Supplementary Online Content

Periyasamy S, John S, Padmavati R, et al. Association of schizophrenia risk with disordered niacin metabolism in an Indian genome-wide association study. *JAMA Psychiatry.* Published online July 3, 2019. doi: 10.1001/jamapsychiatry.2019.1335

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Data Collection and Analysis

Study Sample

All participants gave written informed consent, and the study was approved by relevant institutional ethics committees at each participating institution.
Individuals and families were ascertained via The Schizophrenia Research Foundation, Chennai (SCARF) Medical Records, hospital wards and outpatient clinics, and clinical networks primarily in Chennai, and most recently extending to Tiruchirapalli, Coimbatore, Erode and Hyderabad.

(a) Family sample: The recruitment and diagnostic ascertainment of this sample have been previously described.[1,2] Briefly, recruitment was coordinated by the Schizophrenia Research Foundation (SCARF), Chennai, India. Within Tamil ancestry (defined linguistically and geographically), there was a primary focus on the Brahmin caste (Tamil, and others). Cases and families (Affected Sibling Pair, Trio) were exhaustively recruited from this caste, and then from other geographically proximate castes including Mudaliars and Dalits. Recruitment was conducted via SCARF medical records, inpatient unit, community mental health services, and outpatient clinics, together with a well-established recruitment network of clinicians (at hospitals, clinics in both public and private sectors), who had been previously involved in other SCARF research studies. These facilities included the Voluntary Health Services, Ramachandra Medical College and Hospital, and the Sundaram Medical Foundation.
(b) Case-control samples: The recruitment and diagnostic ascertainment of the cases was identical to the family sample; the controls were assessed using a very similar questionnaire to that used for the healthy individuals within the family sample.
(c) Diagnostic Ascertainment: Once informed consent was obtained and an individual was included in the study, diagnostic interviews were conducted using the Diagnostic Interview for Genetic Studies (DIGS). The research psychiatrists working on the study were able to administer this instrument in Tamil, where necessary, having previously organised its translation, with appropriate checking using standard back translation procedures. The Family Interview for Genetic Studies (FIGS) was also completed.
(d) Inclusion/Exclusion Criteria: (i) Inclusion Criteria for cases. All participants had to be capable of giving signed, informed consent for interview and blood sample collection. Participants had to be over 18 years old at interview. Affected subjects were defined as those whose DSM-IV diagnosis is schizophrenia or schizo-affective disorder based on DIGS, medical records and FIGS sources of information. Control subjects were defined as individuals with no personal or family history (to two degrees) of psychotic disorder, based on FIGS interview. To avoid north-south geographical impact on allele frequency variation, we included only individuals (cases and controls) whose families originated in southern India, based on available interview information.
(ii) Exclusion Criteria for Individuals. Subjects unable to give informed consent to all aspects of the study were excluded, as were those whose psychosis was judged to be secondary to substance use or a known neurological disorder such as epilepsy, based on the consensus diagnostic procedure. Subjects with severe mental retardation (MR) were also excluded, while those with mild MR (IQ>55 or based on clinical and educational history) were included. Subjects with more than 'occasional use' or 'previous use' of alcohol or cannabis (no other illicit drugs) were also excluded.
(iii) Inclusion Criteria for controls. Based on selected recruiting strategies, individuals from the cases' defined geographic region were approached for participation. The FIGS interview plus an additional interview developed for this study, defined "unaffected" status and caste membership.
(e) Best Estimate Diagnosis: In addition to the FIGS and DIGS interviews, comprehensive narrative reports of all historical and interview material were prepared for diagnostic review. Each case was independently reviewed by two experienced research psychiatrists (RP, ST, RT). Any outstanding questions about the case were then resolved (e.g. by the collection of further clinical data) and a final consensus diagnosis was achieved. Ten percent of cases were randomly chosen for review by BJM.

Genotyping and quality control

The study sample was recruited and genotyped in consecutive phases over 15 years. The family sample (n=658) was genotyped using the Illumina CNV370 Beadchip array; unrelated controls (n=199) were recruited early in the study and genotyped on Illumina OmniExpress-12 arrays; subsequently recruited cases and controls were genotyped in two batches (wave 1=1370; wave 2=1008) using the Illumina Infinium PsychArray BeadChip (PsychChip) . Note that the numbers of samples mentioned here reflect pre-QC totals.

MDS analysis

A multi-dimensional scaling (MDS) analysis was performed to test for array effects and whether the Indian samples were genetically similar to the South Asian (SAS) population of 1000 Genomes (1KG) phase 3 reference data (eFigures 1-3). This latest version of 1KG includes the following SAS samples of relevance to our study cohort: Bengali in Bangladesh (BEB); Indian Telugu, UK (ITU); Punjabi in Lahore, Pakistan (PJL); and Sri Lankan Tamil in the UK (STU). (https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html). The MDS analysis was based on all SNPs in common between the Indian family and case-control data sets and the 1000 Genomes data set

(2504 samples; 26231 SNPs after pruning). Strand ambiguous SNPs were removed to avoid strand mismatches. eFigures 1-3 depict a graphical representation of the Indian Tamil samples relative to all the SAS data. In the first dimension (eFigure 1), all four of our component data sets are isolated from the rest of the super population (American, African, European, East Asian) and overlap with SAS population. This overlap is seen in more detail in the second dimension (eFigure 2). The lack of outliers confirms that our samples are of Indian ancestry.

Genotype imputation

Pre-Imputation QC

To check and identify genetically related individuals we generated a set of 26939 LD independent SNPs with minor allele frequency > 0.05 that were common across all genotyping platforms. For each of our four data sets, we identified sample duplications and mis-specified relationships in family samples through IBD analysis and cross-checking with the clinical records. For case-control data sets we excluded related individuals with PI-HAT >= 0.1875 (i.e. individuals beyond 2nd degree relatedness). We then merged the four data sets to identify inter-data set discrepancies such as mis-specified relatedness both within and between families. One hundred and forty-three individuals were removed. Before imputation, each cleaned data set was filtered with the following parameters: --geno 0.02 --maf 0.01 --hwe 0.001 --mind 0.1

Imputation

We imputed each of our four data sets to 1000 Genomes (1KG) phase 3 (see description above). We updated SNP coordinates to human genome build 37 and flipped to the positive strand. Phasing was performed using SHAPEIT v2.72720. Imputation was then conducted using IMPUTE v2.3.021, with each chromosome being split into 5Mb segments with 250kb overlap, and using all 1KG populations (~2,504 individuals) as the reference (as recommended by the authors). Output was converted to 'best guess' genotypes in binary PLINK format.

Post-imputation QC

A total of 81,177,102 imputed SNPs was reduced to ~6.2 - 6.5 million SNPs by removing SNPs with an INFO score <0.8, MAF< 0.05, missing data rate >0.05, or HWE P value $<1 \times 10^{-6}$. The four data sets were merged before further QC was conducted according to Psychiatric Genomics Consortium guidelines (https://sites.google.com/a/broadinstitute.org/psych-chip-resources/qc-methods) by removing SNPs with: (a) a call rate difference between cases and controls with P value $< 1 \times 10^{-5}$; (b) a HWE P-value $< 1 \times 10^{-6}$ in controls; and (c) a HWE P-value $< 1 \times 10^{-10}$ in cases. Finally, as recommended by GCTA, SNPs with a missing data rate >0.005 were removed.

A total of 5,582,932 variants and 3092 individuals (1321 cases; 1771 controls) passed all filters and QC.

SNP-based heritability estimate

Using Genome-wide Complex Trait Analysis (GCTA v1.24.722[3]), we conducted GREML analysis to quantify proportion of variance attributed to genome-wide SNPs or SNP-based heritability. We estimated the genetic relationship matrix (GRM) between all pairs of individuals using data on 5,582,932 autosomal SNPs (by first estimating GRMs using SNPs on each of the autosomes and then merging to create a single GRM). Individuals with relatedness ≥ 0.05 (~3rd cousins or closer) were removed (gcta --grm --grm-cutoff 0.05), leaving 2210 individuals (945 cases, 1265 controls). We estimated the variance explained by all autosomal SNPs (gcta --reml), fitting 20 PCs estimated from the data, sex and SNP array as covariates. Lifetime risk was assumed to be 0.01 for conversion of SNP-based heritability from the observed to the liability scale.

