
	eGenes	703	545	546
Transcript	eQTL	10,186	8,032	8,207
	eFeature	1,700	1,383	1,394
	eGenes	1,577	1,286	1,296
Exon	eQTL	10,439	7,873	7,824
	eFeature	1,737	1,426	1,420
	eGenes	801	665	662
Junction	eQTL	1,740	1,150	1,145
	eFeature	328	228	228
	eGenes	7	5	5

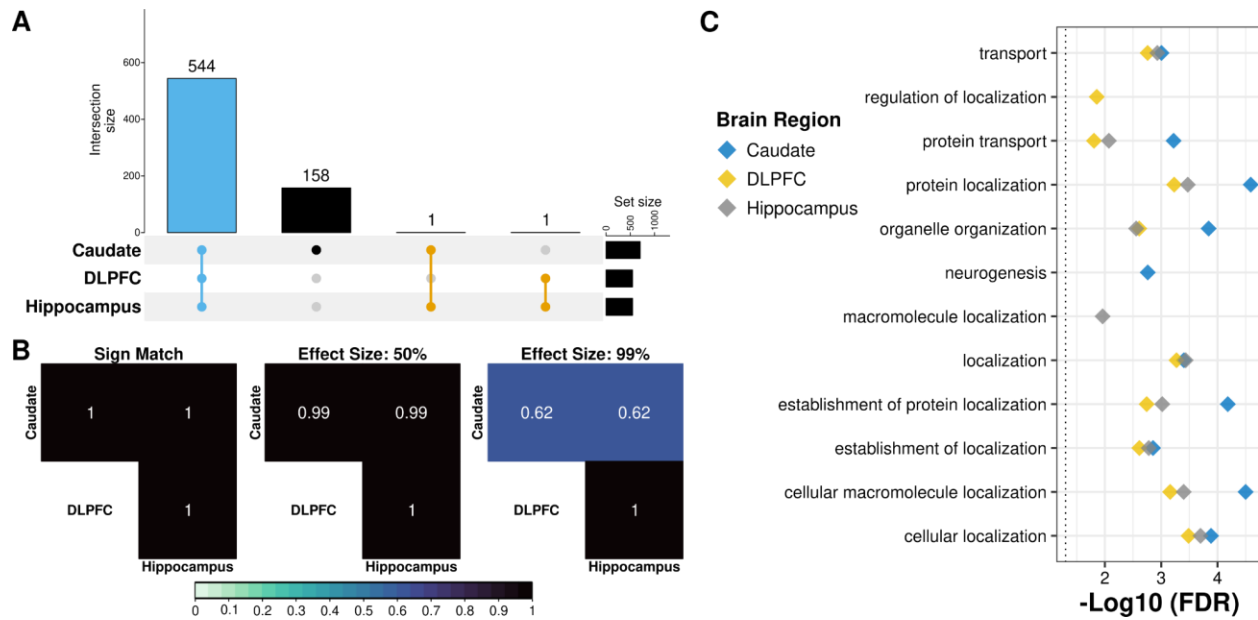


Fig. 4. Sex-interacting eQTL (si-eQTL) are shared across brain regions. **A.** UpSet plot displaying overlap across brain regions for si-eQTL (local false sign rate [lfsr] < 0.05). Blue is shared across the caudate nucleus (n=399; 126 female and 273 male), DLPFC (n=377; 121 female and 256 male), and hippocampus (n=394; 126 female and 268 male); orange, shared between two brain regions; and black, unique to a specific brain region. **B.** Heatmap of the proportion of gene level si-eQTL sharing with sign match (left), within a factor of 0.5 effect size (middle), and within a factor 0.99 effect size (right). **C.** Functional enrichment plot of the ten most significant gene ontology-terms (biological processes) of eGenes for the caudate nucleus (blue), DLPFC (yellow), and hippocampus (gray).

Discussion

Sex has been associated with differential gene expression in the brain and sex-specific effects in neuropsychiatric disorders like schizophrenia. Here, we aimed to take a holistic exploratory analysis approach to sex differences for schizophrenia in the caudate nucleus, DLPFC, and hippocampus. We identified numerous genetic features (genes, transcripts, exons, and exon-exon junctions) that 1) are associated with sex, 2) demonstrate sex-specific expression in schizophrenia, and 3) have expression that interacts with genotypes and sex as si-eQTL. Furthermore, we identified nine genes showing sex-variable association with schizophrenia risk¹¹ (PGC3 GWAS p-value < 5e-8) with colocalization and some overlap with sex-specific schizophrenia DEGs. Additionally, we have demonstrated slight differences between X-chromosome dosage in the hippocampus of male individuals with schizophrenia – and further shown that the degree of dosage compensation varies across the brain. To the best of our knowledge, this is the largest multi-brain region analysis for sex differences in schizophrenia.

In this study, we found 831 unique genes with sex-associated differential expression in the caudate nucleus, DLPFC, and hippocampus. While genes on sex chromosomes have the largest sex-specific effects, we also found that autosomal genes significantly influenced sex differences. Our results support previous findings of sex differences, including differences that are brain-region specific⁷⁰, are primarily located on autosomes^{6,71,72}, and have exhibited shared direction of effect driven by autosomal DEGs⁷⁰. A larger number of samples in our cohort allowed us to identify many more genes than previous analyses conducted across multiple regions, contributing to our knowledge on sex-specific genomic features and sex differences in the brain.

To determine if sex differences observed across brain regions were related to XCI, we further examined DEGs located on the X chromosome. Here, our analysis aligned with previous work showing an enrichment of genes known to escape XCI^{51,87}. Additionally, we found that across the brain (i.e., GTEx and BrainSeq Consortium) X-chromosome dosage showed brain region-specific dosage compensation levels.

The second aim of the study was to identify any sex-specific schizophrenia genomic features across multiple brain regions. As one of the most consistently implicated brain regions for the pathophysiology of schizophrenia, the DLPFC has been the primary focus of postmortem analysis for schizophrenia^{94,95}. However, we have recently shown the importance of analyzing multiple brain regions, specifically for the identification of potentially new therapeutic targets¹². As such, we aimed to identify 1) which brain region showed the most sex-specific schizophrenia-associated transcriptional changes and 2) the degree of sharing of these schizophrenia-associated transcriptional changes across brain regions.

While our interaction model found only junctions (148 exon-exon junctions) significantly dysregulated with a sex-specific interaction across brain regions, comparing female to male expression in schizophrenia allowed us to identify many more sex-specific schizophrenia genomic features than previous studies – even after applying an additional, more stringent filter to our sex-specific schizophrenia DEG results. This was unsurprising as we have previously found the caudate nucleus has substantially more schizophrenia DEGs compared to the DLPFC and hippocampus¹². While we expect the majority of the sex-specific schizophrenia DEGs in the caudate nucleus to be associated with antipsychotic treatment^{12,96}, our results also provide a starting point to examine differential antipsychotic effects by sex for schizophrenia.

As the first study to compare multiple brain regions for sex-specific schizophrenia analysis, we also found that sex-specific schizophrenia differentially expressed features were highly brain-region specific. This also was not unexpected as schizophrenia DEGs, irrespective of sex, are also highly brain-region specific^{12,13,15}. However, the smaller number of identified DE features for transcripts, exons, and junctions was surprising. This might be due to our study being underpowered, as we found twice as many schizophrenia differentially expressed features in male individuals as compared to females.

Given the role of estrogen in the treatment of schizophrenia, we also examined sex hormone-related expression (*AR*, *ESR1*, *ESR2*, and *PGR*) in the brain. Using expression data, we were unable to identify any significant associations after correcting for multiple testing. We found only a nominally significant interaction with diagnosis and age of *ESR2* in female individuals and *PGR* in male individuals in the hippocampus. This may be due to: 1) conventional methods for hormone measurements using serum, or 2) limited sample size for interaction models. Altogether, these findings suggest a need for increased samples from female individuals to further advance our understanding of potential sex differences for schizophrenia across multiple brain regions.

The third aim of our study was to identify si-eQTL across the brain. We annotated hundreds of si-eQTL associated with 704 eGenes across the caudate nucleus, DLPFC, and hippocampus using mash modeling⁹³. With mash modeling, we were able to increase our power to detect si-eQTL. With this study, we have provided the first annotation of si-eQTL in the DLPFC and hippocampus and found that these sex-interacting eGenes were enriched for neurogenesis as well as DEGs from neurological studies (i.e., schizophrenia, autism spectrum disorder, and Alzheimer's). Our results demonstrate the power of tissue-specific sex-interacting eQTL and its potential for identifying genes with sexually dimorphic expression for neurological disorder.

In addition to annotating hundreds of si-eQTL, we have also provided the first annotation of sex-interacting genes with causal variants associated with schizophrenia risk. One limitation of this analysis is our small female sample size. Even so, these results highlight the importance of examining genes associated with schizophrenia risk for potential differences in expression for population covariates (e.g., sex, age, and genetic ancestry).

In summary, we have provided a comprehensive genetic and transcriptional analysis of sex differences in schizophrenia. We have increased the number of annotated features exhibiting sex bias in the brain adding to our current understanding of sex differences in the brain, identified sex-specific schizophrenia genes with indications for new therapeutic targets, and provided the first annotation of si-eQTL for the DLPFC and hippocampus. These results have the potential to direct new therapeutics and new strategies that can address sex-biased responses in the treatment of schizophrenia. Additionally, these results highlight the need for more female samples for schizophrenia analyses.

Acknowledgements

The authors would like to extend their appreciation to the Offices of the Chief Medical Examiner of Washington DC, Northern Virginia, Kalamazoo Michigan, Santa Clara County, University of North Dakota, and Maryland for the provision of brain tissue used in this work. The authors also extend their posthumous appreciation to Dr. Llewellyn B. Bigelow and members of the LIBD Neuropathology Section for their work in assembling and curating the clinical and demographic information and organizing the Human Brain Tissue Repository of the Lieber Institute. Finally,

the authors gratefully acknowledge the families that have donated this tissue to advance our understanding of psychiatric disorders. This work is supported by the Lieber Institute for Brain Development, the National Institutes of Health (NIH) T32 fellowship (T32MH015330) and K99 award (K99MD016964) to KJMB, NIH R01 (R01MH123183) to LC-T, and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation to JAE. RA would like to thank the Hopkins Office for Undergraduate Research (HOUR), Johns Hopkins University for their support through the Summer PURA program as well as the Albstein Research Scholarship for their support.

Author Contributions

Conceptualization, KJMB, RA, ACMP and JAE; Methodology, KJMB, RA, GP, ASF, HG, JMS, LD, ACMP, and JAE; Software, KJMB, RA, GP, ASF, HG, WSU, and ACMP; Formal Analysis, KJMB, RA, GP, ASF, and HG; Investigation, JHS and TMH; Data Curation, KJMB, RA, GP, and JEK; Writing – Original Draft, KJMB, RA, and JAE; Writing – Review & Editing, KJMB, RA, GP, ASF, HG, LC-T, ACMP, DRW, and JAE; Visualization, KJMB, RA, LD, and WSU; Supervision, KJMB, ACMP, and JAE; Project Administration, JAE; Funding Acquisition, KJMB, RA, LC-T, DRW, and JAE.

These authors contributed equally: KJMB and RA.

Conflict of interest

DRW serves on the Scientific Advisory Boards of Sage Therapeutics and Pasithea Therapeutics. All other authors declare no competing interests.

Data availability

Publicly available BrainSeq Consortium total RNA DLPFC and hippocampus RangedSummarizedExperiment R Objects with processed counts are available at <http://eqtl.brainseq.org/phase2/>. Publicly available BrainSeq Consortium total RNA caudate nucleus RangedSummarizedExperiment R Objects with processed counts are available at http://erwinpaquolalab.libd.org/caudate_eqtl/. Analysis-ready genotype data will be shared with researchers that obtain dbGaP access. FASTQ files are available for total RNA DLPFC and hippocampus via [Globus](#) collections [jhpce#bsp2-dlpfc](#) and [jhpce#bsp2-hippo](#). For the caudate nucleus, FASTQ files are available via dbGaP. PGC3 GWAS summary statistics are available at <https://figshare.com/articles/dataset/scz2022/19426775>. Supplementary data files are also available at <https://doi.org/10.5281/zenodo.8415382> or http://erwinpaquolalab.libd.org/3region_sex/.

Code availability

All code and Jupyter Notebooks are available through GitHub at https://github.com/LieberInstitute/sex_differences_sz with more detail⁸⁰.

References

1. Goldstein, J. M. *et al.* Are there sex differences in neuropsychological functions among patients with schizophrenia? *Am. J. Psychiatry* **155**, 1358–1364 (1998).
2. Eranti, S. V., MacCabe, J. H., Bundy, H. & Murray, R. M. Gender difference in age at onset of schizophrenia: a meta-analysis. *Psychol. Med.* **43**, 155–167 (2013).
3. Faraone, S. V., Chen, W. J., Goldstein, J. M. & Tsuang, M. T. Gender differences in age at onset of schizophrenia. *Br. J. Psychiatry* **164**, 625–629 (1994).
4. Khashan, A. S. *et al.* Higher risk of offspring schizophrenia following antenatal maternal exposure to severe adverse life events. *Arch. Gen. Psychiatry* **65**, 146–152 (2008).
5. Migeon, B. R. X-linked diseases: susceptible females. *Genet. Med.* **22**, 1156–1174 (2020).
6. Hoffman, G. E. *et al.* Sex differences in the human brain transcriptome of cases with schizophrenia. *Biol. Psychiatry* **91**, 92–101 (2022).
7. Qin, W., Liu, C., Sodhi, M. & Lu, H. Meta-analysis of sex differences in gene expression in schizophrenia. *BMC Syst. Biol.* **10 Suppl 1**, 9 (2016).
8. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, (2020).
9. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
10. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
11. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
12. Benjamin, K. J. M. *et al.* Analysis of the caudate nucleus transcriptome in individuals with

- schizophrenia highlights effects of antipsychotics and new risk genes. *Nat. Neurosci.* **25**, 1559–1568 (2022).
13. Collado-Torres, L. *et al.* Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. *Neuron* **103**, 203–216.e8 (2019).
 14. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
 15. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
 16. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
 17. Kukurba, K. R. *et al.* Impact of the X Chromosome and sex on regulatory variation. *Genome Res.* **26**, 768–777 (2016).
 18. Yao, C. *et al.* Sex- and age-interacting eQTLs in human complex diseases. *Hum. Mol. Genet.* **23**, 1947–1956 (2014).
 19. Shen, J. J., Wang, Y.-F. & Yang, W. Sex-Interacting mRNA- and miRNA-eQTLs and Their Implications in Gene Expression Regulation and Disease. *Front. Genet.* **10**, 313 (2019).
 20. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
 21. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

22. Graubert, A., Aguet, F., Ravi, A., Ardlie, K. G. & Getz, G. RNA-SeQC 2: efficient RNA-seq quality control and quantification for large cohorts. *Bioinformatics* **37**, 3048–3050 (2021).
23. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
24. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
25. Cotto, K. C. *et al.* Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat. Commun.* **14**, 1589 (2023).
26. Morgan, M., Obenchain, V., Hester, J. & Pagès, H. SummarizedExperiment: SummarizedExperiment container. (2022).
27. Eagles, N. J. *et al.* SPEAQeasy: a scalable pipeline for expression analysis and quantification for R/bioconductor-powered RNA-seq analyses. *BMC Bioinformatics* **22**, 224 (2021).
28. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
29. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
30. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
31. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
32. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium

- panel. *Nat. Genet.* **48**, 1443–1448 (2016).
33. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
 34. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
 35. Purcell, S. & Chang, C. *PLINK*. (2021).
 36. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
 37. Leek, J. T. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* **42**, e161 (2014).
 38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
 39. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
 40. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
 41. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
 42. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
 43. Smyth, G. K. Linear models and empirical bayes methods for assessing differential

- expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
44. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
 45. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
 46. Benjamin, K. J. M., Katipalli, T. & Paquola, A. C. M. dRFEtools: Dynamic recursive feature elimination for omics. *Bioinformatics* (2023) doi:10.1093/bioinformatics/btad513.
 47. Breiman, L. Random Forests. *Springer Science and Business Media LLC* (2001) doi:10.1023/a:1010933404324.
 48. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
 49. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
 50. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
 51. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
 52. Jue, N. K. *et al.* Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach. *BMC Genomics* **14**, 150 (2013).
 53. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
 54. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).

55. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
56. Piñero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).
57. Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).
58. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
59. Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978 (2010).
60. Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. & Chen, C.-F. A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23**, 1274–1281 (2007).
61. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
62. Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228 (2019).
63. Uebachs, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
64. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021).
65. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian

- false discovery rate control. *Ann. Appl. Stat.* **10**, 1619–1638 (2016).
66. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. *Am. J. Hum. Genet.* **98**, 1114–1129 (2016).
 67. Lee, Y., Francesca, L., Pique-Regi, R. & Wen, X. Bayesian Multi-SNP Genetic Association Analysis: Control of FDR and Use of Summary Statistics. *BioRxiv* (2018) doi:10.1101/316471.
 68. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13**, e1006646 (2017).
 69. Pividori, M. *et al.* PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci. Adv.* **6**, (2020).
 70. Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol.* **15**, 7 (2017).
 71. Mayne, B. T. *et al.* Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. *Front. Genet.* **7**, 183 (2016).
 72. Trabzuni, D. *et al.* Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771 (2013).
 73. Hoffman, G. E. *et al.* CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data* **6**, 180 (2019).
 74. Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. (2020).
 75. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in

- multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
76. Wickham, H. *ggplot2 - Elegant Graphics for Data Analysis*. (Springer International Publishing, 2016). doi:10.1007/978-3-319-24277-4.
 77. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
 78. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).
 79. Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W. & Seney, M. L. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Sci. Rep.* **8**, 9588 (2018).
 80. Benjamin, K. J. LieberInstitute/sex_differences_sz: Updates for revision. *Zenodo* (2023) doi:10.5281/zenodo.8410992.
 81. Sun, L., Wang, Z., Lu, T., Manolio, T. A. & Paterson, A. D. eXclusionarY: 10 years later, where are the sex chromosomes in GWASs? *Am. J. Hum. Genet.* **110**, 903–912 (2023).
 82. Vernet, R. *et al.* Identification of novel genes influencing eosinophil-specific protein levels in asthma families. *J. Allergy Clin. Immunol.* **150**, 1168–1177 (2022).
 83. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
 84. A Gadd, D. *et al.* The genetic and epigenetic profile of serum S100 β in the Lothian Birth Cohort 1936 and its relationship to Alzheimer’s disease. *Wellcome Open Res.* **6**, 306 (2021).
 85. Alliey-Rodriguez, N. *et al.* NRXN1 is associated with enlargement of the temporal horns

- of the lateral ventricles in psychosis. *Transl. Psychiatry* **9**, 230 (2019).
86. Carlson, J. C. *et al.* Genome-wide interaction studies identify sex-specific risk alleles for nonsyndromic orofacial clefts. *Genet. Epidemiol.* **42**, 664–672 (2018).
 87. Lopes-Ramos, C. M. *et al.* Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep.* **31**, 107795 (2020).
 88. Brand, B. A., de Boer, J. N. & Sommer, I. E. C. Estrogens in schizophrenia: progress, current challenges and opportunities. *Curr. Opin. Psychiatry* **34**, 228–237 (2021).
 89. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
 90. Marques-Coelho, D. *et al.* Differential transcript usage unravels gene expression alterations in Alzheimer’s disease human brains. *npj Aging Mech. Dis.* **7**, 2 (2021).
 91. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* (2020).
 92. Leon, A. C. & Heo, M. Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model. *Comput. Stat. Data Anal.* **53**, 603–608 (2009).
 93. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).
 94. Goldman-Rakic, P. S., Carroll, B. J. & Barrett, J. E. Psychopathology and the brain. *New York: Raven* (1991).
 95. Selemon, L. D. Regionally diverse cortical pathology in schizophrenia: clues to the etiology of the disease. *Schizophr. Bull.* **27**, 349–377 (2001).
 96. Perzel Mandell, K. A. *et al.* Molecular phenotypes associated with antipsychotic drugs in the human caudate nucleus. *Mol. Psychiatry* **27**, 2061–2067 (2022).

97. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
98. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
99. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
100. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
101. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
102. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
103. Astle, W. J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016).

Figure legends

Fig. 1. Sex-biased expression across the caudate nucleus, DLPFC, and hippocampus. **A.** Circos plot showing significant differentially expressed genes (DEGs) for the caudate nucleus (blue; n=393; 121 female and 272 male), DLPFC (red; n=359; 114 female and 245 male), and hippocampus (green; n=375; 121 female and 254 male) across all chromosomes. Female bias (upregulated in female individuals) in red, and male bias (upregulated in male individuals) in blue. **B.** Gene set enrichment analysis (GSEA) of sex differential expression analysis across brain regions, highlighting the top ten most significant terms upregulated in females (female bias) or males (male bias). NES: normalized enrichment score. XCI: X-chromosome inactivation. **C.** Scatterplots of the estimated proportion of expression variance explained by sex within the 100 most significant autosomal DEGs (i.e., adjusted p-value) for the caudate nucleus (DEGs, n=60), DLPFC (DEGs, n=25), and hippocampus (DEGs, n=42). **D.** UpSet plot showing overlap of DEGs across the caudate nucleus, DLPFC, and hippocampus. Blue is shared across the caudate nucleus, DLPFC, and hippocampus; orange, shared between two brain regions; and black, unique to a specific brain region. * Indicating p-value < 0.0001 for two-sided, Fisher's exact test. **E.** Scatterplots of effect size (logFC) for all genes tested showing concordant positive directionality with significant two-sided, Spearman correlation (R^2) of all genes. **F.** Example box plots of genes showing an interaction between sex and brain region. FC= fold change log2 (male / female). Female individuals in red and male individuals in blue. Adjusted P-value (P) annotation using dream⁴² (default of Satterthwaite approximation) generated statistics annotation. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range.

Fig. 2. X-linked gene expression and dosage compensation observed across the caudate nucleus, DLPFC, and hippocampus. **A.** Enrichment of significant sex-biased genes relative to genes known to escape X-chromosome inactivation (XCI) across the caudate nucleus (n=393; 121 female and 272 male), DLPFC (n=359; 114 female and 245 male), and hippocampus (n=375; 121 female and 254 male). Dot plot of enrichment (two-sided, Fisher's exact test, two-tailed) of differentially expressed genes (DEGs) for XCI genes by brain region separated by male bias (upregulated in male individuals), female bias (upregulated in female individuals), and all DEGs. Size of dots denotes -log10 of Bonferroni corrected p-values. Color relates to log10 odds ratio (OR) with depletion in blue and enrichment in red. **B.** Box plot showing relative X-chromosome expression (RXE) comparison between female (red) and male (blue) individuals with two-sided, Mann-Whitney U p-values annotated. **C.** Box plot showing significant differences between neurotypical controls (CTL; gray) and individuals with schizophrenia (SZ; gold) in the caudate nucleus, DLPFC, and hippocampus for female (left) and male (right) individuals with annotation of two-sided, Mann-Whitney U p-values. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range.

Fig. 3. Transcriptional changes for schizophrenia shared between sexes within brain regions. RRHO (rank-rank hypergeometric overlap) maps comparing schizophrenia transcriptional changes for all genes between brain region pairs in **A.** females and **B.** males. The panel presents the overlap relationship between two brain regions. The color bars represent the degree of significance [$-\log_{10}(\text{p-value})$] of overlap between two brain regions. Arrows show the direction of effect for schizophrenia (up- or down-regulated in schizophrenia) by brain region. **C.** RRHO map comparing female and male schizophrenia transcriptional changes within brain regions for all genes. The color bar represents the degree of significance [$-\log_{10}(\text{p-value})$] of the overlap between the sexes. **D.** Venn diagram showing overlap within brain regions for sex-stratified schizophrenia DEGs (female in red and male in blue; $\text{FDR} < 0.05$). Female-specific schizophrenia DEGs in red, male-specific schizophrenia DEGs in blue, and schizophrenia DEGs shared between female- and male-specific schizophrenia analyses in purple. SZ: schizophrenia.

Fig. 4. Sex-interacting eQTL are shared across brain regions. **A.** UpSet plot displaying overlap across brain regions for si-eQTL ($\text{lfsr} < 0.05$). Blue is shared across the caudate nucleus ($n=399$; 126 female and 273 male), DLPFC ($n=377$; 121 female and 256 male), and hippocampus ($n=394$; 126 female and 268 male); orange, shared between two brain regions; and black, unique to a specific brain region. **B.** Heatmap of the proportion of gene level si-eQTL sharing with sign match (left), within a factor of 0.5 effect size (middle), and within a factor 0.99 effect size (right). **C.** Functional enrichment plot of the ten most significant GO-terms (biological processes) of eGenes for the caudate nucleus (blue), DLPFC (yellow), and hippocampus (gray).

Supplementary information

Figures

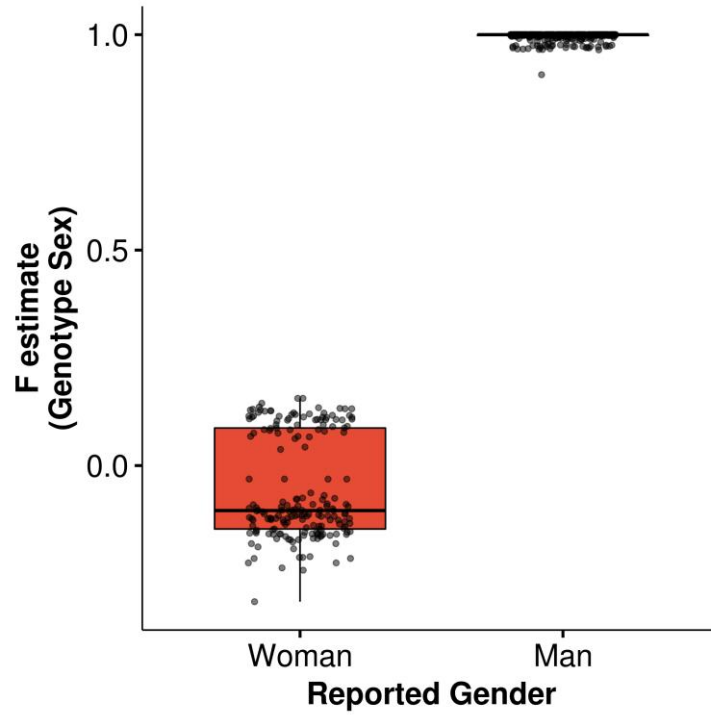


Fig. S1. Completed overlap of reported gender with sex genotype. Box plot showing F estimates of genotype sex for female and male individuals correlate with reported gender (i.e., woman or man). Box plots show the median and first and third quartiles, and whiskers extend to $1.5\times$ the interquartile range.

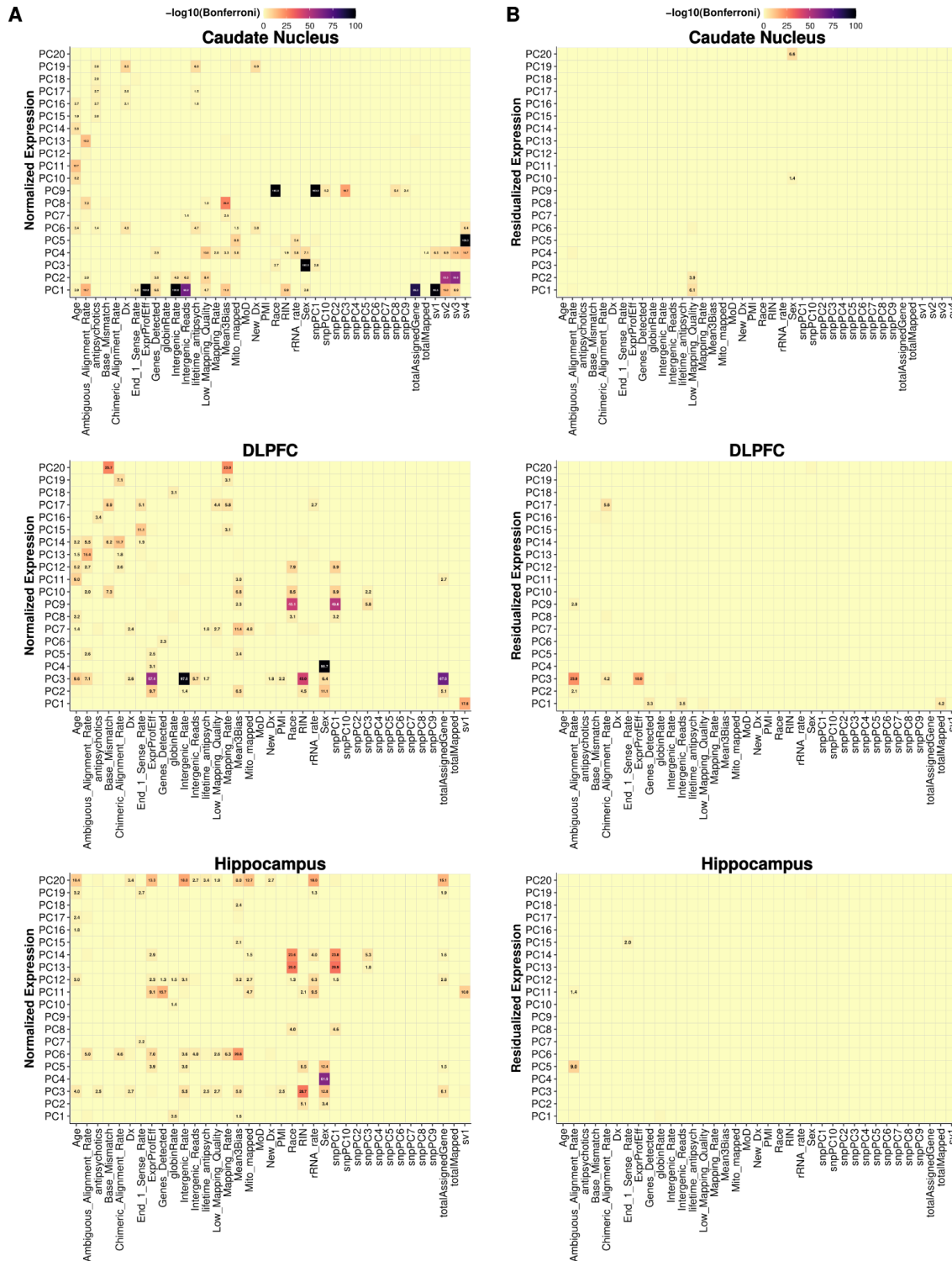


Fig. S3. Spurious correlations reduced after regressing out general model covariates. Heatmaps showing correlation between principal component analysis of **A.** normalized gene expression or **B.** residualized gene expression. A value of 1.3 or greater is significant and equivalent to Bonferroni corrected p-value < 0.05. Significant correlations are denoted within each tile.

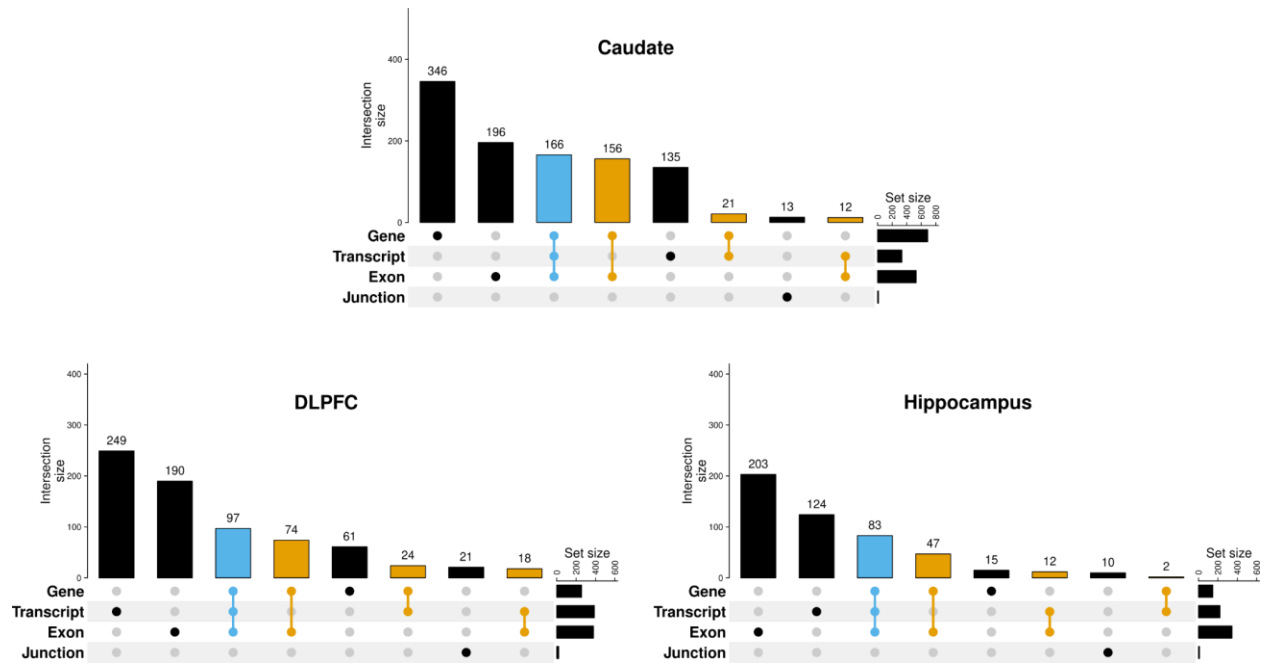


Fig. S4. Isoform level analysis increases the detection of sex-biased genes. Blue, shared across three brain regions; orange, shared between two brain regions; and black, unique to a specific brain region. Novel junctions not annotated to unique gene ID.

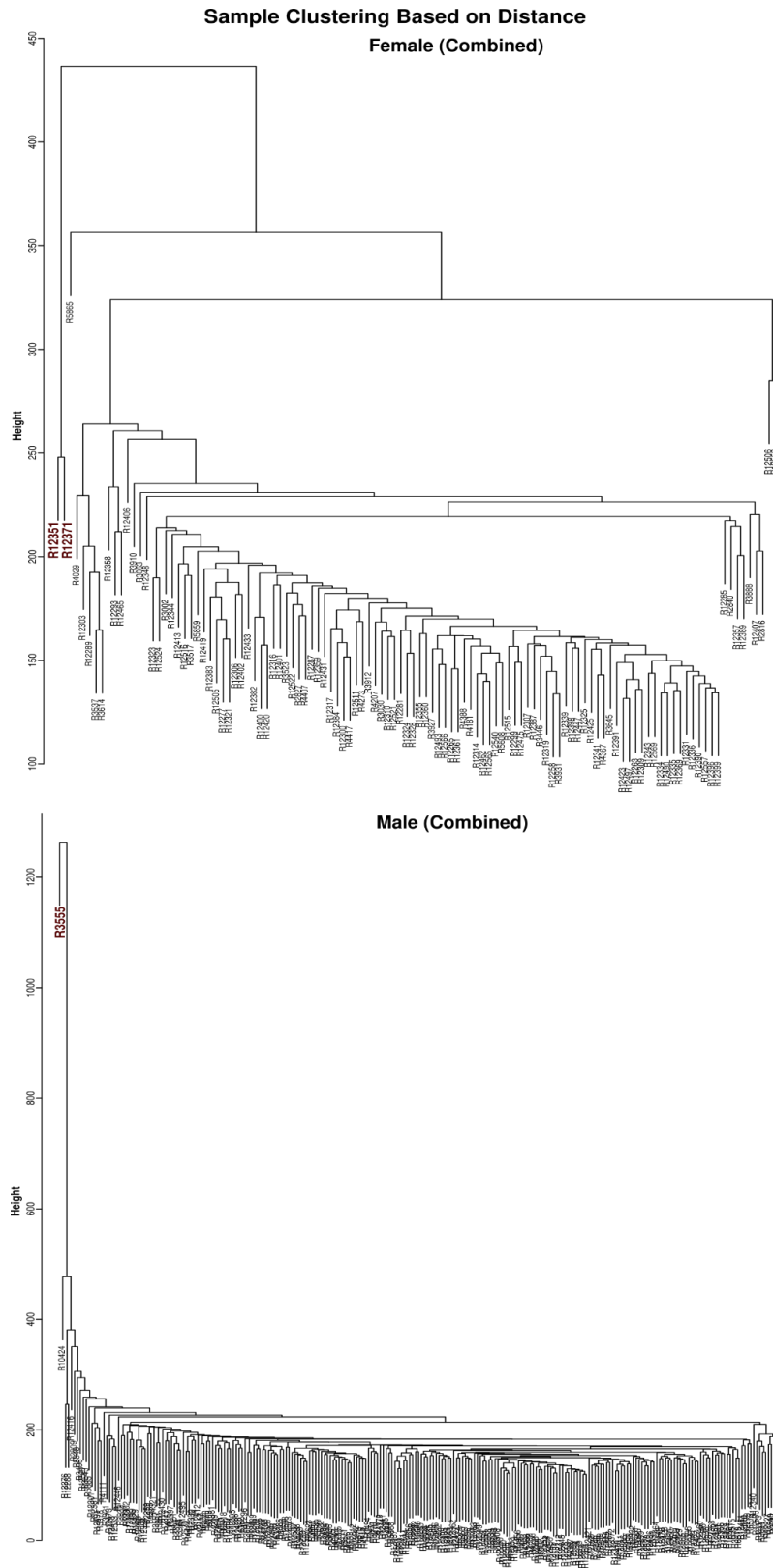


Fig. S5. Sample dendrogram showing outlier individuals for the DLPFC. Removed samples highlighted in red.