We performed additional analyses in which the genetic variance was partitioned by chromosome: one in which the GRMs for the 22 autosomes were fitted simultaneously in a single model (GCTA's --reml-no-constrain --mgrm-bin), and another in which the GRM for each chromosome was analysed separately (GCTA --reml-no-constrain --grm). If the variance explained by the 22 individually estimated variances was greater than that for the 22 simultaneously estimated variances then this would be evidence for stratification in the data[4] (eFigure 4 in the Supplement).

Genome-wide association analyses

The data set consisting of 3092 samples was split into two data sets for association analysis. The first was a case -control sample of unrelated individuals (Ncase=816; Ncontrol=900). The second comprised the family samples (Ncase=505, Ncontrol=871, Nfamily=405). For both samples we used GCTA's –mlma-loco (leaving-one-chromosome-out) function which conducts association analysis for each SNP in turn fitting a GRM based on all chromosomes but leaving out the chromosome on which the

candidate SNP is located. Ten principal components (PCs) were utilised to control for population/family stratification/cryptic relatedness calculated using PC-AiR,[5] . Genotyping chip-type was fitted as a covariate to control for batch effects. The genomic inflation factor, *lambda*, was calculated individually for each data set. GWAS test statistics were corrected for inflation before meta-analysis with GWAS test statistics from thecase-control samples using METAL (2011-03-25).[6]

Chromosome X analysis

Chromosome X imputation was conducted for subjects passing quality control for the autosomal analysis. A unified data set was generated for subsequent analyses that include all SNPs and individuals passing the above filtering criteria in both the male and female QC groups. To generate a unified data set we applied additional exclusions of chromosomal X SNPs with missingness ≥ 0.05 or HWE $P < 10^{-6}$ in females. To consider differences in genotyping between hemizygous males and diploid females, we used XWAS[7] to apply all the aforementioned QC steps of samples separately for males and females.

X-specific QC steps were performed using XWAS 2.0[7] on the unified data set. These included: removing SNPs with significantly different MAF between male and female samples in the controls; removing SNPs with significantly different missingness rates between male and female controls; and the removal of SNPs in the pseudoautosomal regions (PARs). Analysis was performed using GCTA separately for related males and females and unrelated males and females, and they were meta-analysed using METAL.

Verification of Indian top locus

Imputation batch-based meta-analysis

As a sensitivity analysis, to check that the significant association in our top locus is not an artefact of batch effects, we split the 3092 samples based on array types in which they were genotyped and imputed. Of the 3092 samples, 640 were genotyped using Illumina CNV370, 1361 were genotyped in PsychChip Wave 1 and 895 were genotyped in PsychChip Wave 2. These samples were analysed separately using GCTA mlma-loco. Samples from PsychChip Wave 2 were split into family (N=625) and unrelated (N=270) samples since the GWAS results were deflated. In this sensitivity analysis, 196 control samples genotyped as a stand-alone batch using Illumina OmniExpress-12 arrays were excluded. The meta-analysis for the included data sets (minus the 196 control samples) were performed using METAL.

Post-imputation batch-associated QC

The samples were genotyped in 4 batches over 10 years. The samples were imputed in 4 batches at the end of the 10 years. Since this process could lead to batch specific artefacts during genotyping and imputation, we performed four GWASs by setting each batch in turn as "case" and the remaining batches as "controls" to identify SNPs associated with a particular batch. To avoid confounding by case status, each of the four GWASs was performed using only the controls. SNPs with a p-value $< 5 \times 10^{-8}$ across 4 GWASs were removed, since each such result reflected a batch effect, in line with recently published protocols. [8]

Genomic profile risk scoring (PRS) analyses

Many SNPs are associated with disease at a level that does not reach genome-wide significance. Capturing their contribution can deepen our understanding of schizophrenia's polygenic architecture. We conducted PRS[9] using PGC2 meta-analyses results as the "discovery" sample and our Indian summary statistics as the independent "target" sample to determine whether European PRS could predict schizophrenia in our Indian sample. Results are presented using both the observed scale (0 or 1 i.e. binary trait) (eFigure 11) and the liability scale (included in eResults), adjusted for potential ascertainment bias including disease prevalence, and interpretable across studies.

The PGC2 schizophrenia association analysis was based on the "discovery" sample of 35,476 cases and 46,839 controls, affording the highest precision to date to the individual SNP estimates and thus good power in the profile scoring analyses.[10] In order to obtain a highly informative SNP set with the least statistical noise, we excluded uncommon SNPs (MAF < 5%), low-quality variants (imputation INFO < 0.9), indels, and SNPs in the extended MHC region (chr6:25-34 Mb).
Since we used a summary statistics-based analysis in PRSice v1.25,[11] the following stringent clumping parameters were used as per gtx (Genetics ToolboX)(R package) recommendation.
--clump-p1 0.5 --clump-p2 0.5 --clump-r2 0.05 --clump-kb 300

Only SNPs that were in common between Indian samples and the PGC2 schizophrenia association analyses were included in the profile scoring (953849 SNPs overlapping; 32176 SNPs after clumping).

We performed high resolution PRS of our target summary data to identify the best fit P_T and to obtain results calculated at broad p-value thresholds ($5x10^{-8}$, $1x10^{-6}$, $1x10^{-5}$, $1x10^{-4}$, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5).
Individual profile scores in the "target" samples were calculated as the sum of the count of the risk alleles weighted by the log of the odds ratio from the PGC2 association analysis, divided by the total number of SNPs.

These results reflect the variance explained on the observed probability scale,[12] quantified as Nagelkerke's R2 ($NK\text{-}R^2$) and transformed to the liability scale[4] assuming a baseline population risk of 1%. We examined whether the observed fraction of results at $P_T = 0.05$ displaying the same direction of allelic effects across PGC2 and India was significantly greater than expected by chance (i.e. 50%) using the binomial Sign Test. eTable 9 shows the total number of SNPs, the SNPs sharing the same direction of effects, and their respective sign test for all the SNPs included below the p-value threshold ($P_T = 0.05$) and MAF ranges <0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.5.

We used sign tests to compare the patterns of results between our Indian GWAS results and those of the PGC2. We used the clumping settings above to derive a filtered set of SNPs. We then determined the number of SNPs whose logistic regression beta coefficient signs were the same between two independent samples. The significance of the observed proportion differing from chance (50%) was evaluated using the binomial distribution.

Post-GWAS extended bioinformatics analyses

Statistical Fine mapping

PAINTOR (ver 3.1)[13,14] was used for statistical fine mapping analysis by integrating functional genomic data (Transcription Factor Binding Sites; Methylation sites in brain regions - Hippocampus, Cingulate, Frontal Lobe) in the top Indian locus. We used our Indian data set as the LD reference data set.

Natural Selection

Using the Geography of Genetic Variants browser[15] we identified worldwide patterns of allele frequency distribution for our top SNP. We then used the R package Rehh 2.0.2 to detect signatures of selection using Extended Homozygosity Hapelotype(EHH) based test. We used Rsb (i.e. the observed of standardised log-ratio of the integrated EHHS(iES) between pairs of populations) to detect signals of positive selection.

Transethnic Meta-analysis

We transformed the MLMA-loco effect sizes generated in our Indian data set using a recently published method.[16] The transformed odds ratios were utilised for transethnic meta-analysis with the PGC2 data set.

Regional plots were generated using LOCUSZOOM (http://locuszoom.org/genform.php?type=yourdata) to investigate our top locus with 100Kb and 1000Kb windows, and a threshold of $r^2>0.6$ to define the LD block. To investigate the regulatory effects of our top SNP, we interrogated the expression quantitative trait loci (eQTL) databases, Brain eQTL Almanac (BRAINEAC)[17] (http://braineac.org/), Genotype-Tissue Expression (GTEx v7)[18] (https://www.gtexportal.org/) and the Common Mind Consortium (CMC)[19](https://www.synapse.org/#!Synapse:syn2759792/wiki/). We also utilized the Human Brain Transcriptome (HBT)[20](http://hbatlas.org/) to examine spatio-temporal patterns of genes and transcription factors located in our top locus.

Trans-ethnic genetic correlation for SCZ between India and Europe

We used POPCORN[21] to estimate trans-ethnic genetic correlation between PGC2 and Indian Summary Statistics; there was no significant finding because the sample size of our Indian data set was comparatively small (n=3092).

SMR analysis of top Indian locus

SMR (Summary-data-based Mendelian Randomization) ver. 0.712 and Brain-eMeta eQTL summary data (n =1,194 samples) were used for SMR single/multi SNP based methods and HEIDI (HEterogeneity In Dependent Instruments) analysis.[22] This analysis was conducted to identify genes within our top locus whose expression levels were associated with schizophrenia due to pleiotropy in our Indian data set. We used our top SNP and genes in our top locus to interrogate the Brain-eMeta eQTL data (n =1,194 samples).[23]

Pathway analyses

Gene-set based analyses were performed using MAGMA v1.06.[24] SNPs were mapped to genes according to their position in the NCBI 37.3 build and with a 35 kb 5′ and 10 kb 3′ window around each gene. Gene-sets from REACTOME-MSigDB version 6.0 were used for pathway analysis. For genes, the summary statistics from the imputed GWAS were used to derive gene-based statistics using the South Asian 1000 genomes reference panel (SAS phase 3) to model linkage disequilibrium. MAGMA tests for gene set enrichment by first generating a gene-wide statistic from the GWAS results files, adjusting for gene size, SNP density and linkage disequilibrium effects (eFigure 18). MAGMA then performs a competitive test of gene set association to investigate if associative enrichment within a user-defined gene set outperforms other gene sets from across the genome of similar size.[24] SNP-wise=top was used as the gene analysis model since a smaller percentage of SNPs had p-values approximating the GWAS significance threshold. The analysis accounted for synonymous SNP IDs and an adaptive permutation procedure was used to account for varying gene p-values since a large number of permutations only has added value if the p-value is very low.

Network connectivity-enrichment analysis

We used MAGNUM[25] to evaluate whether genes perturbed by schizophrenia-associated variants are enriched for tissue-specific gene regulation. The steps involved aggregating GWAS summary statistics at the level of genes using gene-based analysis; we used MAGMA to generate the gene-based statistics. We then defined the proximity of genes within the network, computing and summarizing connectivity-enrichment curves, then computing empirical p-values and finally defining a connectivity enrichment score for our GWAS summary statistics as the negative $\log_{10}$ of the empirical p-value. We interrogated 32 higher level regulatory networks from FANTOM 5, GTEx expression networks and brain tissue specific regulatory networks to identify enriched networks perturbed by schizophrenia associated variants from our Indian summary statistics.

Post-GWAS functional analyses

Cell line assays

To validate the imputed genotypes for our top locus, we sequenced the region between *MROH6* and *NAPRT1*. DNA haplotypes in the sample population were extracted from genotype data. The four haplotypes of interest, representing 98.7% of the alleles in the Chennai Indian population, were PCR amplified from genomic DNA from three individuals. (eTable 5).

To assess the effect of allele-specific expression, we measured expression of genes at the rs10866912 locus in lymphoblastoid cell lines established from 60 individuals within the study population. We randomly selected cell lines from 20 individuals from the population for each of the AA, CA, and CC rs10866912 genotypes and measured expression of genes within our locus defined by r^2>0.6: *NAPRT1, EEF1D, PYCRL, TIGD5, GSDMD,* and *MROH6*.

Establishment of lymphoblastoid cell lines

Lymphocytes were extracted from fresh, whole blood samples using a Ficoll gradient, and transformed with Epstein Barr Virus (EBV). Briefly, lymphocytes were incubated with a virus preparation from the EBV-producing marmoset B-cell line B95-8 and treated with phytohaemagglutinin (Sigma). Cells were plated out in 96-well plates and incubated at 37° in 95% air/5% CO2 for up to 5 weeks with twice weekly media changes until transformed lymphoblastoid cells became apparent.

RNA expression analysis

Total RNA was extracted from lymphoblastoid cell lines using the RNeasy mini kit (Qiagen). One µg of RNA from each cell line was reverse transcribed using the SuperScript VILO cDNA Synthesis Kit (ThermoFisher). cDNA levels were measured in triplicate using exon-spanning Taqman probes, normalised to expression of GAPDH using the ΔΔCt method. Expression results for each genotype were pooled.

Zebrafish maintenance and transgenic lines

Adult Zebrafish and embryos were maintained by standard protocols approved by the University of Queensland Animal Ethics Committee. Ethics committee approval IMB/266/15/NHMRC.

To select target sequences on the *dre-naprt1* mRNA, we first used the Invitrogen BLOCK-IT RNAi Designer (miR) using the full 3'UTR of *dre-naprt1* (ENSDART00000038157.5). Among the different hits obtained, we ran BLAST to select synthetic miRNA

that did not show off-target on the zebrafish genome. Three synthetic miRNAs were selected based on their BLOCK-IT score, no off-target and no overlap on the 3'UTR of dre-naprt1 mRNA. Forward and reverse primers generated are presented in eTable 4.

Generation of dre-naprt1 loss-of-function transgene

Each forward and reverse primers oligonucleotides (artificial pri-miR, eTable 4 were first annealed. 5µl of 200µM of each forward and reverse primer were mixed in a 20µl reaction including 2µl NEB10x buffer2. The mixture was heated to 95°C for 5min using a thermocycler and left to cool down in the thermocycler's plate for 30 min. Annealed artificial pri-miR was mixed and briefly spun down before being diluted 5,000-fold in water at room temperature. Diluted samples were stocked at room temperature until ligation. The pME-RNAi652 plasmid[26] was digested with BsmBI enzyme and gel-extracted/purified for inserting the annealed primers (artificial pri-miR), 10 ng of linearized pME-RNAi652 construct was mixed with 4µl of the aforementioned 1:5,000-dilution and ligated following the manufacturer's instruction (NEB T4 DNA ligase). We obtained 3 different plasmids that were sequenced, one for each artificial pri-miR, named pME-R-NAPRT1-1, pME-R-NAPRT1-2 and pME-R-NAPRT1-3. We then chained the different miRNAs as presented previously.[27] We finally generated a plasmid carrying the 3 synthetic anti-NAPRT1 miRNAs in the following order: 1-2-3, named pME-R-NAPRT1-123. To generate a final plasmid for integration into the zebrafish genome, we proceeded to a gateway LR reaction with the following plasmids, p5E-UBIQUTUIN, pME-R-NAPRT1-123, and destination tol2 plasmid 1456[26]. The final plasmid was named UBI-R-NAPRT1-123 and drives i) the ubiquitous co-expression of dsRED (red fluorescence) with the synthetic anti-NAPRT1 miRNA, ii) the tissue-specific lens expression of GFP for assisting in the selection of transgenic animals that will have integrated the transgene. For the control transgene, we follow the same methods using an empty pME-RNAi.

Loss-of-function transgene injections and generation of F0 fish

To integrate DNA constructs into the zebrafish genome, 1 to 2nl of a mix containing 30ng/µl of DNA of interest plus 25ng/µl of Transposase mRNA and phenol red were injected into one-cell stage wt zebrafish embryos (F0 fish).