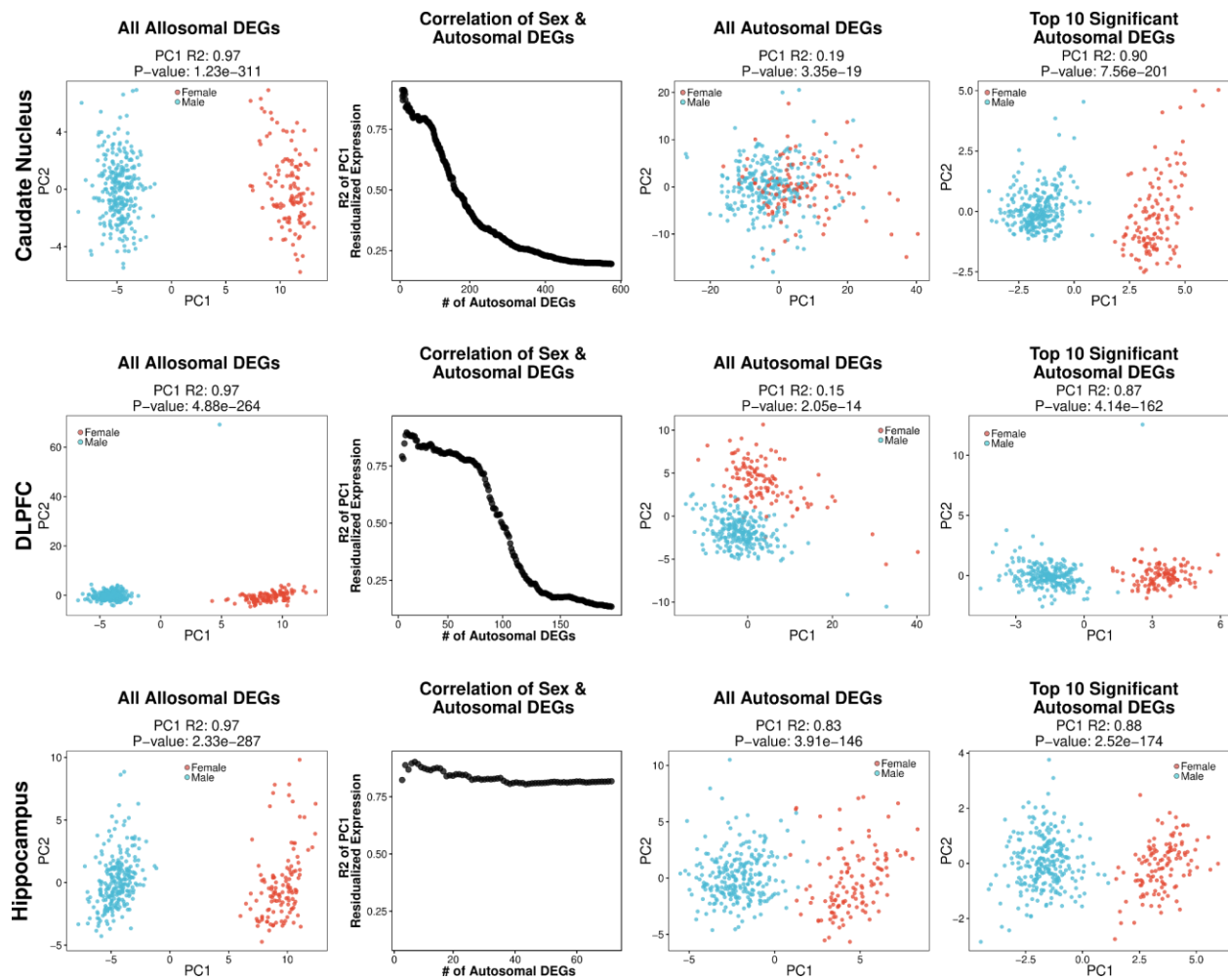


Fig. S6. Autosomal sex-specific DEGs show significant correlation with sex and strong predictive power for sex in the brain. Scatterplots of principal components (PC) 1 and 2 from dimensionally reduced expression of all allosomal and autosomal DEGs for the caudate nucleus, DLPFC, and hippocampus.

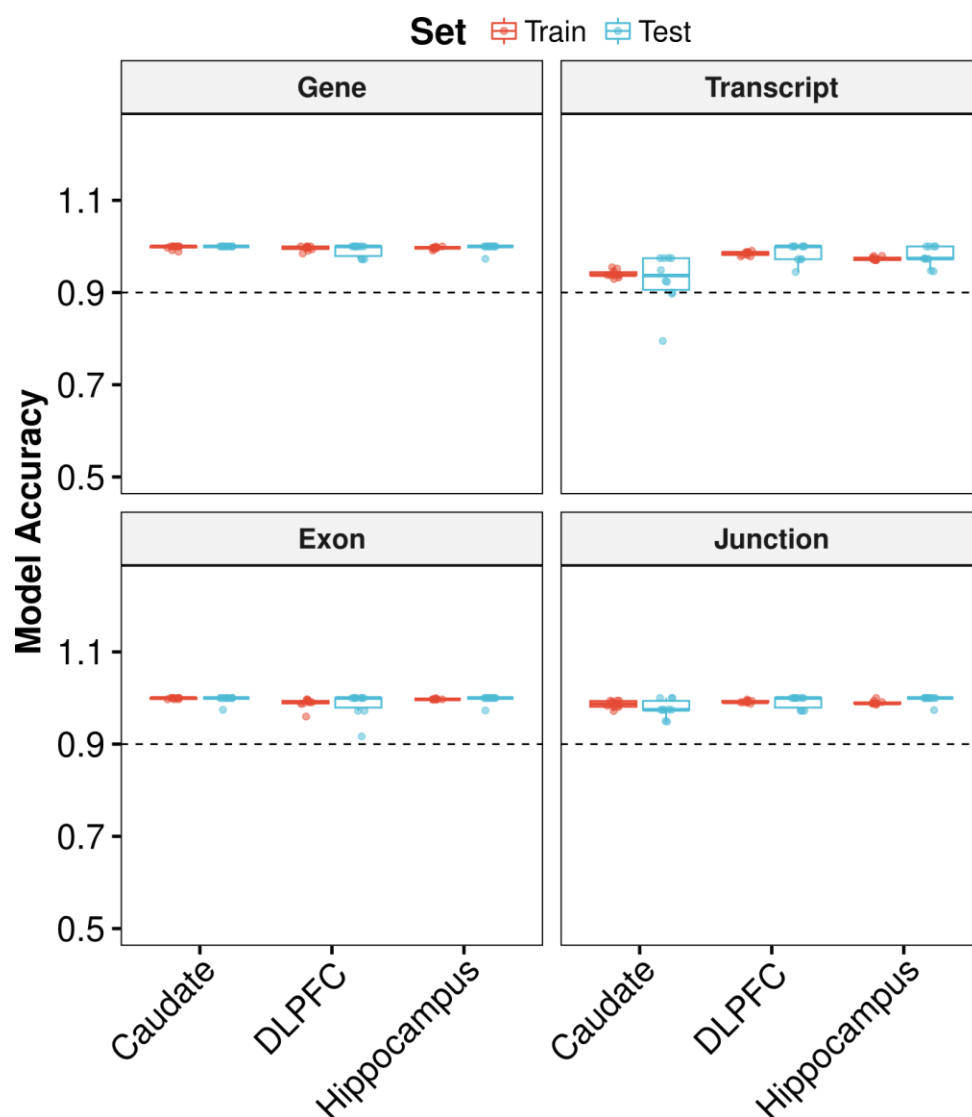


Fig. S7. Metric summary across brain regions and features show high classification accuracy for sex using autosomes. Box plot of train (red) and test (blue) accuracy for sex classification in the caudate nucleus, DLPFC, and hippocampus for genes, transcripts, exons, and junctions. Each point represents results from one fold within the 10-fold cross-validation. Dashed line denotes 90% accuracy. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range.

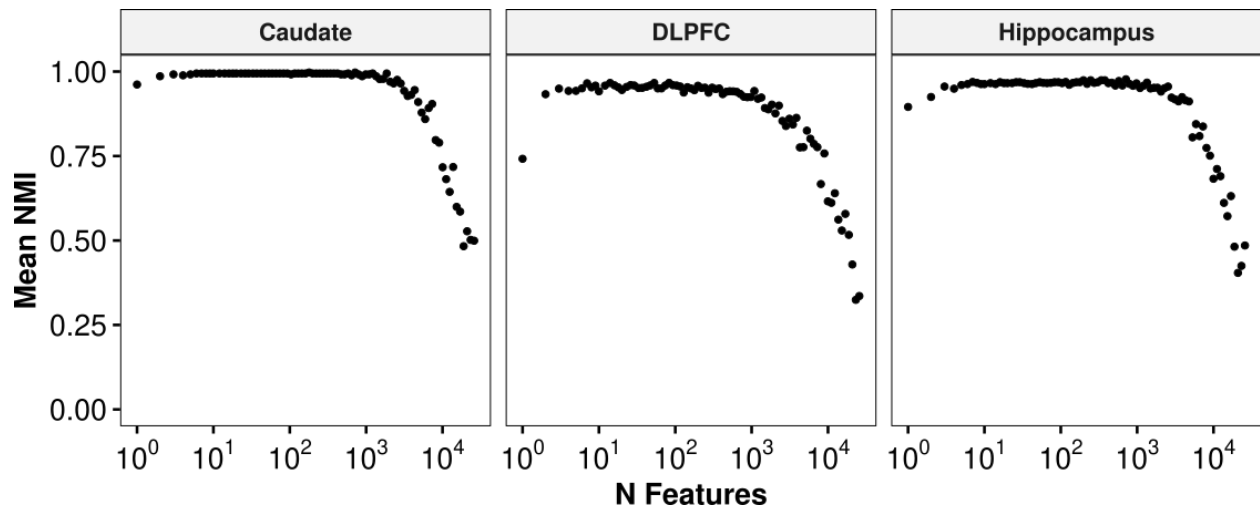


Fig. S8. Dynamic recursive feature elimination curve shows prediction is driven primarily by one gene. Scatterplot showing mean normalized mutual information (NMI) training score across ten folds as a function of N features (x-axis, \log_{10} scale).

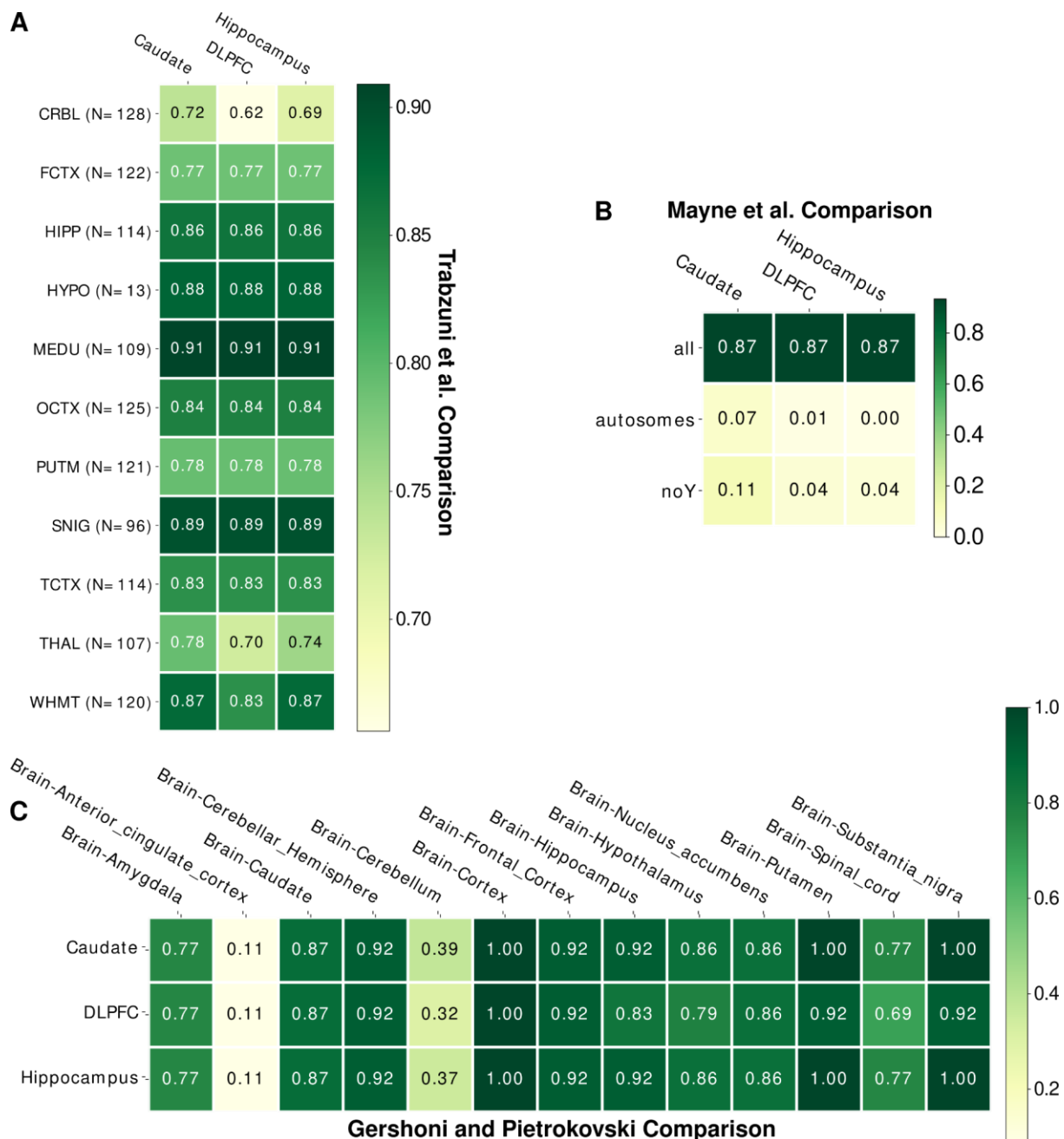


Fig. S9. High overlap of sex-specific differentially expressed genes (DEGs) across multiple datasets and brain regions. Heatmaps comparing BrainSeq Consortium brain regions that show the ratio of DEGs overlap with **A.** Trabzuni et al. brain regions, **B.** Mayne et al. meta analysis with significant expression in at least one brain region using all chromosomes, only autosomes, or no Y chromosomes, and **C.** Gershoni and Pietrokovski brain regions. Abbreviations: CRBL: cerebellum, FCTX: frontal cortex, HIPP: hippocampus, HYPO: hypothalamus, MEDU: medulla, OCTX: occipital cortex, PUTM: putamen, SNIG: substantia nigra, TCTX: temporal cortex, THAL: thalamus, and WHMT: white matter.

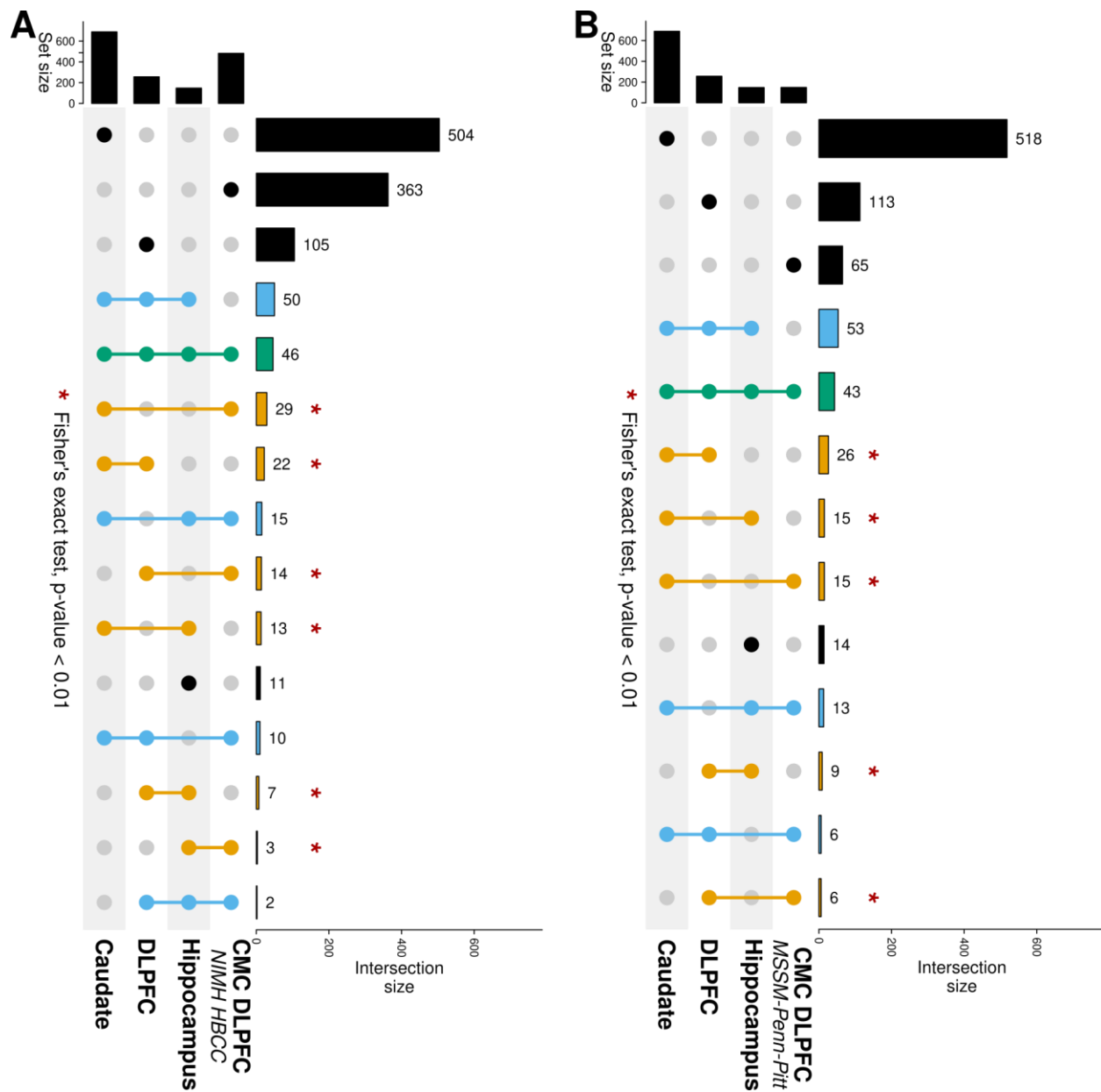


Fig. S10. Significant sharing of sex-specific differentially expressed genes (DEGs) across brain regions replicate in the CommonMind Consortium (CMC) DLPFC. UpSet plot showing number of DEGs shared across brain regions with the CMC DLPFC cohort **A.** NIMH HBCC and **B.** MSSM-Penn-Pitt. * Indicating p-value < 0.01 for two-tailed, Fisher's exact test. Green is shared across the four brain regions; blue, shared across three brain regions; orange, shared between two brain regions; and black, unique to a specific brain region.

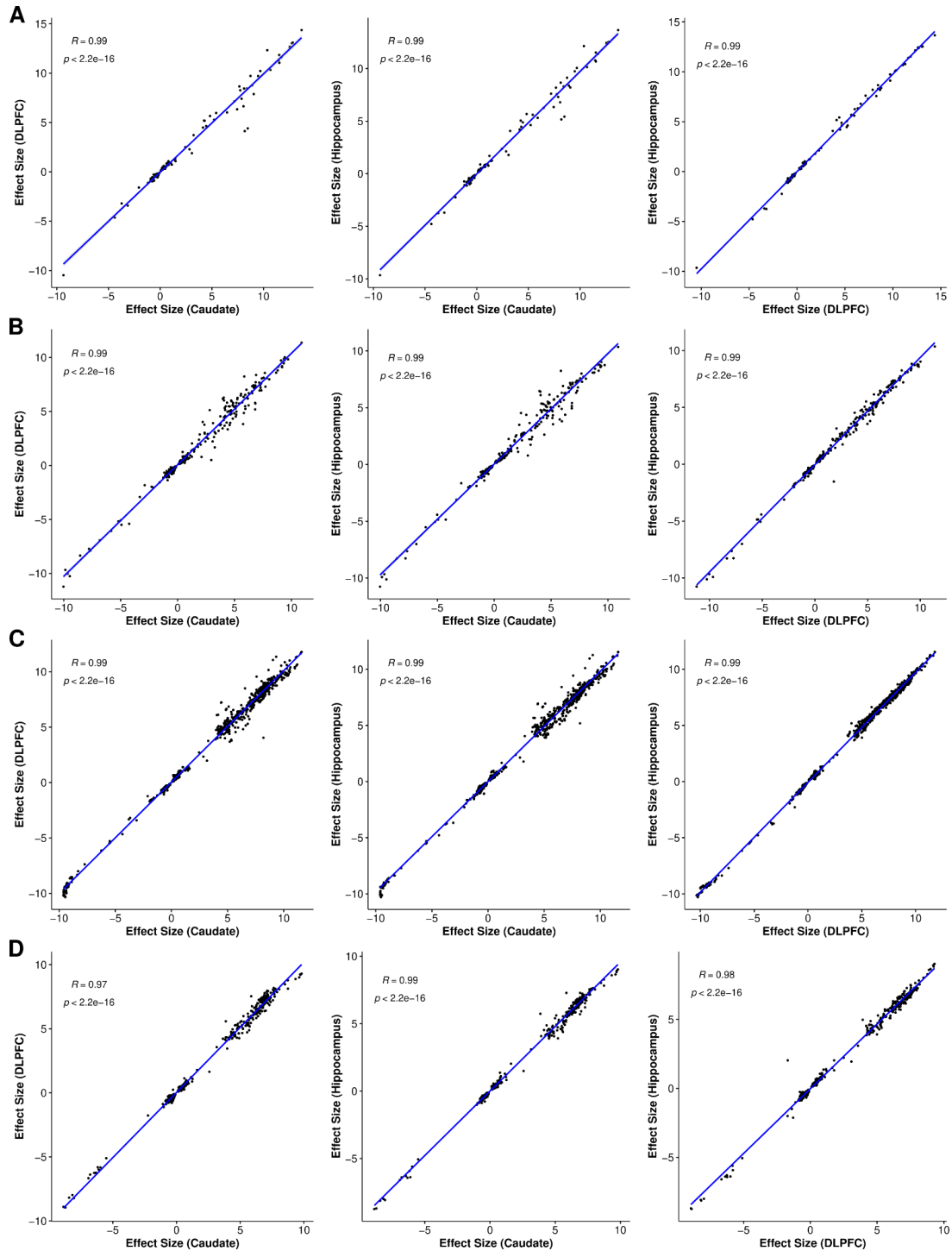


Fig. S11. All significantly differentially expressed features (FDR < 0.05) have concordant directionality for sex differences across the three brain regions. Scatterplot of Spearman correlation comparing effect size (logFC) from sex differentially expressed features (adjusted p-value < 0.05) between brain region pairs for **A.** genes, **B.** transcripts, **C.** exons, and **D.** exon-exon junctions. A fitted trend line is presented in blue as the mean values +/- standard deviation.