F1 transgenic line selection

F0 injected fish were selected at 72 hours-post-fertilization (hpf) based on both strong mosaic ubiquitous dsRED expression and GFP expression in their lens. Sixty fish were raised until adulthood. Individual F0 adult were then outcrossed with wt (AB) fish for selecting the founder giving birth to F1 transgenic animals who have integrated the construct into their genome. Seven lines have been isolated and the F0 fish giving birth to animals with the strongest dsRED expression was selected for the analysis and for establishing the UBI-R-NAPRT1-123 line.

dre-naprt1 mRNA injections

To synthetize *dre-naprt1* mRNA for rescue experiments, we first amplified *dre-naprt1* cDNA using primers dreNAPRT1-Forward and dreNAPRT1-Reverse flanked by compatible gateway BP sequences (eTable 4). The corresponding PCR product was then inserted into pDONR221 through a gateway BP reaction following manufacturer instructions. The resulting plasmid was named pME-NAPRT1. We then generated a final construct for RNA synthesis through a gateway LR reaction mixing pME-NAPRT1, p3E-polyA (tol2kit 302 plasmid) and a custom R1R3 gateway pDEST presenting a T3 promoter. The final construct was named T3_NAPRT1_PA. *dre-naprt1* mRNA was then synthetized using a mMESSAGE mMACHINE T3 transcription kit following manufacturer instructions, and purified using a MEGAclear kit following manufacturer instructions. *dre-naprt1* mRNA was injected at 150pg into one-cell stage zebrafish embryos.

eResults. Findings

SNP heritability

SNP heritability on the liability scale was estimated to be 0.287 (SE: 0.073), assuming a disease prevalence of 0.01 ($P = 3.61 \times 10^{-5}$). This estimate is higher than the estimate for schizophrenia in Europeans 0.23 (SE: 0.01), reported by Lee *et al*.,[4] but comparable to the estimates for individual cohorts in their study.
The variance explained by all chromosomes fitted independently (0.309) was only slightly higher than that explained by all chromosomes fitted simultaneously (0.296), suggesting little evidence for stratification.

Verification of Indian top locus

Imputation batch-based meta-analysis

The meta-analysis q-q plot showed a mild deflation (lambda= 0.98) showing no evidence for residual populations stratification, but rather marginal evidence for overfitting/overcorrection conservative results (see eTable 4; eFigure 6). SNP rs10866912 was the most associated locus as reported for the main analysis. The magnitude and direction of the effect size was unchanged (OR 1.27), but as expected, given the smaller sample size, the association P-value was slightly increased (P-value= 3.53×10^{-7}).

Post-imputation batch-associated QC

Of 5.5 million SNPs, 228 SNPs were removed for being observed to have genome-wide significant batch associations ($P < 5 \times 10^{-8}$). The most significant SNP in this list of 228 SNPs had a p-value of 3.66×10^{-3} in the Indian GWAS. None of the SNPs in our top Indian locus were in this list, confirming that the batch effect was not confounding this locus. We have regenerated the Manhattan plot excluding these SNPs.

SMR analysis

Since the eQTL data were predominantly European and our GWAS data were of Indian ancestry, there were significant haplotype frequency differences between these data sets: in particular, the frequency of the haplotype containing allele "C" of our top SNP, rs10866912 was ~0.26 in Indians compared with ~0.47 in Europeans. For rs10866912, only four genes in the eQTL data set surpassed the eQTL threshold ($P < 5 \times 10^{-8}$); of these, *NAPRT1* had the most significant SMR p-value ($P = 2.63 \times 10^{-07}$). The SNPs in high LD with rs10866912 (rs10866911, rs4873803, rs4873804) also gave the most significant results with *NAPRT1*.

Trans-ethnic replication of PGC2 loci in Indian data set

We analysed all (108) genome-wide significant PGC2 loci in our Indian data set. We mention here:

(a) the top PGC2 locus (MHC region on chromosome 6p22): Using LDlink[28] we identified proxy SNPs for the top PGC2 SNP (rs115329265-Old ID, rs1233578 –New ID according to dbSNP142) in this region, because this SNP did not survive QC in our Indian data set. For this SNP, there were four SNPs in LD > 0.9 (~28Kb region) and 302 SNPs in LD > 0.5 and <0.9 (~995Kb region) in the 1000 Genomes CEU population. In contrast, for the 1000g STU population there were 13 SNPs in LD >0.9 (~48Kb region) and 58 SNPs in LD > 0.5 and <0.9 (~937Kb region). For this 48Kb region defined by LD >0.9, we observed no evidence for replication (p<0.05) in Indian samples (eFigure 20). Possible explanations for this finding include: (i) relatively small Indian sample size reducing power to detect the variation; (ii) trans-ancestry allele frequency differences between Europeans and Indians; (iii) pathogen-driven selection of the risk allele in Europeans;[29] (iv) a pure schizophrenia phenotype (few cases of schizoafftective disorder) with negligible substance abuse in the Indian data set compared with PGC2 schizophrenia which is attended by significant co-morbid substance abuse.

(b) For all other PGC2 genome-wide significant loci, 96/107 regions had representative SNPs within the STU LD block in our Indian data set. Of these, 31 PGC2 regions replicated (*P*-value < 0.05) in our Indian data set, with the chromosome 2q32.3 locus (rs59979824; ranked 66 in PGC2) having the best p-value (0.0007298) in our data set. (eTable 12 and eFigure 21)

Genomic profile risk scoring (PRS)

European PGC2 SNPs achieving a threshold p-value ($P_T = 0.05$) were most predictive of schizophrenia in India, explaining 5% (maximum Nagelkerke's $R^2 = 0.05$; $P = 2.2 \times 10^{-37}$) of the genetic variation when measured on the observed scale (zero or 1, i.e. binary trait), and 3% on the liability scale. The majority (53%) of SNPs below $P_T = 0.05$ had the same direction of (genetic) effect

between PGC2 and India (53.48%, binomial $P = 1.31 \times 10^{-15}$) (eTable 9), confirming that this result is highly unlikely to reflect chance.

Post-GWAS extended Bioinformatics analyses

Gene-set analyses

There were no Reactome gene-sets (including NAD^+ specific pathways) significantly enriched for association based on Bonferroni correction ($0.05/630=7.9 \times 10^{-05}$). However, the top pathways with p-value < 0.01 were DSCAM interactions, Dopamine Neurotransmitter Release Cycle, Unblocking of NMDA Receptor Glutamate Binding and activation, and signal transduction by L1. These pathways are associated with brain development. The *DSCAM* (DS Cell Adhesion Molecule) gene is a member of the immunoglobulin superfamily of cell adhesion molecules involved in human central and peripheral nervous system development and mediates axon pathfinding. Among the other leading pathways, L1 plays an adhesive role in cell-cell interaction and also functions as a signal transducing receptor assisting neurons in axonal growth and guidance (http://www.reactome.org/content/detail/R-HSA-445144). Collectively these pathway-based findings are consistent with the neurodevelopmental hypothesis of schizophrenia.[30]

Network connectivity-enrichment analysis

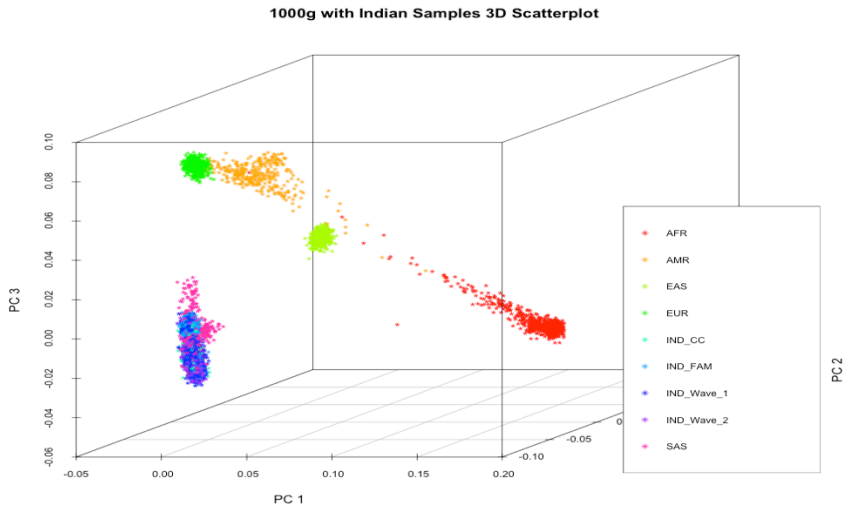
Of the 32 FANTOM 5 higher-level gene regulatory networks in humans, mesenchymal stem smooth muscle cells, and nervous system and hind brain networks generated the highest scores for our Indian summary statistics. When only brain regions were analysed, the putamen generated the highest enrichment score, similar to that observed in the PGC2 schizophrenia data set. Moreover, consistent with these results, the top four networks in humans identified by GTEx co-expression network analyses of our Indian data set were located in the brain (eFigure 19).