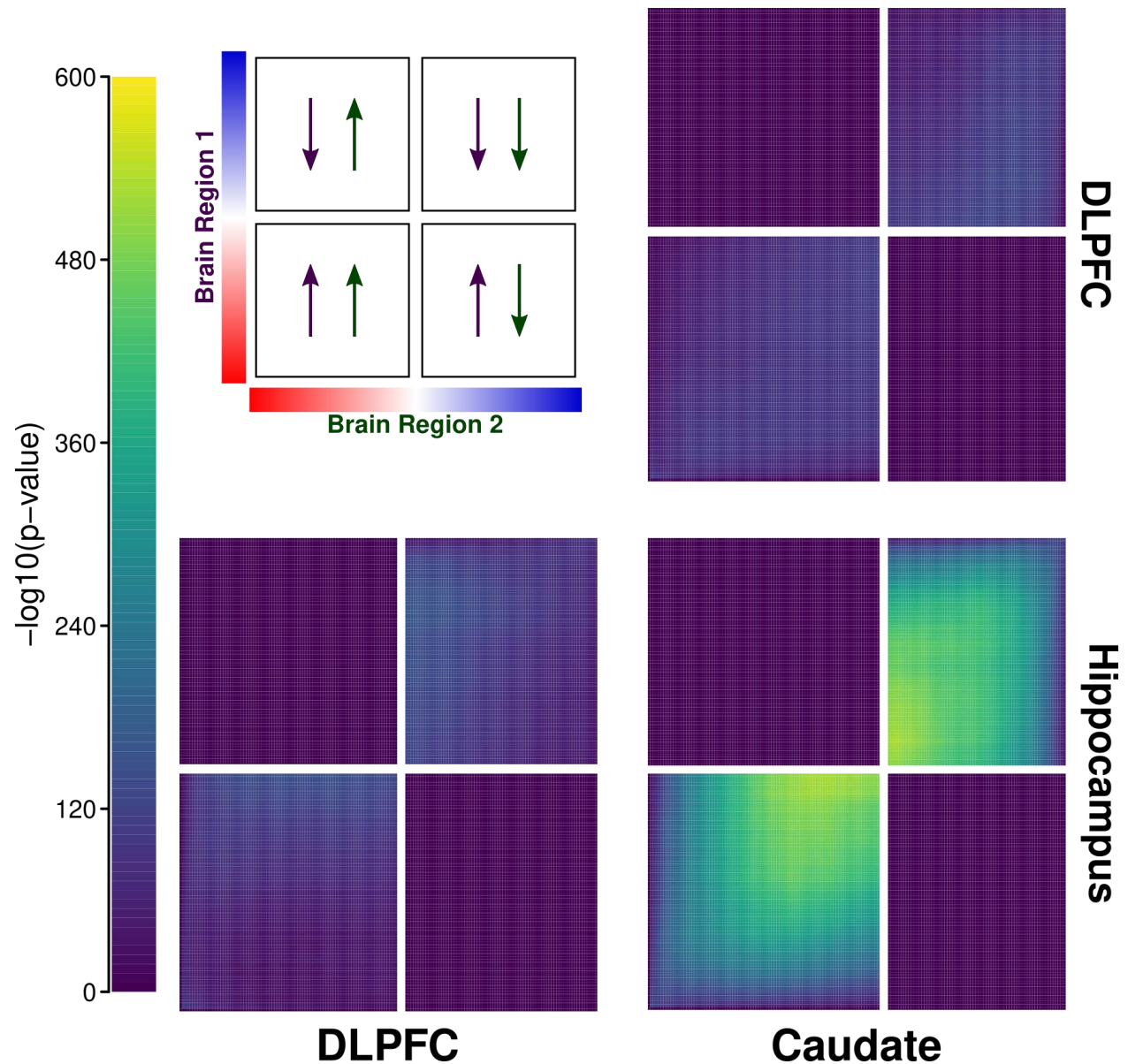


Fig. S12. Concordant transcriptional changes for sex-specific expression across the brain. RRHO (rank-rank hypergeometric overlap) maps comparing sex-specific transcriptional changes for all genes between brain region pairs stratified by direction of effect. There are no genes with discordant direction of effect. The panel presents the overlapping relationship between two brain regions. The color bar represents the degree of significance $[-\log_{10}(\text{p-value})]$ of overlap between two brain regions. Arrows show the direction of effect for sex (female biased [downregulated] or male biased [upregulated]) by brain region.

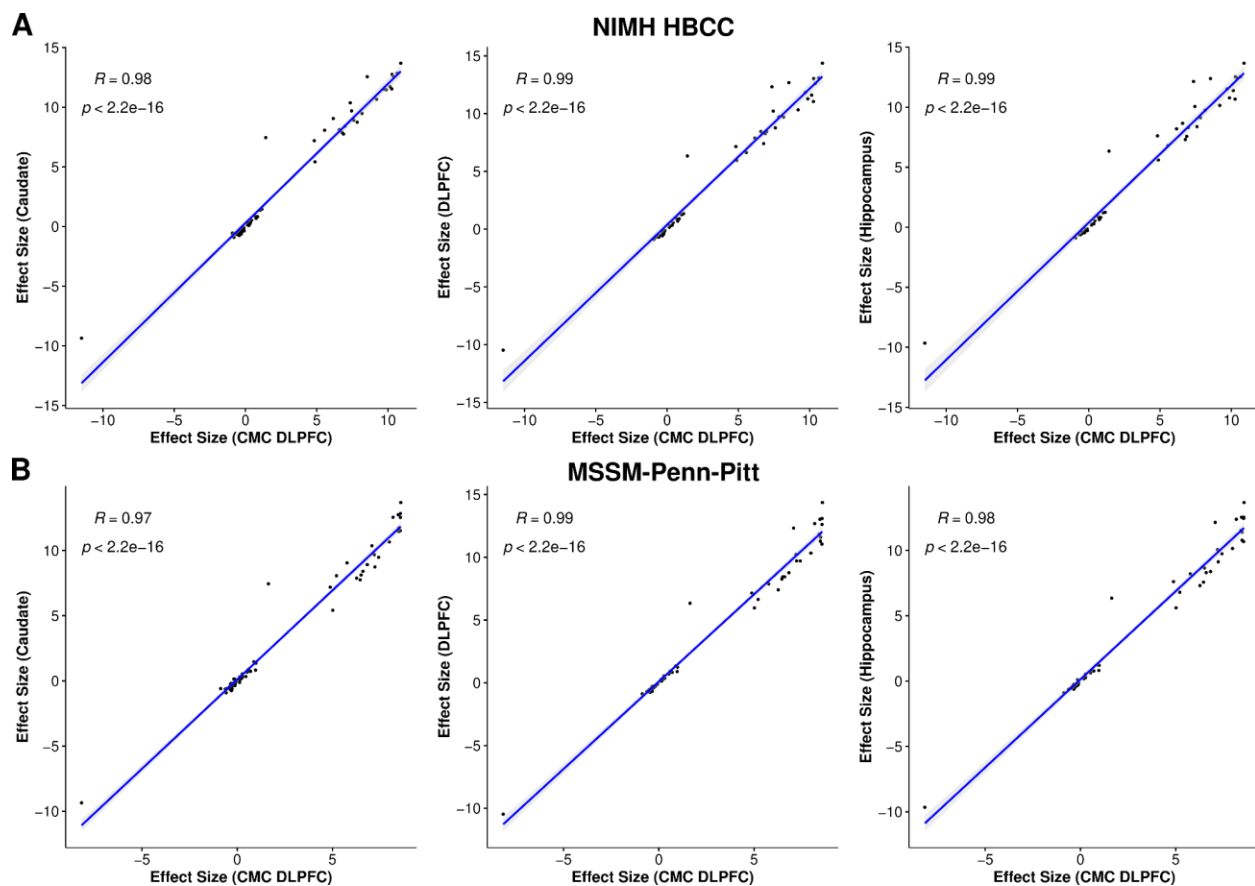


Fig. S13. Replication of concordant directionality of sex DEGs with CommonMind (CMC) DLPFC. Scatterplot comparing effect size (logFC) between BrainSeq Consortium brain regions (caudate nucleus, DLPFC, and hippocampus) and CMC DLPFC cohort **A**. NIMH HBCC and **B**. MSSM-Penn-Pitt. A fitted trend line is presented in blue as the mean values \pm standard deviation. The standard deviation is shaded in light gray.

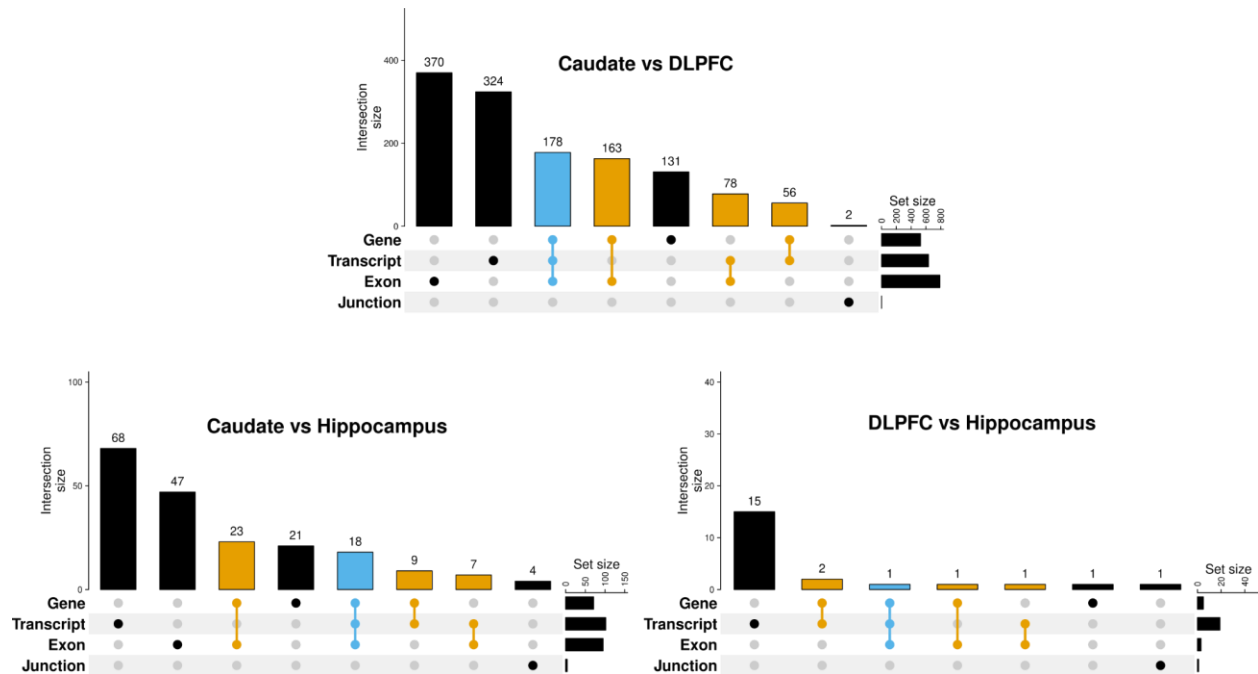


Fig. S14. Unique differentially expressed genes (DEGs) double with the inclusion of isoform-level analysis. Blue, shared across three brain regions; orange, shared between two brain regions; and black, unique to a specific brain region. Novel junctions not annotated to unique gene ID.

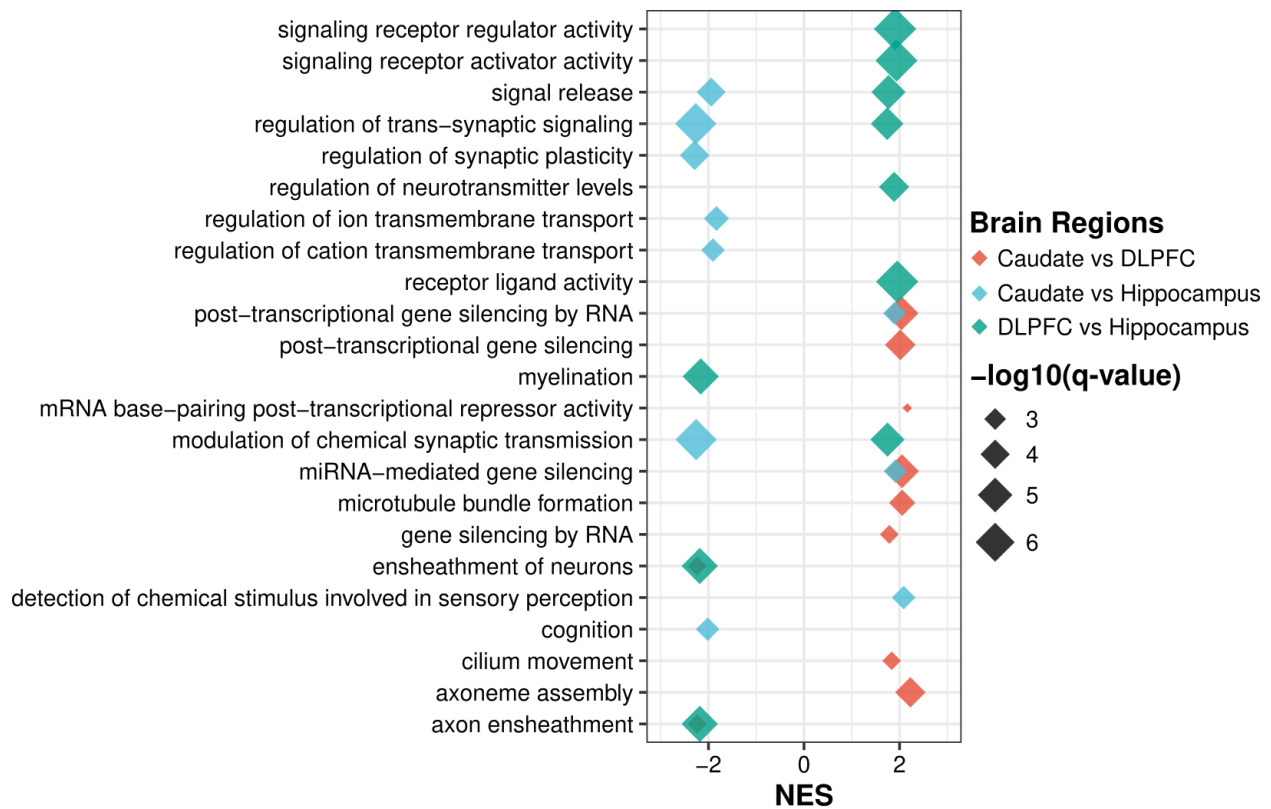


Fig. S15. Brain region-specific transcriptional sex differences show significant enrichment for myelination, cognition, and gene silencing. Gene set enrichment analysis (GSEA) of brain region-specific sex differential expression analysis between brain regions. NES: normalized enrichment score.

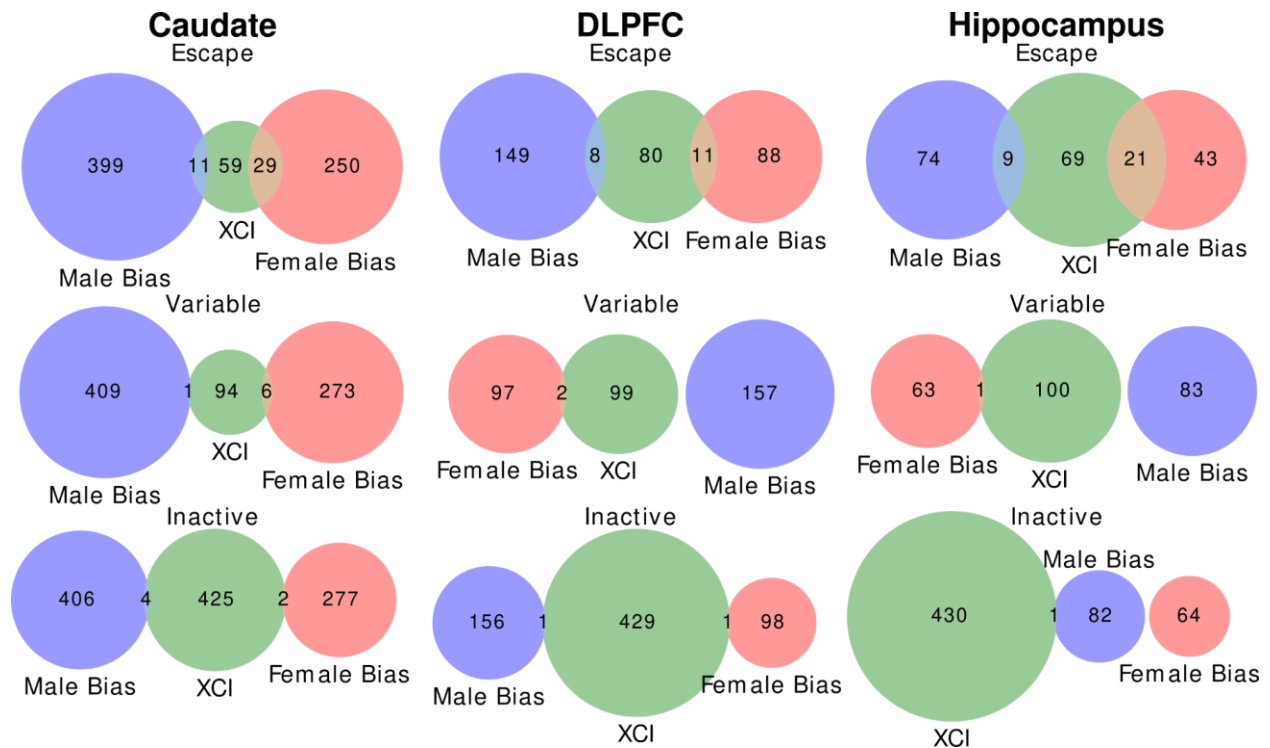


Fig. S16. Large overlap of escaping X-chromosome inactivation (XCI) with sex-specific differentially expressed genes (DEGs). Venn diagram demonstrating large overlap of escaping X-chromosome inactivation (XCI) genes with sex-specific differentially expressed genes (DEGs) upregulated in female individuals (female bias; red) compared with variable and inactive XCI genes for the caudate nucleus, DLPFC, and hippocampus. DEGs upregulated in male individuals (male bias; blue). XCI genes in green.

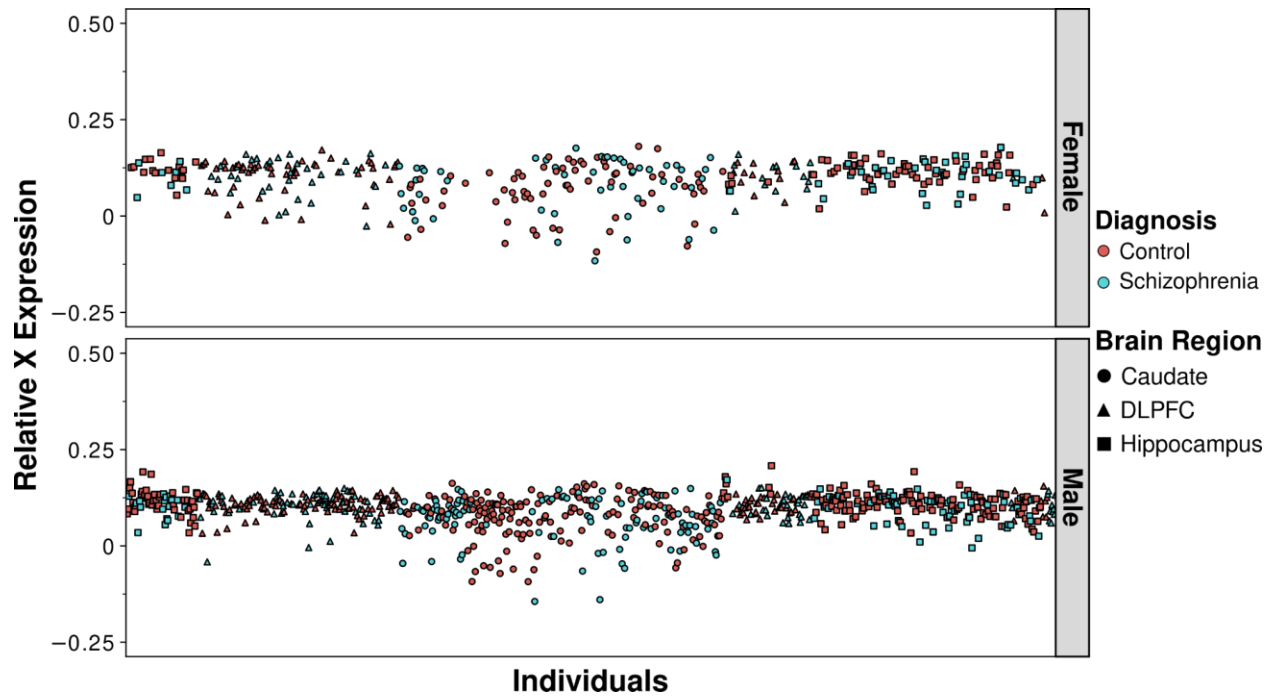


Fig. S17. Dosage is properly compensated in the caudate nucleus, DLPFC, and hippocampus. Scatterplot of relative X expression (RXE) across brain regions separated by sex. Control (red), Schizophrenia (blue), caudate nucleus (circle), DLPFC (triangle), and hippocampus (square).

A

$$\text{Relative X Expression (RXE)} = \log_2(X) - \log_2(A)$$

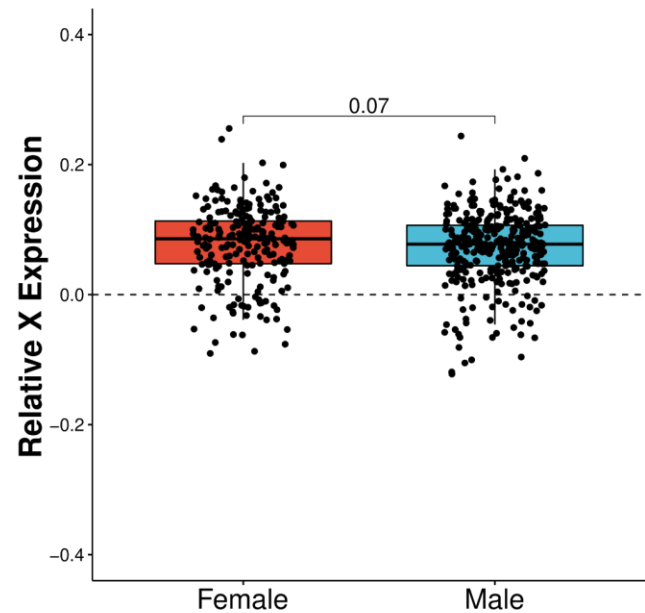
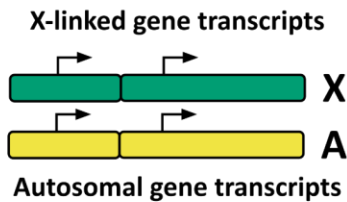
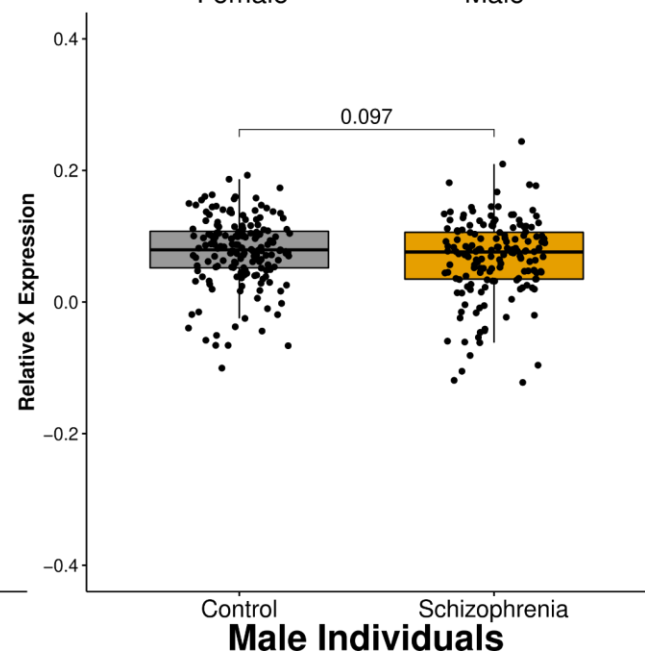
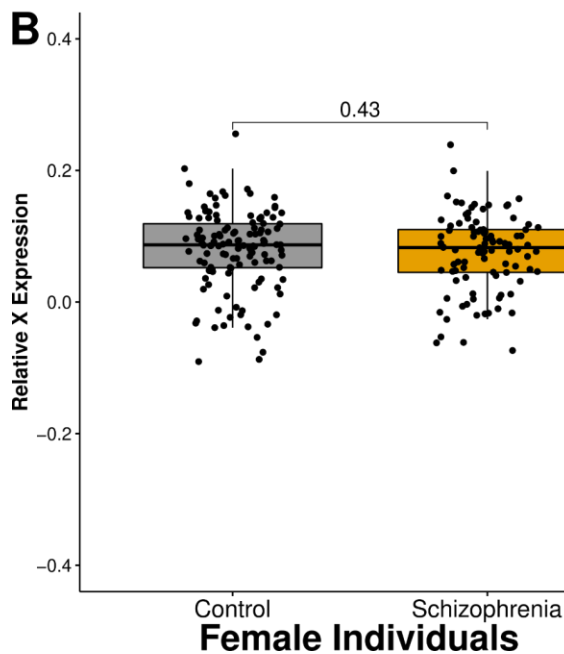
**B**

Fig. S18. Replication of relative X expression (RXE) sex differences within the DLPFC of CommonMind Consortium (CMC). **A.** Schematic of RXE (left) and box plots showing RXE comparison between female (red) and male (blue) individuals for the CMC DLPFC (right). **B.** Box plots showing RXE comparison between neurotypical controls (gray) and schizophrenia (gold) individuals for female (left) and male (right) individuals in the CMC DLPFC. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range. Annotations are p-values from Mann-Whitney U two-tailed test.

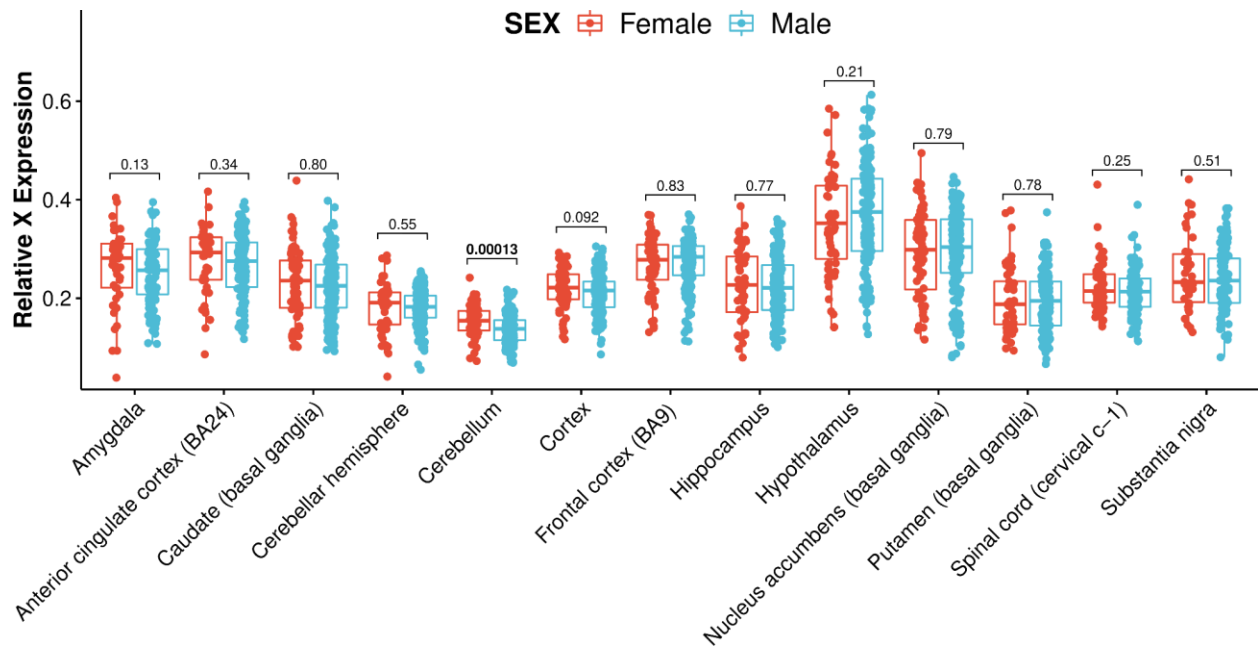


Fig. S19. Replication of relative X expression (RXE) sex differences within the 13 brain regions of GTEx shows RXE variation within GTEx brain regions. Box plots showing RXE comparison between female (red) and male (blue) individuals for the 13 GTEx brain regions. Box plots show the median and first and third quartiles, and whiskers extend to 1.5× the interquartile range. Annotations of p-values are derived from Mann-Whitney U two-tailed test. Significant differences are bolded.

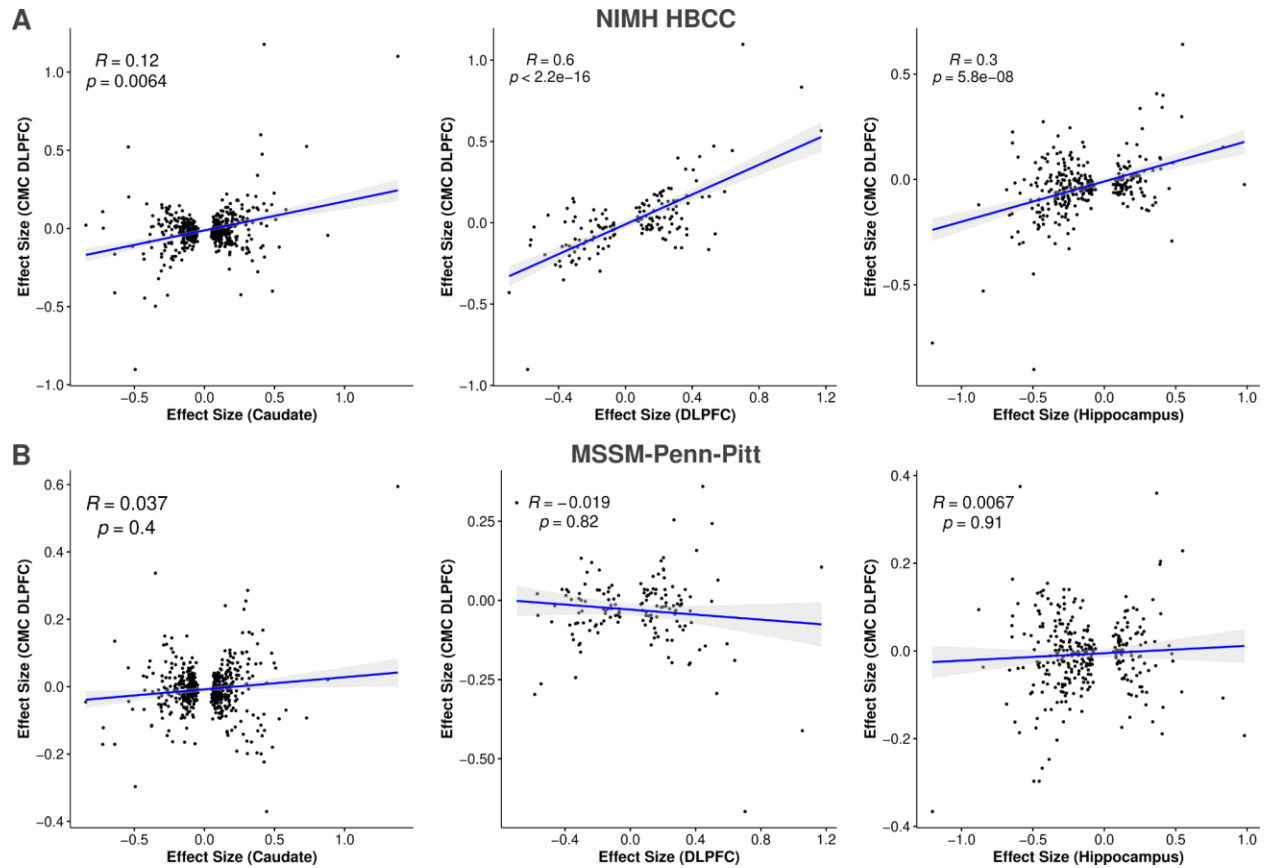
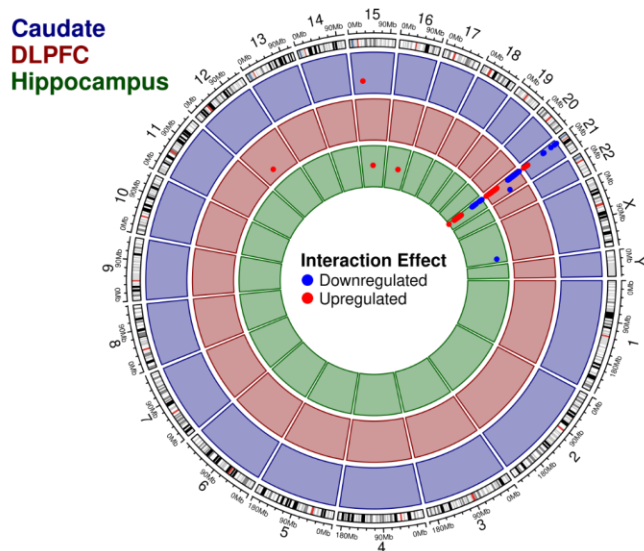


Fig. S20. Limited replication of nominally significant, sex-interacting, schizophrenia-associated DEGs (differentially expressed genes) shared with CommonMind Consortium (CMC) DLPFC (NIMH HBCC cohort). Scatterplot comparing effect size (logFC) between BrainSeq Consortium brain regions (caudate nucleus, DLPFC, and hippocampus) and CMC DLPFC cohort **A.** NIMH HBCC and **B.** MSSM-Penn-Pitt. A fitted trend line is presented in blue as the mean values \pm standard deviation. The standard deviation is shaded in light gray.

A



B

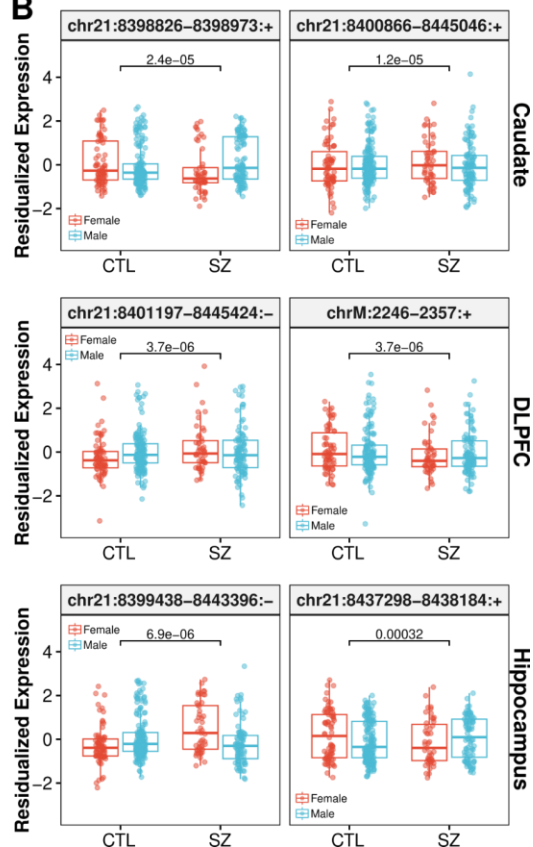


Fig. S21. Differentially expressed junctions demonstrating sex and diagnosis interaction in the caudate nucleus. **A.** Circos plot showing significant DE junctions for the caudate nucleus (blue), DLPFC (red), and hippocampus (green) across all chromosomes. **B.** Box plots of the most significant upregulated and downregulated DE junction by brain region. Neurotypical controls (CTL), schizophrenia (SZ), female (red), and male (blue). Adjusted p-value from the DE-interacting model is annotated on the box plots. Box plots show the median and first and third quartiles, and whiskers extend to $1.5 \times$ the interquartile range.