Post-GWAS functional analyses

mRNA expression associated with the rs10866912 locus

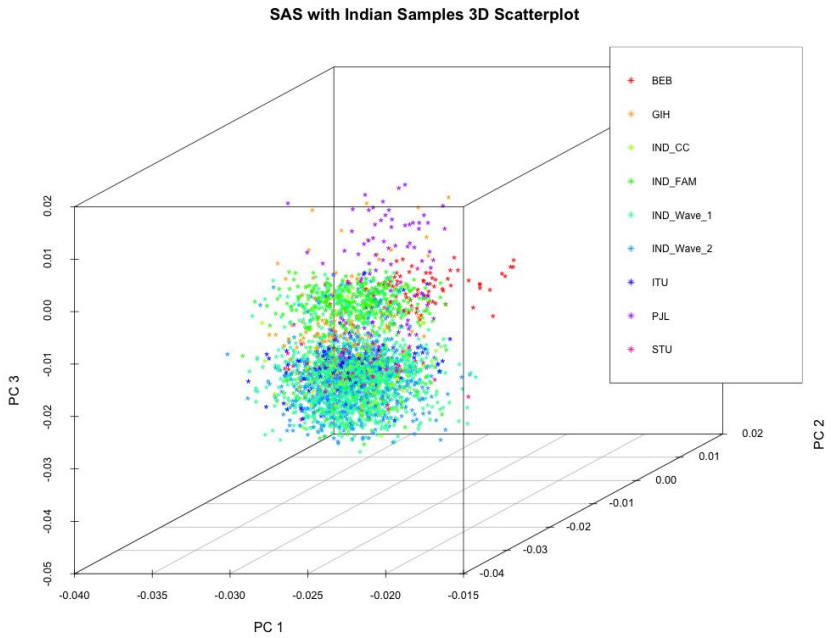
Our index SNP is located in an intergenic region spanning ~2Kb between *MROH6* and *NAPRT1* genes. Since our top locus was imputed, we sought to validate this sequence (2Kb) experimentally. Genotyping data indicated the presence of 4 distinct haplotypes in the 2Kb of genomic DNA between the *MROH6* and *NAPRT1* genes on chromosome 8, covering 98.8% of the alleles present in the study population. Haplotypes differed at 6 SNPs and one indel (eTable 5). The risk allele (A) of rs19866912 was present in haplotypes A, B, and C, with a combined frequency of 72% while the protective allele (C) of our index SNP was present in Haplotype D with a frequency of 26.8%. These haplotypes confirm that the genotyped alleles (A, C) and the alleles predicted by imputation (A, C) are identical for our index SNP.

eFigure 1. MDS Analysis Using 1000g SAS Population



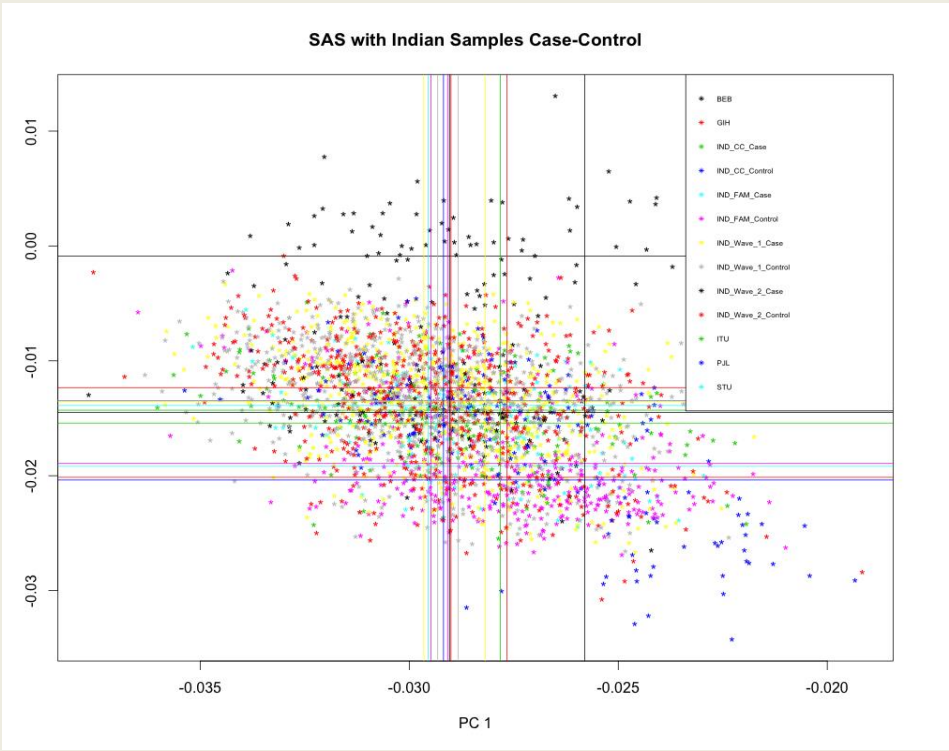
Super populations include African (AFR), Admixed American (AMR), European (EUR), East Asian (EAS) and South Asian (SAS). The four Indian data sets include IND_CC, IND_FAM, IND_Wave1 and IND_Wave_2. The figure shows the 4 Indian data sets clustering with the SAS population in the bottom left corner.

eFigure 2. MDS Analysis Using 1000g SAS Populations and Our Indian Data Set



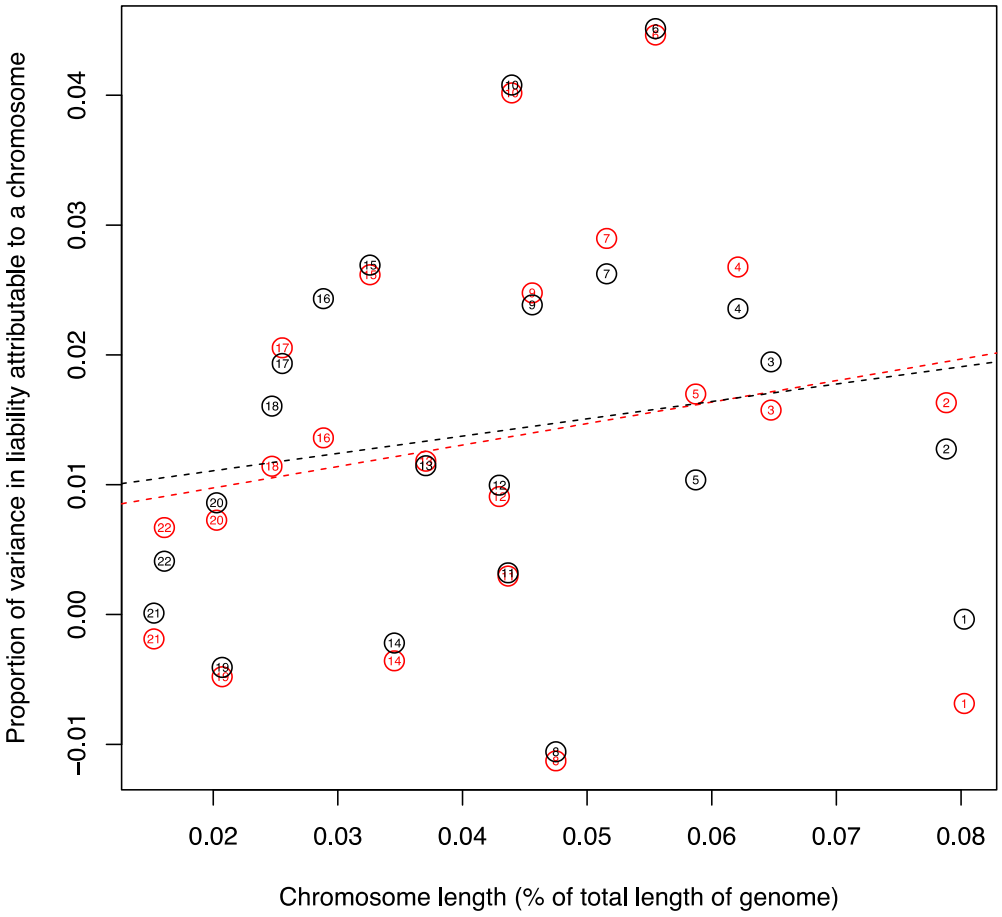
Populations of SAS include Bengali from Bangladesh (BEB), Gujarati Indians from Houston, Texas (GIH), Indian Telugu from the UK (ITU), Punjabis from Lahore, Pakistan (PJL) and Sri Lankan Tamils from the UK (STU). The four Indian data sets (IND_CC, IND_FAM, IND_Wave1 and IND_Wave_2) were genotyped using three different Illumina arrays and cluster accordingly in this figure.