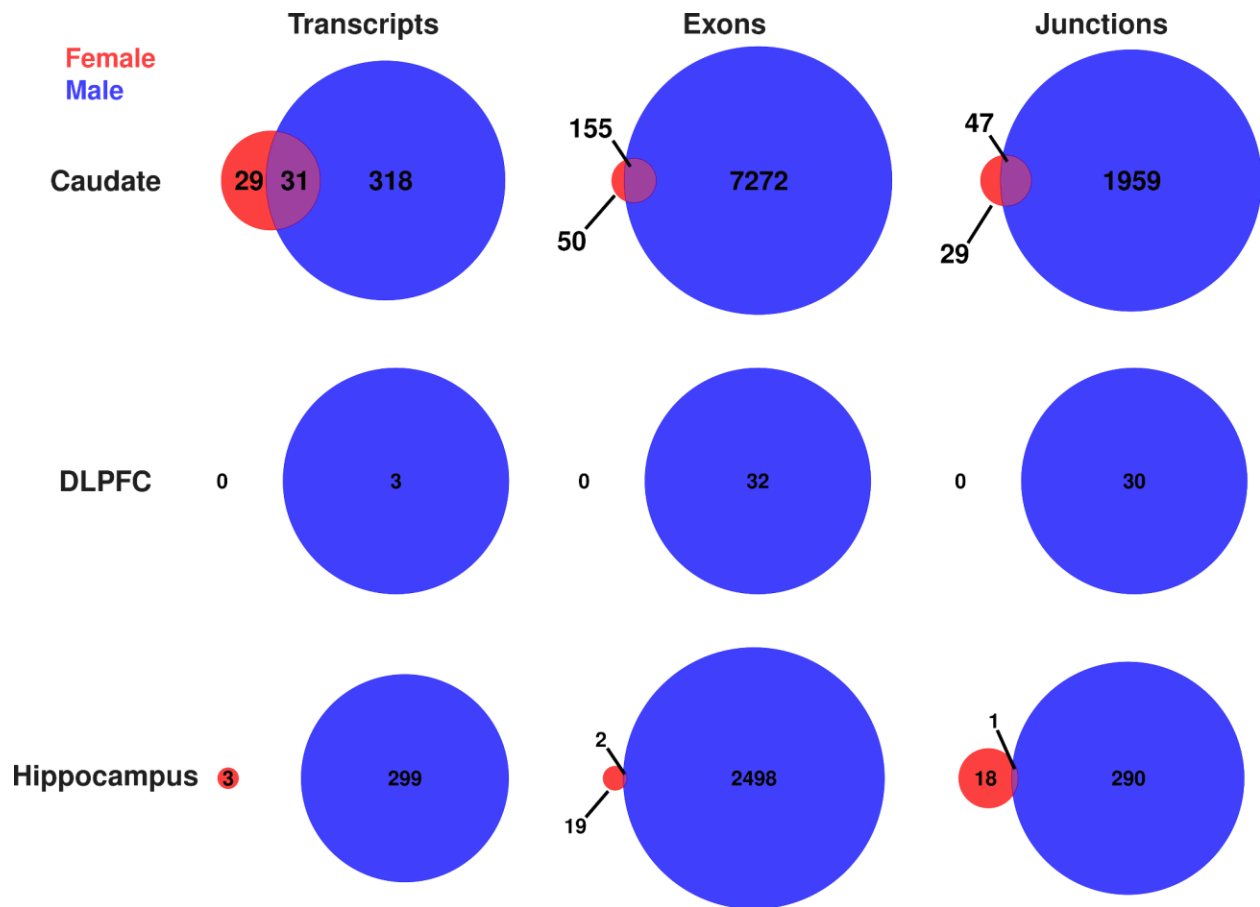


Fig. S22. Little identification of female-specific schizophrenia differential features (transcripts, exons, and junctions) across the caudate nucleus, DLPFC, and hippocampus. Venn diagram showing little to no overlap for DLPFC and hippocampus between female (red) and male (blue) schizophrenia differentially expressed transcripts, exons, and exon-exon junctions.

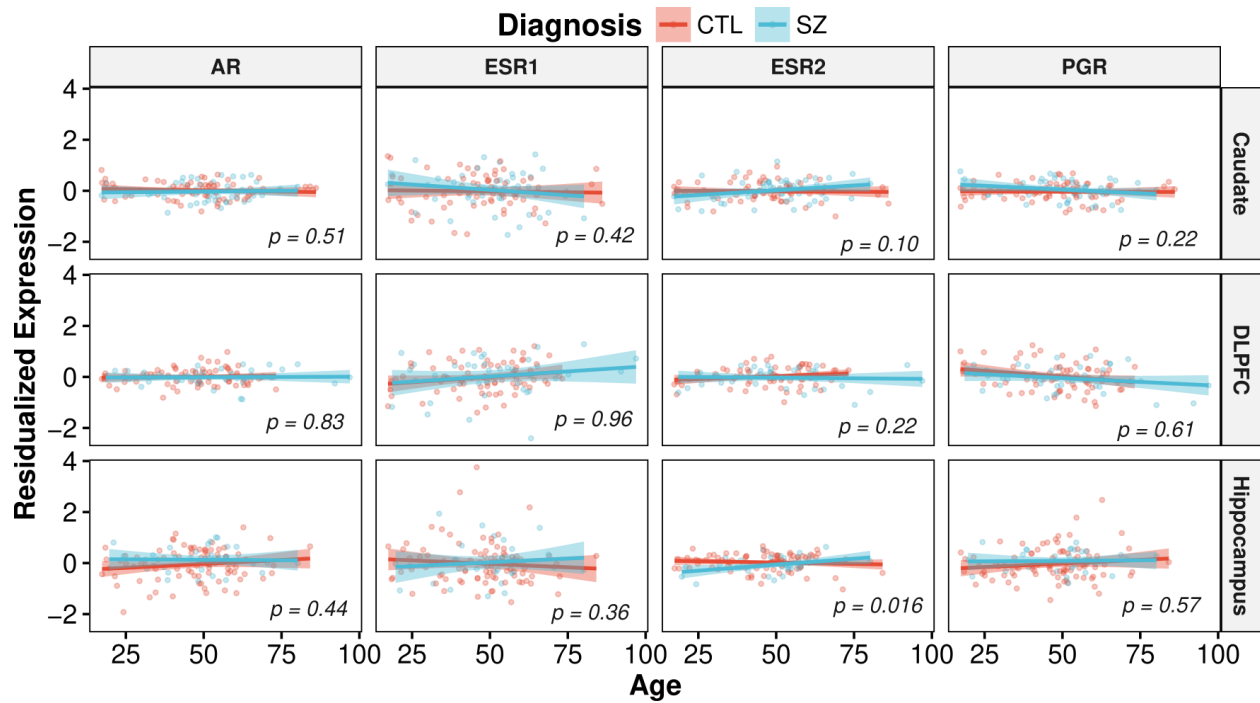


Fig. S23. Nominal significant interaction of age and diagnosis in the hippocampus for *ESR2* for female individuals. Scatterplot of residualized expression showing correlation with age as a function of diagnosis status (control [CTL] in red and schizophrenia [SZ] in blue). A fitted trend line is presented as the mean values \pm standard deviation separated by diagnosis status (control in red and schizophrenia in blue). The standard deviation is shaded in by diagnosis status (control in red and schizophrenia in blue).

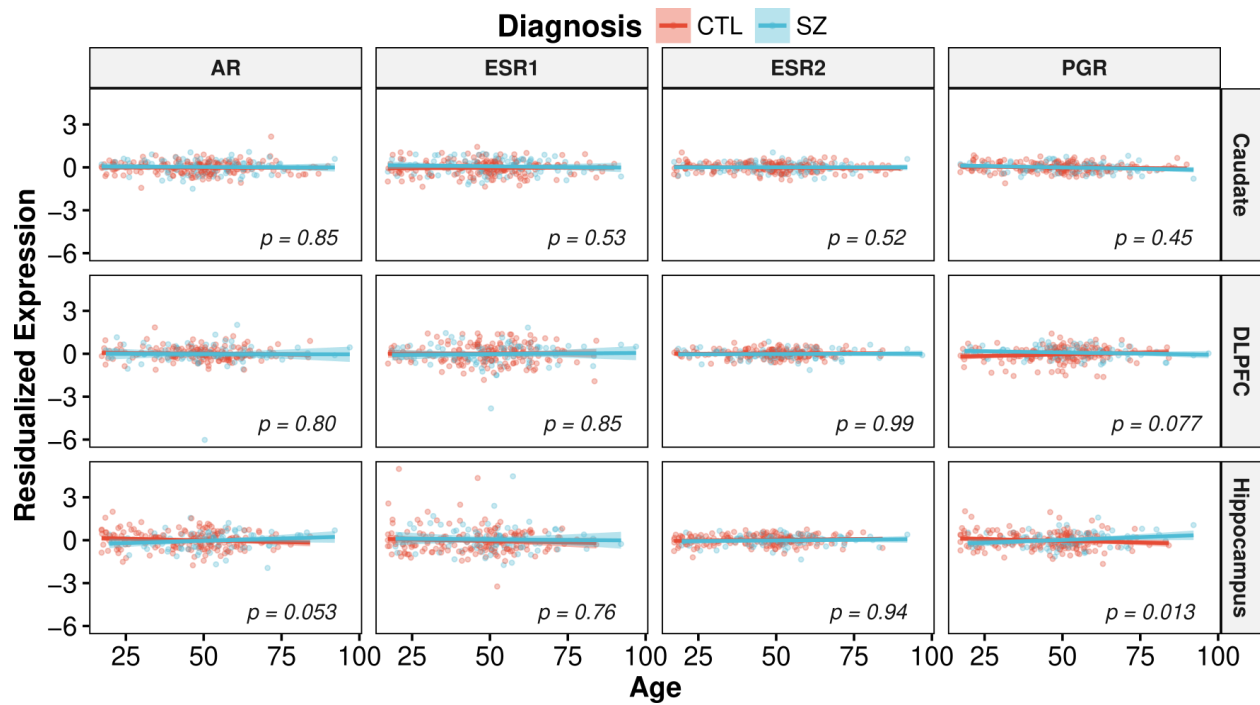


Fig. S24. No significant interaction of age and diagnosis in the hippocampus for *ESR2* for male individuals. Scatterplot of residualized expression showing no significant correlation with age as a function of diagnosis status (control [CTL] in red and schizophrenia [SZ] in blue). A fitted trend line is presented as the mean values \pm standard deviation separated by diagnosis status (control in red and schizophrenia in blue). The standard deviation is shaded in by diagnosis status (control in red and schizophrenia in blue).

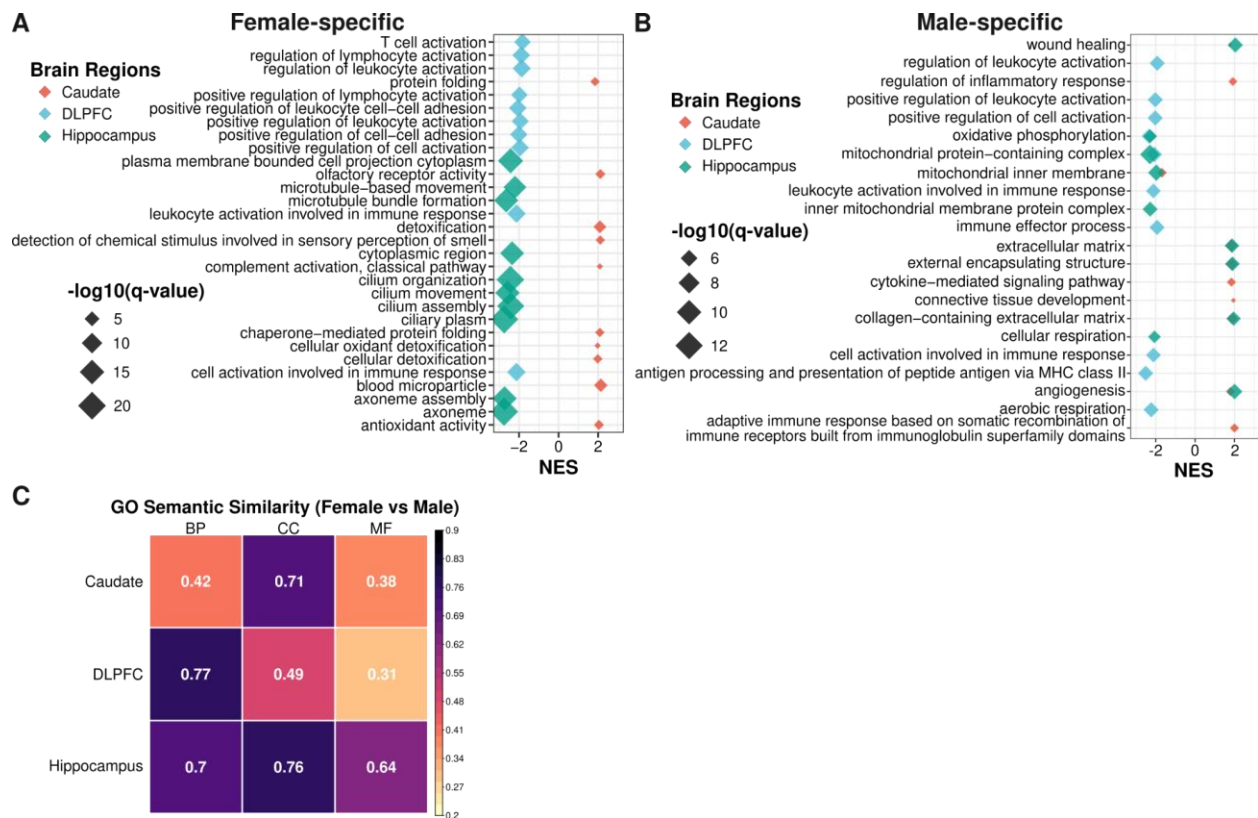


Fig. S25. Sex-specific schizophrenia DEGs show large overlap and enrichment for immune-related pathways. Plot of the most significant by p-value gene terms from the gene term enrichment analysis. NES > 0: upregulated in individuals with schizophrenia and NES < 0: downregulated in individuals with schizophrenia.

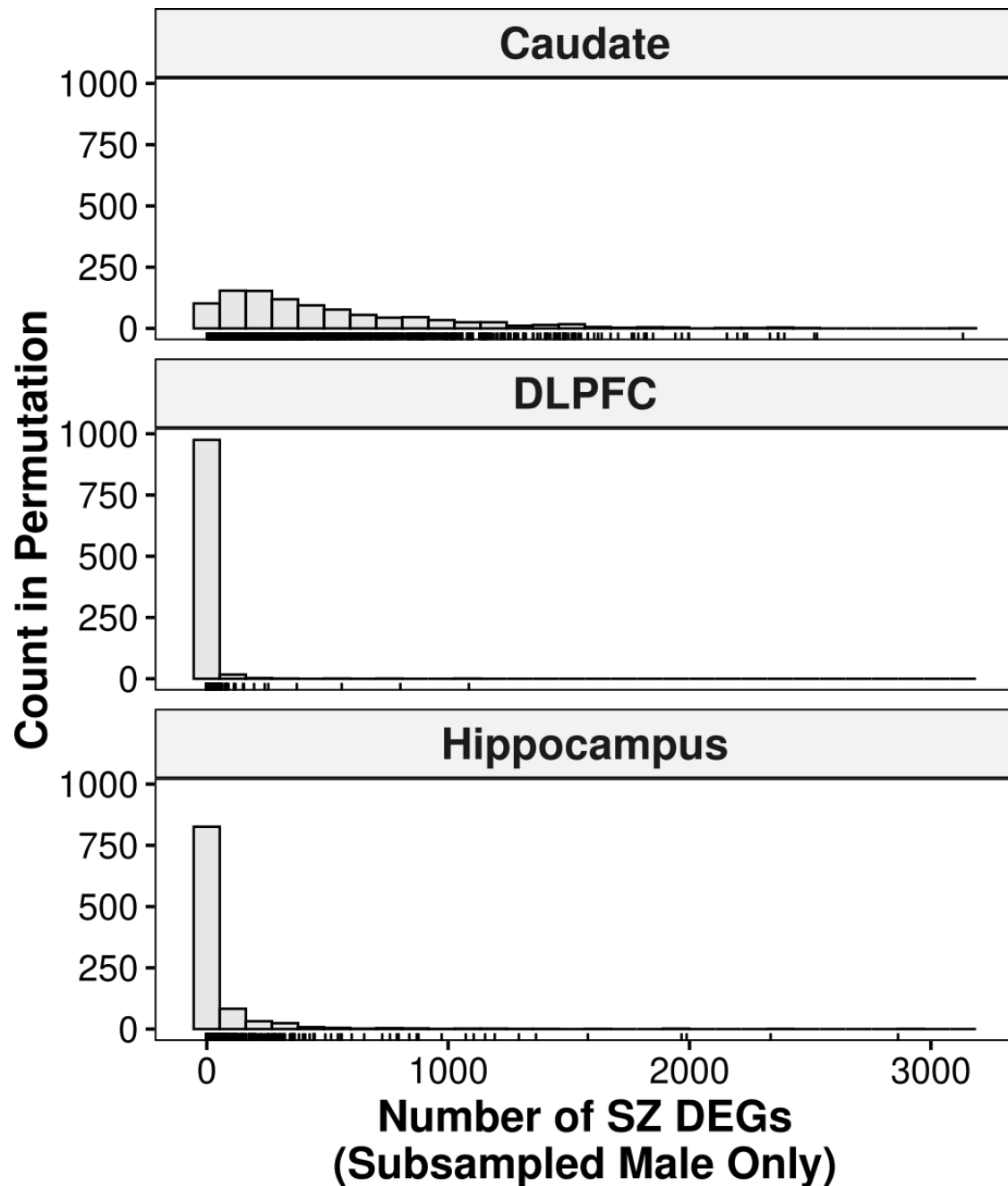


Fig. S26. Reduction of the detected schizophrenia differentially expressed genes (DEGs) in male-only analysis at smaller sample sizes. Histogram of permutation analysis of male-only differential expression analysis at female sample size levels ($n=121$, 114 , and 121 for the caudate nucleus, DLPFC, and hippocampus, respectively) shows a reduction of detected schizophrenia DEGs similar to female-only schizophrenia analysis.

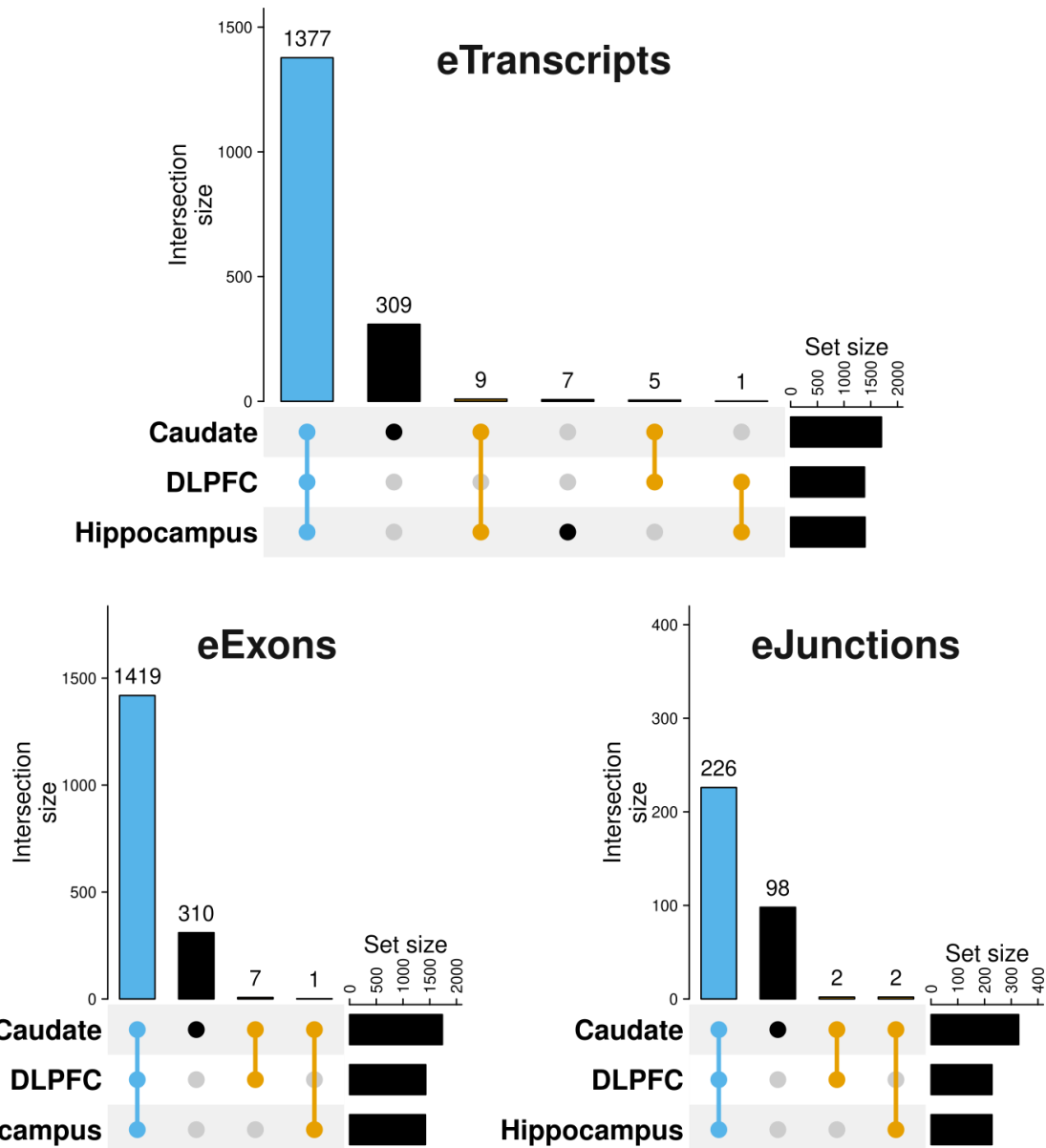


Fig. S27. si-eQTLs are shared across brain regions. UpSet plots showing the majority of si-eQTL are shared across features for eTranscripts (si-eQTL associated with unique transcripts), eExons (si-eQTL associated with unique exons), and eJunctions (si-eQTL associated with unique junctions). Blue is shared across three brain regions; orange, shared between two brain regions; and black, unique to a specific brain region.