eFigure 3. MDS Analysis Using 1000g Super Populations



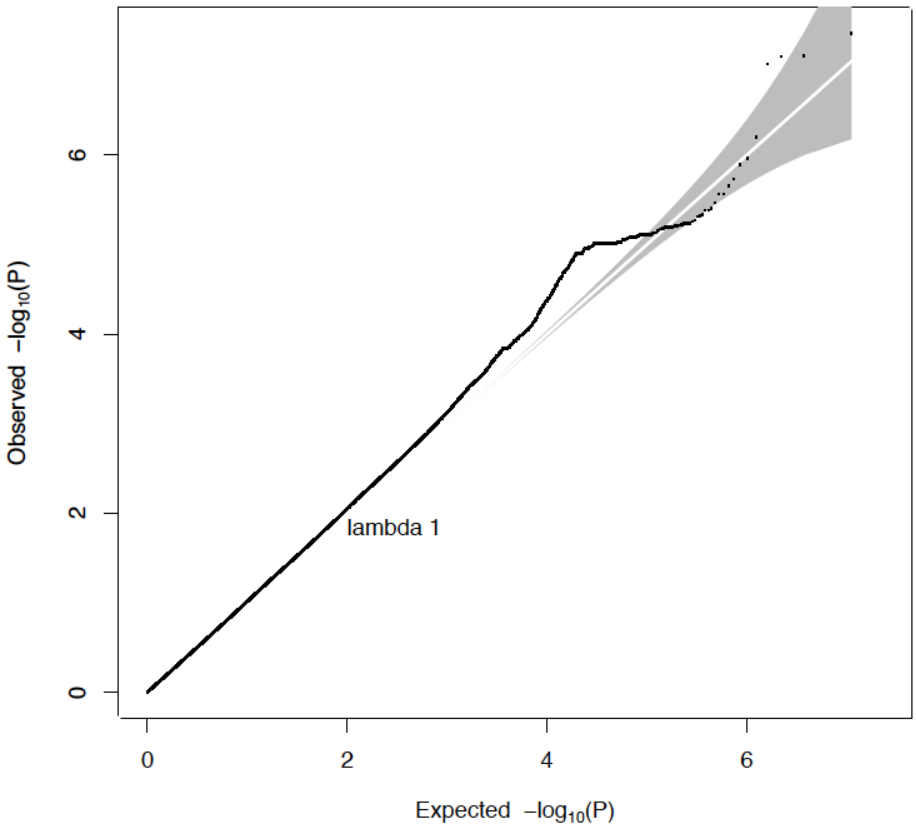
Each of the four Indian data sets is divided into cases and controls. Based on the Principal Component means, there is no case-specific bias.

eFigure 4. SNP Heritability



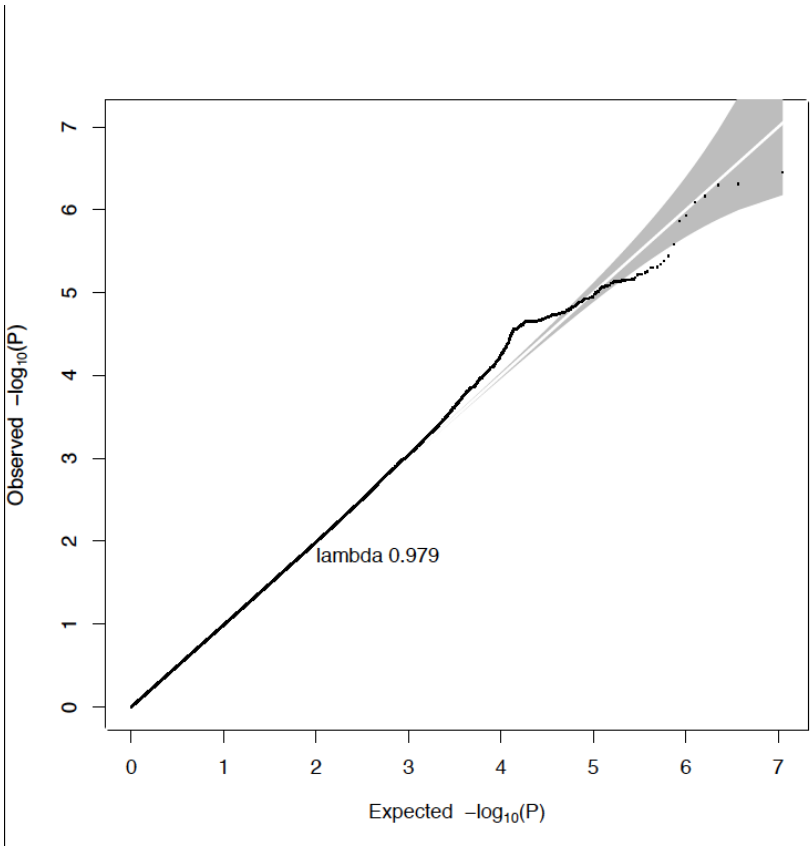
Estimated proportion of variance explained by SNPs on individual chromosomes from analysis of all chromosomes simultaneously (red) or separate analyses of each chromosome (black).

eFigure 5. Q-Q Plot for Indian Meta-analysis



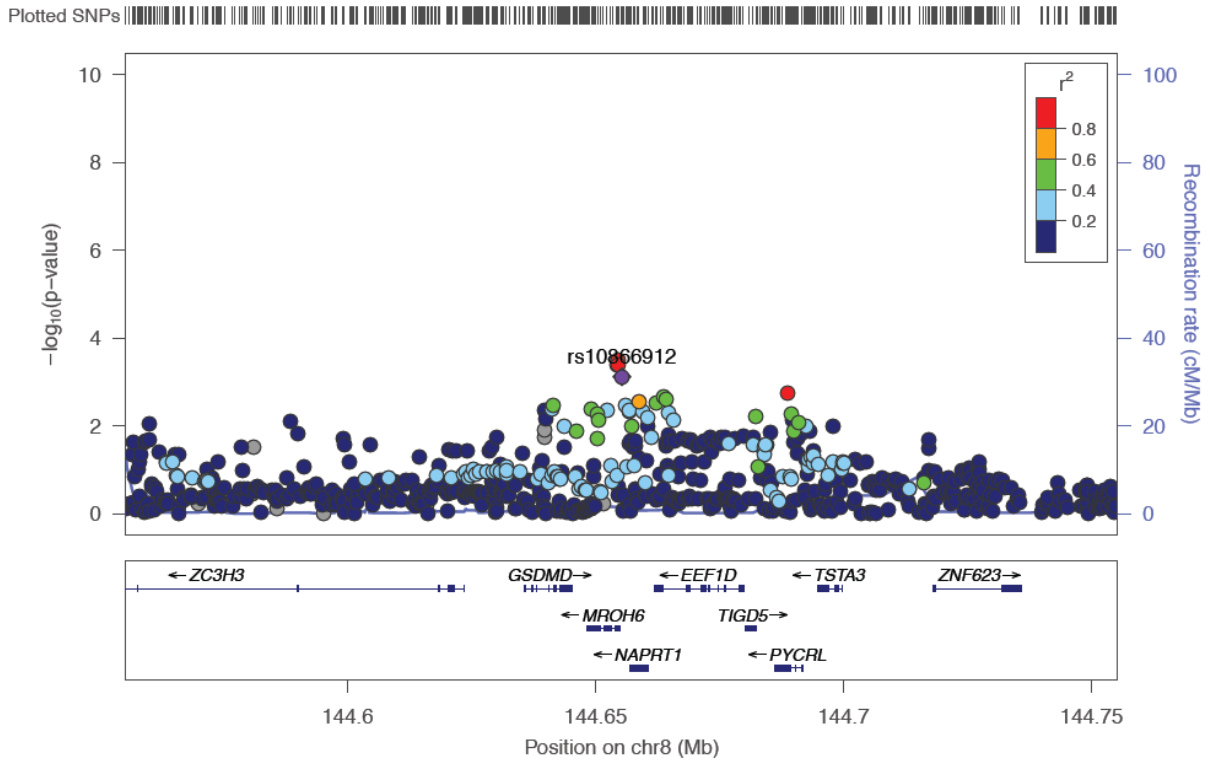
A meta-analysis of Indian case-control and family data sets. The observed *P*-values (y-axis) were compared with the expected p-values under the null distribution (x-axis). The plot was generated by genomic correction for family data which was then meta-analysed with case-control data. The resultant $\lambda_{GC} = 1$.