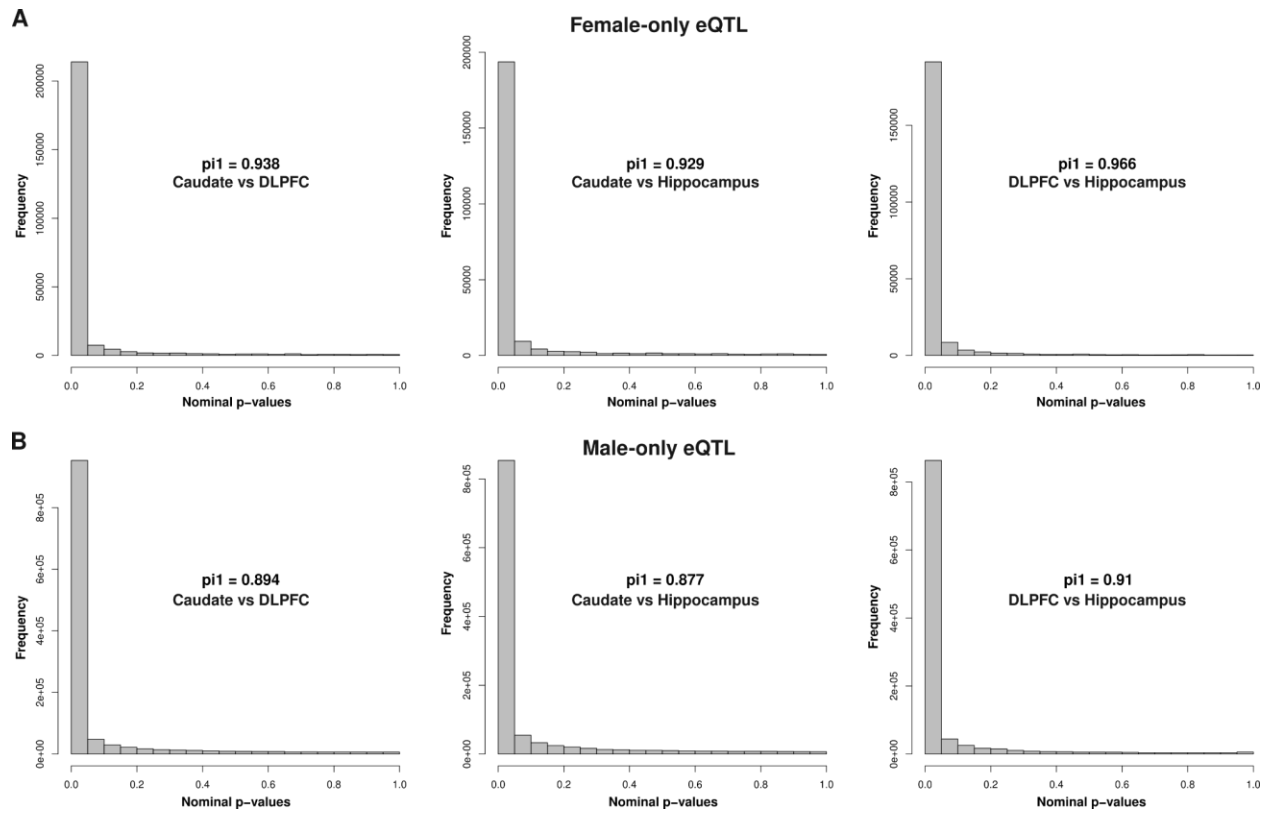


Fig. S28. High level of replication of si-eQTL across brain regions. Histogram of significant si-eQTL ($lfsr < 0.05$) from nominal p-values generated from **A.** female-only and **B.** male-only eQTL analyses. π_1 (π_1) statistic annotated on histograms.

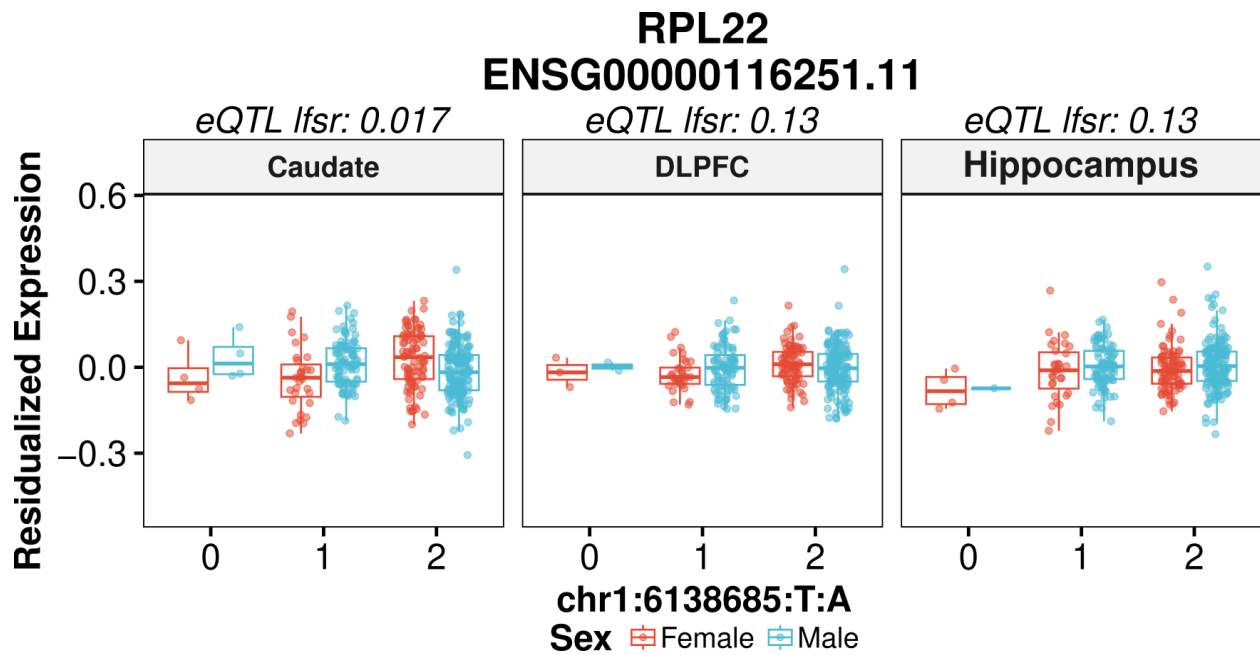


Fig. S29. Example box plot of caudate nucleus-specific si-eQTL. Example of caudate nucleus-specific si-eQTL ($lfsr < 0.05$). All other si-eQTL are shared with the caudate nucleus. Box plots show the median and first and third quartiles, and whiskers extend to $1.5 \times$ the interquartile range.

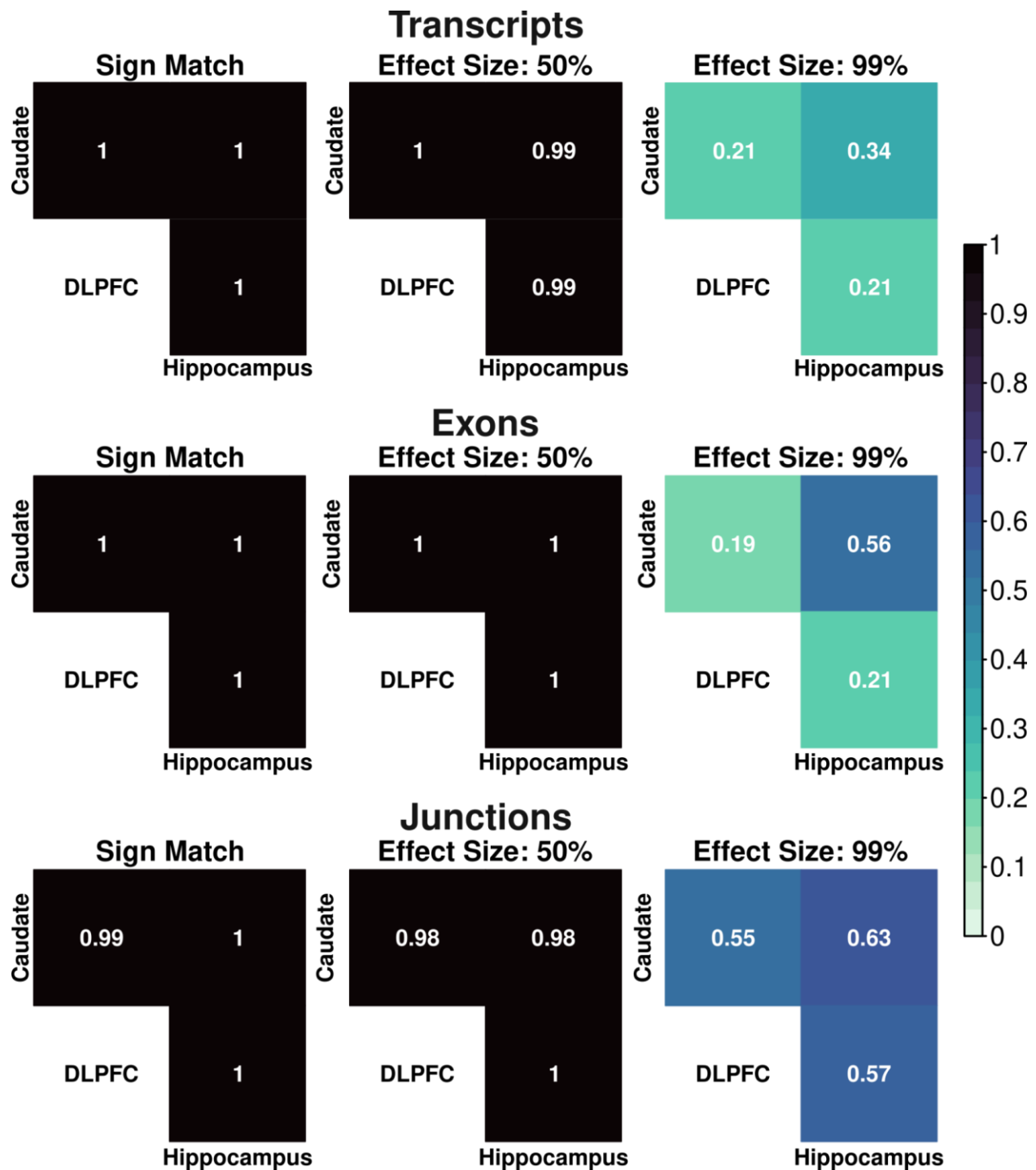


Fig. S30. si-eQTL SNP-gene pairs shared across brain regions demonstrate concordant directionality. Heatmap of proportion of feature level (transcripts, exons, and exon-exon junctions) si-eQTL sharing with the sign of si-eQTL matching (left), the same sign of si-eQTL and within a factor of 0.5 effect size (middle), and the same sign of si-eQTL and within a factor 0.99 effect size (right) using mashr⁶³. A factor of one is a perfect effect size match.

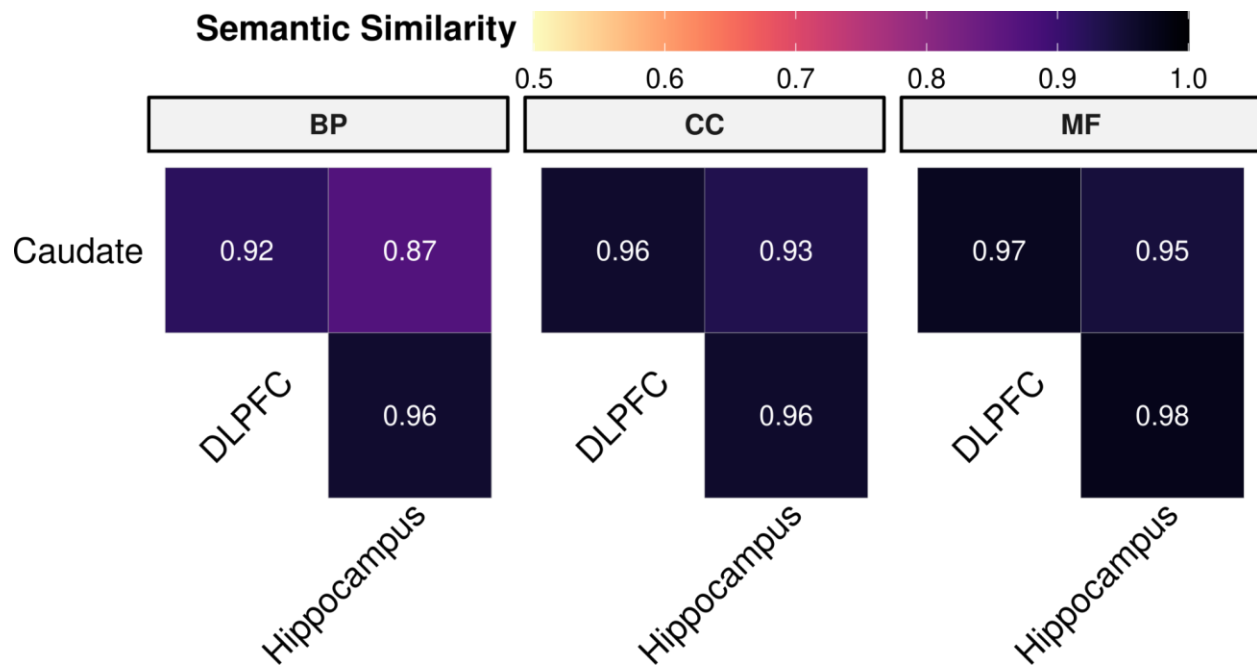


Fig. S31. Sex-interacting eGenes share functional semantic similarity across brain regions. Heatmap of significant term enrichment for the Gene Ontology database (BP: Biological Process, CC: Cellular Component, and MF: Molecular Function) across brain regions.

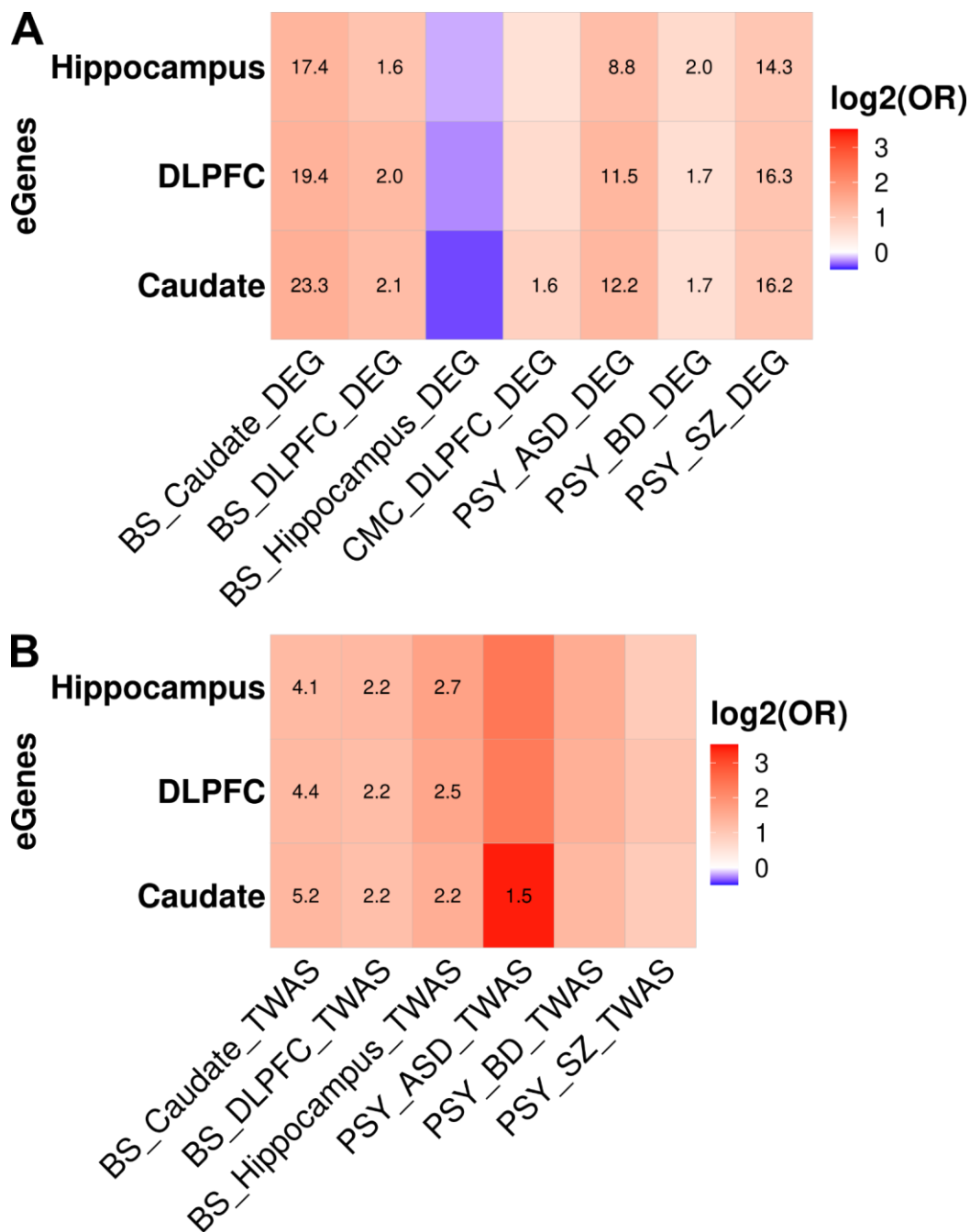


Fig. S32. Sex-interacting eGenes are enriched for neuropsychiatric disorders. Heatmap of enrichment (red) / depletion (blue) for sex-interacting eGenes with **A.** differential expression genes (DEG) or **B.** transcriptome-wide association studies (TWAS) for neuropsychiatric disorders for the caudate nucleus, DLPFC, and hippocampus. Significant enrichments (Fisher's exact test, FDR corrected p-values, $-\log_{10}$ transformed) annotated within tiles. BS: BrainSeq Consortium, CMC: CommonMind Consortium, PSY: psychENCODE. For BrainSeq Consortium and CMC, DEG and TWAS are for schizophrenia (SZ). ASD: Autism spectrum disorder, BD: Bipolar disorder. eGenes: si-eQTL associated with unique genes. psychENCODE results are for DLPFC.