eFigure 6. Q-Q Plot for Indian Batch–Based Meta-analysis



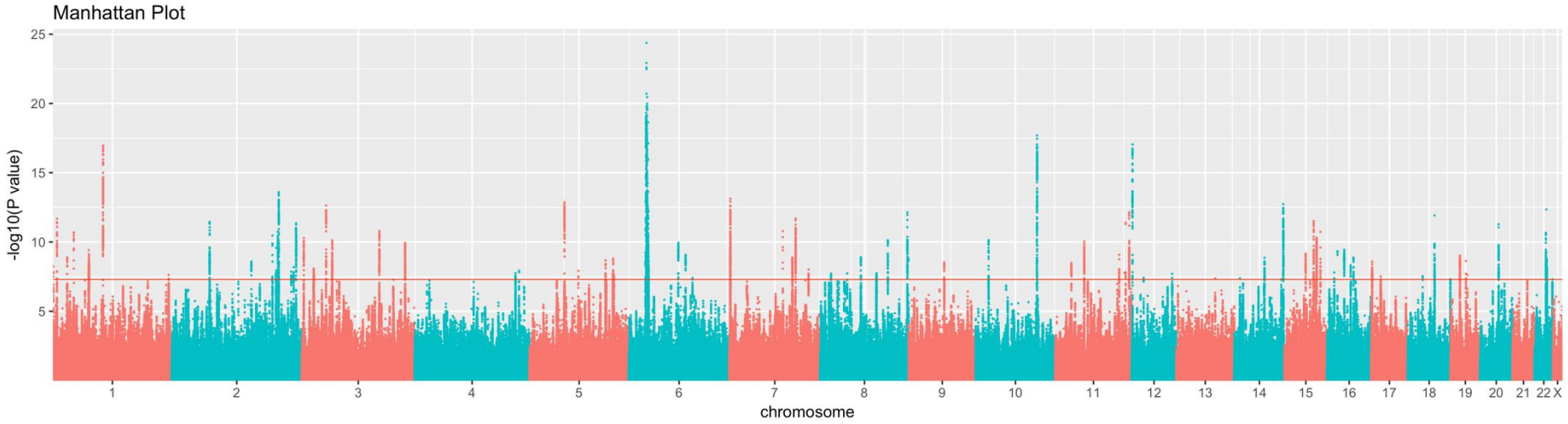
A meta-analysis of Indian CNV370, PsychChip Wave 1, PsychChip Wave 2 Family and PsychChip Wave 2 unrelated data set. The observed *P*-values (y-axis) were compared with the expected p-values under the null distribution (x-axis). The plot was generated by genomic correction for family data which was then meta-analysed with case-control data. The resultant λ_{GC} ~ 0.98.

eFigure 7. Regional Plot of Indian Chromosome 8q24.3 Locus (100-Kb Window) in PGC2 Data Set



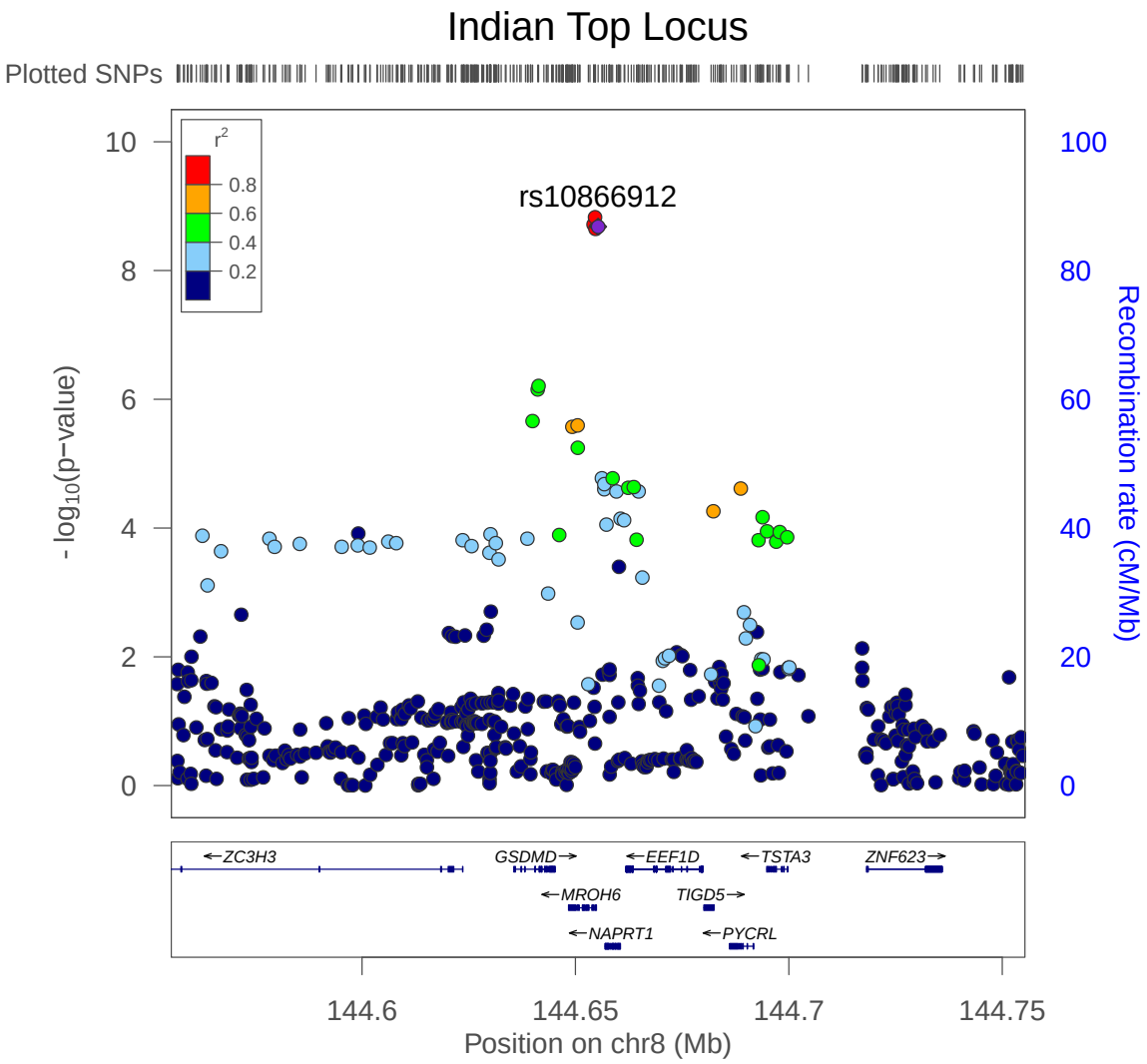
The index SNP, rs10866912 (coloured purple) was replicated (P = 7.56 ×10^{-4}) in PGC2. The other SNPs are coloured according to the degree of linkage disequilibrium (measured by r^2) with the index SNP. The Thousand Genome Phase 3 EUR population was used to calculate LD. The x-axis shows the SNP locus position on chromosome 8 (GRCh37/hg19 build). The y-axis shows the significance of association (-$\log_{10}P$). Nine genes are located within the 100Kb window, the direction of transcription (upstream/downstream) being annotated with arrows.