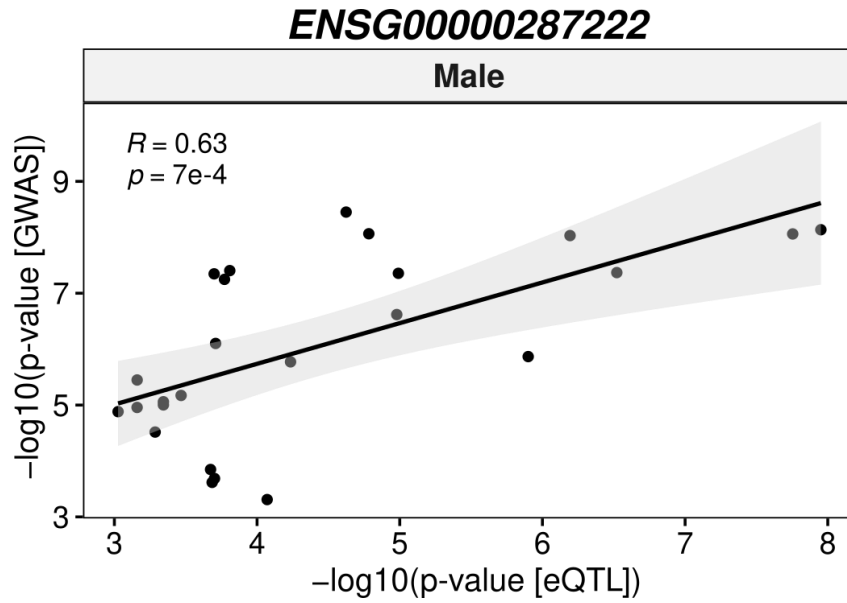


Fig. S33. Colocalization of si-eQTL and schizophrenia risk in the caudate nucleus. P-P plot of the significant schizophrenia GWAS associations ($RCP > 0.5$). Gene name annotated on top of scatterplot. A fitted trend line is presented in black as the mean values \pm standard deviation. The standard deviation is shaded in light gray.

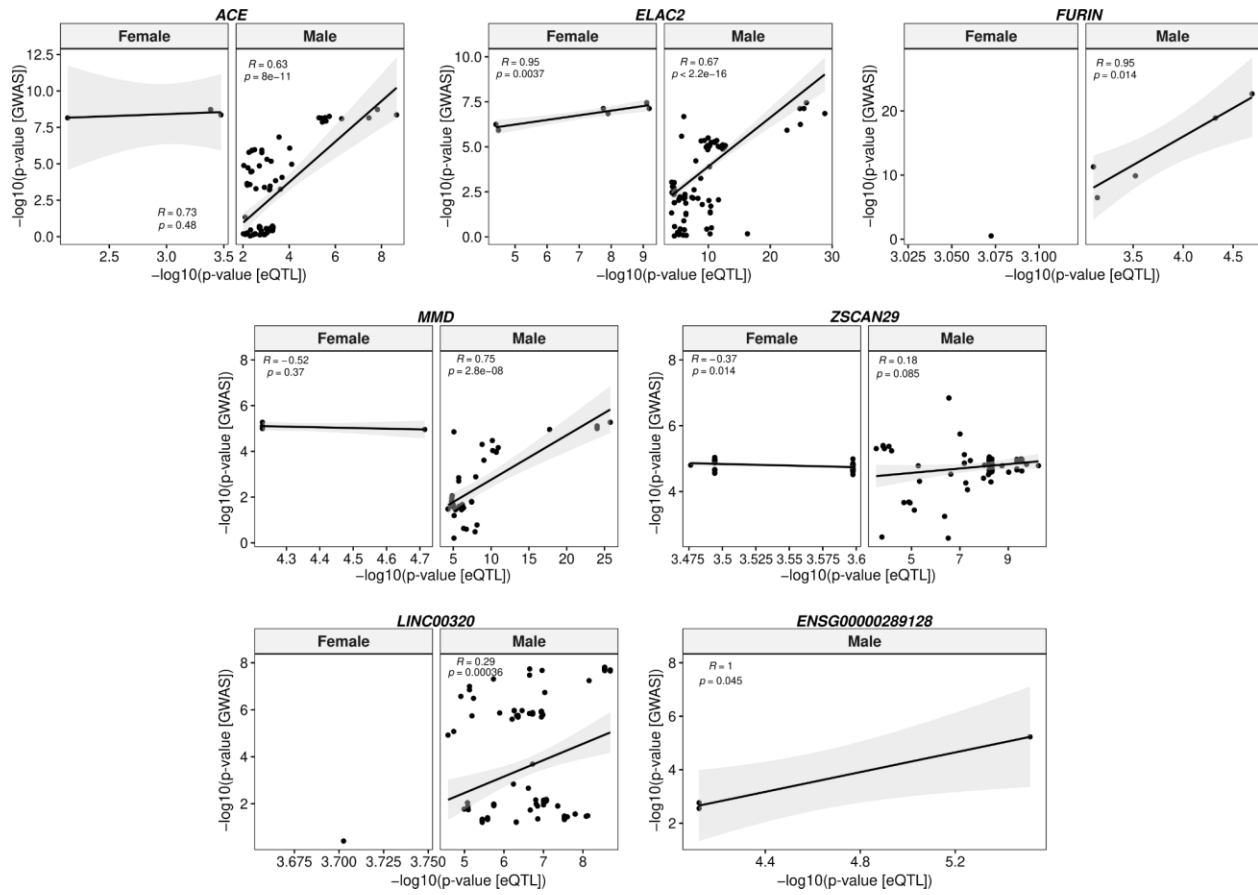


Fig. S34. Colocalization of si-eQTL and schizophrenia risk in the DLPFC. P-P plots of the significant schizophrenia GWAS associations (RCP > 0.5). Gene names annotated on top of each scatterplot. A fitted trend line is presented in black as the mean values \pm standard deviation. The standard deviation is shaded in light gray.

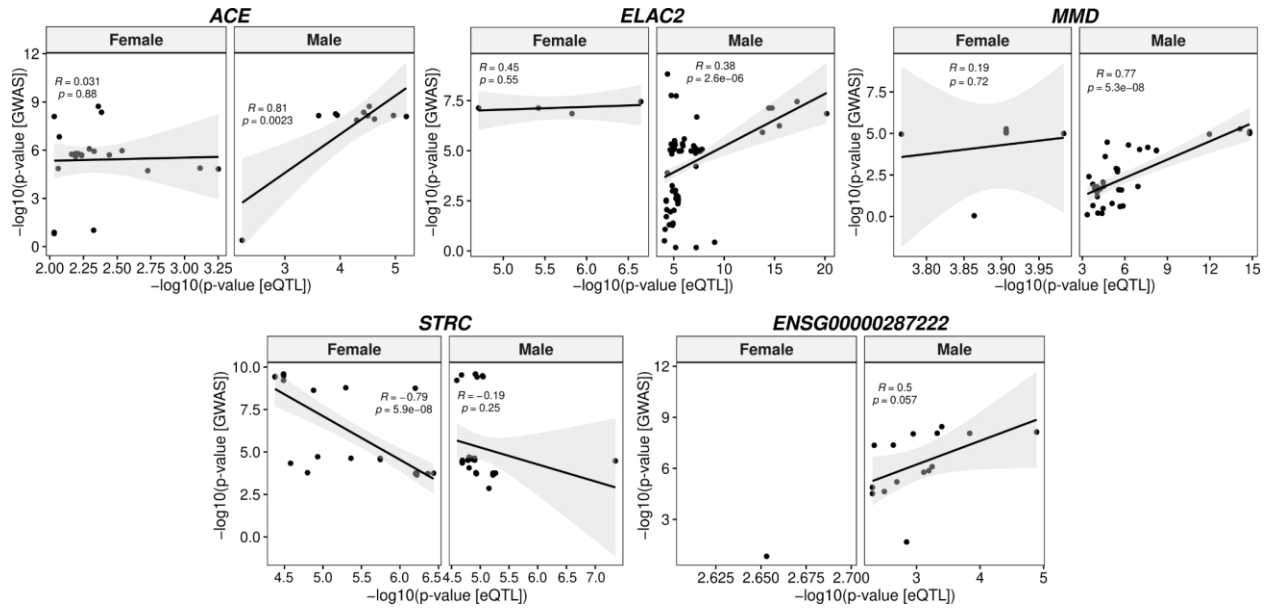


Fig. S35. Colocalization of si-eQTL and schizophrenia risk in the hippocampus. P-P plots of the significant schizophrenia GWAS associations ($RCP > 0.5$). Gene names annotated on top of each scatterplot. A fitted trend line is presented in black as the mean values \pm standard deviation. The standard deviation is shaded in light gray.

Tables

Table S1. Summary of GWAS trait download information including website and download filename.

GWAS Trait	Website	Filename
Schizophrenia ¹¹	https://www.med.unc.edu/pgc/download-results/	PGC3_SCZ_wave3.european.autosome.public.v3.vcf.tsv.gz
Depression ⁹⁷	https://datashare.ed.ac.uk/handle/10283/3203	Summary statistics of 10,000 variants from a meta-analysis of PGC, UK Biobank and 23andMe (444.2Kb)
Bipolar Disorder ⁹⁸	https://www.med.unc.edu/pgc/download-results/	daner_PGC_BIP32b_mds7a_0416a.gz
Autism ⁹⁹	https://www.med.unc.edu/pgc/download-results/	iPSYCH-PGC_ASD_Nov2017.gz
ADHD ¹⁰⁰	https://www.med.unc.edu/pgc/download-results/	daner_adhd_meta_filtered_NA_iPSYCH23_PGC11_singPCs_woSEX_2ell6sd_EUR_Neff_70.meta.gz
Anorexia ¹⁰¹	https://www.med.unc.edu/pgc/download-results/	pgcAN2.2019-07.vcf.tsv.gz
BMI ¹⁰²	https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files	Meta-analysis_Locke_et_al+UKBiobank_2018.txt.gz
Height ⁸³	https://cnsgenomics.com/data/giant_2022/	https://cnsgenomics.com/data/giant_2022/GIANT_HEIGH_T_YENGO_2022_GWAS_SUMMARY_STATS_EUR.gz
Basophil count ¹⁰³	https://ftp.sanger.ac.uk/project/humangen/summary_statistics/human/2017-12-12/	baso_N171846_narrow_form.tsv.gz
Neutrophil count ¹⁰³	https://ftp.sanger.ac.uk/project/humangen/summary_statistics/human/2017-12-12/	neut_N170702_narrow_form.tsv.gz
Eosinophil count ¹⁰³	https://ftp.sanger.ac.uk/project/humangen/summary_statistics/human/2017-12-12/	eo_N172275_narrow_form.tsv.gz
Monocyte count ¹⁰³	https://ftp.sanger.ac.uk/project/humangen/summary_statistics/human/2017-12-12/	mono_N170721_narrow_form.tsv.gz

Table S2. Summary of differential expression results (adjusted p-value < 0.05) by feature (genes, transcripts, exons, and exon-exon junctions) for sex differences in the caudate nucleus, DLPFC, and hippocampus.

Number of differentially expressed features is separated by chromosome location. The number of unique genes associated with transcript, exon, or exon-exon junction in parenthesis. *Novel junctions not annotated to unique gene ID.

Brain Region	Chr Location	Gene	Transcript (Geneid)	Exon (Geneid)	Junction (Geneid)
Caudate nucleus	Allosomal	112	333 (100)	1523 (111)	622 (11)*
	Autosomal	576	252 (232)	1291 (419)	399 (5)*
	Mitochondria	1	2 (2)	0 (0)	3 (1)*
DLPFC	Allosomal	70	298 (97)	725 (69)	609 (10)*
	Autosomal	186	307 (291)	903 (310)	1421 (14)*
	Mitochondria	0	0 (0)	0 (0)	0 (0)*
Hippocampus	Allosomal	79	290 (88)	1361 (101)	516 (9)*
	Autosomal	68	144 (133)	669 (244)	370 (4)*
	Mitochondria	0	0 (0)	0 (0)	1 (1)*

Table S3. Summary of random forest classification with dynamic recursive feature elimination prediction for sex using autosomes accuracy and smallest number of features. Train, test, and smallest set metrics of mean, median, and standard deviation (std) across 10 folds for each feature (gene, transcript, exon, and exon-exon junction) and brain region (caudate nucleus, DLPFC, and hippocampus).

Feature	Brain Region	N Features		Train Accuracy (%)		Test Accuracy (%)	
		Median	Mean \pm Std	Median	Mean \pm Std	Median	Mean \pm Std
Gene	Caudate nucleus	77	349 \pm 609	100	99.8 \pm 0.4	100	100 \pm 0
	DLPFC	32.5	50.6 \pm 29.8	100	99.6 \pm 0.5	100	99.2 \pm 1.3
	Hippocampus	221	355 \pm 538	100	99.7 \pm 0.2	100	99.7 \pm 0.8
Transcript	Caudate nucleus	51	51.9 \pm 7.9	93.8	94.0 \pm 0.8	93.7	92.9 \pm 5.6
	DLPFC	84	91 \pm 20.7	98.6	98.4 \pm 0.4	100	98.6 \pm 2.0
	Hippocampus	60.5	55.8 \pm 41.4	97.3	97.4 \pm 0.3	97.4	97.9 \pm 2.1
Exon	Caudate nucleus	219	196.6 \pm 135	100	99.9 \pm 0.1	100	99.7 \pm 0.8
	DLPFC	56.5	75.4 \pm 62.9	99.1	98.9 \pm 1.1	100	98.6 \pm 2.7
	Hippocampus	324.5	352 \pm 324	99.7	99.7 \pm 0.1	100	99.7 \pm 0.8
Junction	Caudate nucleus	66.5	69.3 \pm 36.5	98.7	98.6 \pm 0.8	97.4	97.7 \pm 1.9
	DLPFC	28	31.6 \pm 16.1	99.1	99.2 \pm 0.3	100	99.2 \pm 1.3
	Hippocampus	48.5	42.3 \pm 18.3	98.8	99.0 \pm 0.4	100	99.7 \pm 0.8

Table S4. Summary results of differential expression analysis (adjusted p-value < 0.05) of interaction for sex and brain region for genes, transcripts, exons, and exon-exon junctions. The number of unique genes associated with transcript, exon, or exon-exon junction in parenthesis. *Novel junctions not annotated to unique gene ID.

Brain Regions	Gene	Transcript (Geneid)	Exon (Geneid)	Junction (Geneid)
Caudate nucleus and DLPFC	528	728 (636)	2780 (789)	694 (3)*
Caudate nucleus and Hippocampus	71	137 (102)	256 (95)	127 (5)*
DLPFC and Hippocampus	5	25 (19)	6 (3)	80 (2)*

Table S5. Summary of π_1 statistic for sex-specific schizophrenia nominally significant DEGs (p-value < 0.05) between BrainSeq (caudate nucleus, DLPFC, and hippocampus) and CMC DLPFC by cohort.

CMC DLPFC cohort	BrainSeq Region	π_1 statistic
MSSM-Penn-Pitt	Caudate nucleus	0
	DLPFC	0
	Hippocampus	0
NIMH HBCC	Caudate nucleus	0.07
	DLPFC	0.51
	Hippocampus	0

Table S6. Summary of stringent sex-specific differential expression analysis (FDR < 0.05) for schizophrenia by sex and feature (gene, transcript, exon, and exon-exon junction) for the caudate nucleus, DLPFC, and hippocampus. The number of unique genes associated with transcript, exon, or exon-exon junction in parenthesis.
*Novel junctions not annotated to unique gene ID.

Brain Region	Sex	Gene	Transcript (Geneid)	Exon (Geneid)	Junction (Geneid)
Caudate nucleus	Female	194	16 (16)	22 (17)	15 (1)*
	Male	1130	190 (178)	3645 (864)	1007 (3)*
DLPFC	Female	0	0 (0)	0 (0)	0 (0)*
	Male	5	1 (1)	23 (3)	17 (1)*
Hippocampus	Female	0	0 (0)	7 (7)	4 (2)*
	Male	149	105 (104)	865 (53690)	72 (3)*

Table S7. Summary of shared genes between two brain regions for stringent male-specific schizophrenia DEGs.

Brain Regions	# of Genes	Gene Names
Caudate nucleus and DLPFC	2	<i>EDN3</i> and <i>PLD4</i>
Caudate nucleus and Hippocampus	23	<i>MIDEAS</i> , <i>IL1R1</i> , <i>SLC11A1</i> , <i>DCP1A</i> , <i>OSMR</i> , <i>ENSG00000272601</i> , <i>TMEM52B</i> , <i>CHD2</i> , <i>ACSL5</i> , <i>MACORIS</i> , <i>GPD1</i> , <i>BCL6</i> , <i>JAK3</i> , <i>MYC</i> , <i>ZBTB16</i> , <i>TAC1</i> , <i>MMD2</i> , <i>IL18BP</i> , <i>PLRG1</i> , <i>ETV6</i> , <i>ZNF395</i> , and <i>SPAG7</i>
DLPFC and Hippocampus	0	Not available

Data

Data S1. BrainSeq_bySex_genes_3region_eQTL_results.tar.gz: Compressed tar file containing gene eQTL results by sex per brain region including all nominal p-values as well as permutation analysis.

Data S2. differential_expression_analysis_4features_sex.txt.gz: Compressed text file of differential expression analysis for sex across the caudate nucleus, DLPFC, and hippocampus for four features (gene, transcript, exon, and exon-exon junction).

Data S3. BrainSeq_MAGMA_enrichment_analysis.xlsx: Excel file of magma enrichment results of all DEGs (i.e., sex-specific, sex interacting with brain region, and sex-specific schizophrenia) separated by direction of effect across the caudate nucleus, DLPFC, and hippocampus.

Data S4. BrainSeq_sex_specific_functional_enrichment_3brain_regions.xlsx: Excel file of GO-term enrichment and GSEA for sex across the caudate nucleus, DLPFC, and hippocampus.

Data S5. BrainSeq_sex_WGCNA_results.tar.gz: Compressed tar file containing with WGCNA results for autosomal only and all genes sex networks including eigengenes, module membership text files, GO enrichment text and excel file results, and enrichment results with sex DEG analysis for the caudate nucleus, DLPFC, and hippocampus.

Data S6. BrainSeq_module_preservation_zsummary.xlsx: Excel file of Z summary for module preservation across brain regions for all individuals, control only, and schizophrenia only.

Data S7. differential_expression_region_interaction_sex_4features.txt: Text file of differential expression analysis for the interaction between brain region and sex for the four features (gene, transcript, exon, and exon-exon junction).

Data S8. BrainSeq_region_interaction_sex_functional_enrichment.xlsx: Excel file of GSEA for sex and region interaction pairwise enrichment.

Data S9. BrainSeq_male_biased_genes_XCI_status.tsv: Text file of male-biased (upregulated in male individuals) DEGs across the caudate nucleus, DLPFC, and hippocampus annotated for XCI status.

Data S10 differential_expression_interaction_jxn.txt: Text file of differentially expressed exon-exon junctions for interaction model of schizophrenia and sex across the caudate nucleus, DLPFC, and hippocampus.

Data S11. differential_expression_schizophrenia_by_sex_4features.txt.gz: Compressed text file of differential expression analysis for schizophrenia by sex across the caudate nucleus, DLPFC, and hippocampus for four features (gene, transcript, exon, and exon-exon junction).

Data S12. functional_enrichment_analysis_sexSpecificSZ_3brain_regions.txt: Functional enrichment of male-specific schizophrenia across three brain regions using stringent gene list (all DEGs, upregulated and downregulated in schizophrenia).

Data S13. BrainSeq_sexGenotypes_4features_3regions.txt.gz: Compressed text file of sex-interacting eQTL results across the caudate nucleus, DLPFC, and hippocampus for four features (gene, transcript, exon, and exon-exon junction) generated using mash modeling.

Data S14. BrainSeq_sex_interacting_eGene_functional_enrichment.xlsx: Excel file of functional enrichment of sex-interacting eQTL associated with unique genes for the caudate nucleus, DLPFC, and hippocampus.

Data S15. BrainSeq_siEQTL_public_comparison.xlsx: Excel file of si-eQTL shared with the public datasets^{19,91}.

Data S16. dapg_sex_interacting_eQTL_results.tar.gz: si-eQTL results after fine-mapping across the three brain regions for four features (gene, transcript, exon, and exon-exon junction).

Data S17. BrainSeq_colocalization_3regions.xlsx: Excel file of gene level colocalization between DAP-G fine-mapped sex-interacting eQTL and schizophrenia GWAS¹¹ on the signal level and individual SNP level using fastENLOC.