eFigure 8. Manhattan Plot for PGC2-Indian Meta-analysis Using Metasoft-RE2C



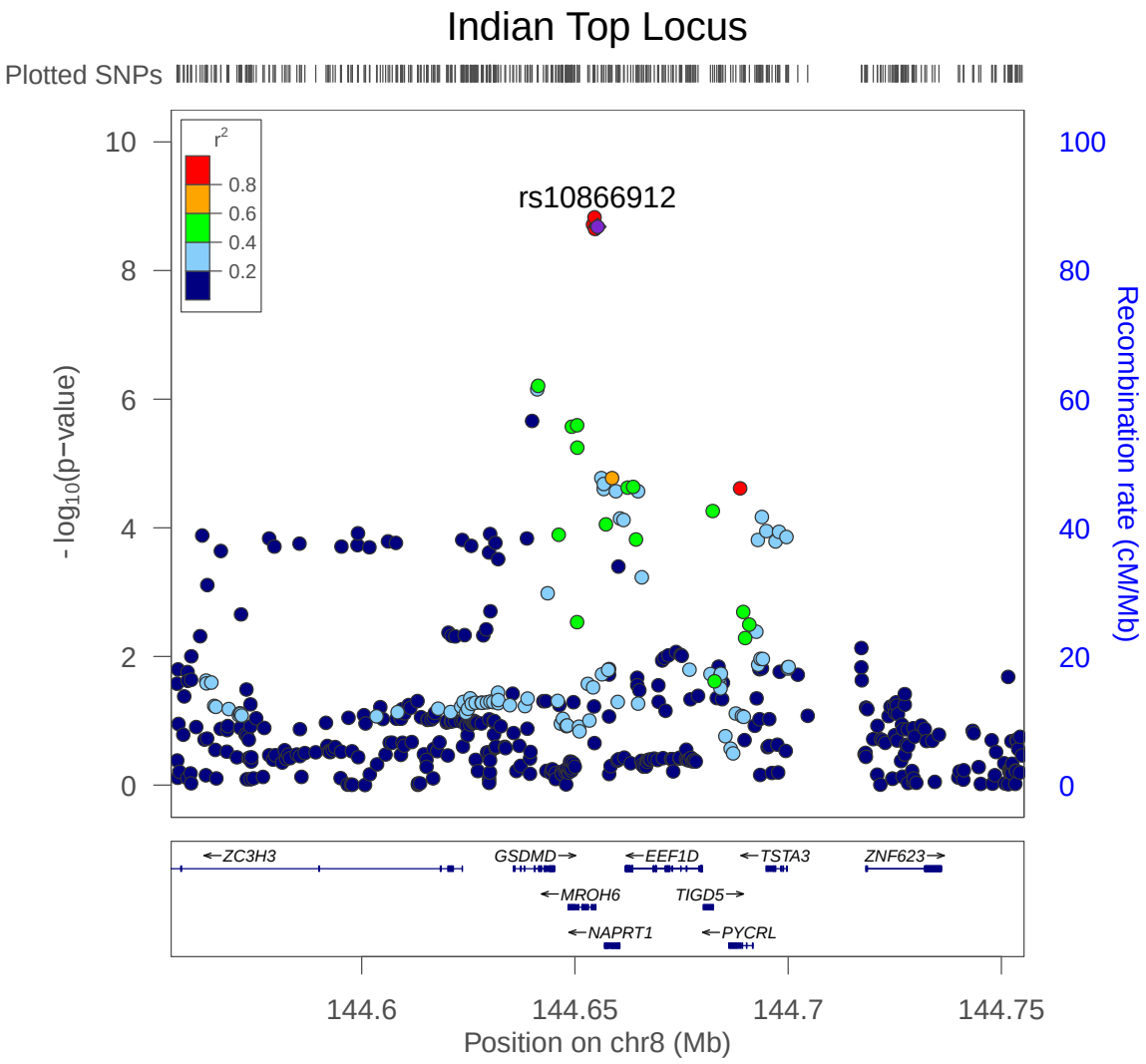
Manhattan plot showing genome-wide schizophrenia associations of PGC2 and Indian samples. The x-axis shows the chromosomal position and the y-axis the significance of association ($-\log_{10}P$). The horizontal red line represents the level of genome-wide significance ($<5 \times 10^{-8}$). Our top genome-wide significant locus is situated on chromosome 8q24.3 (rs10866912, $P = 2.09 \times 10^{-9}$).

eFigure 9. Regional Plot of Chromosome 8q24.3 Locus (100-Kb Window) in Cross Population Meta-analysis Results With SAS LD



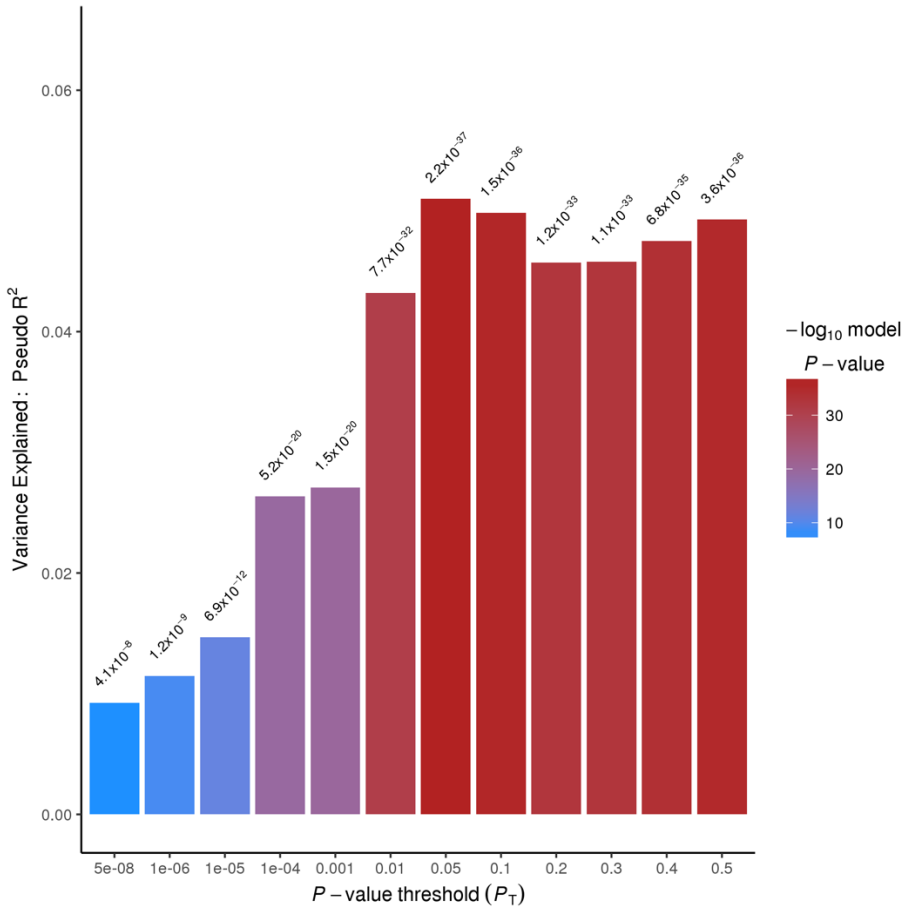
The index SNP, rs10866912 (coloured purple) was significant (P = 1.58 ×10^{-09}) in trans ethnic meta-analysis. The other SNPs are coloured according to the degree of linkage disequilibrium (measured by r^2) with the index SNP. The Thousand Genome Phase 3 SAS population was used to calculate LD. The x-axis shows the SNP locus position on chromosome 8 (GRCh37/hg19 build). The y-axis shows the significance of association (-log_{10}P). Nine genes are located within the 100Kb window, the direction of transcription (upstream/downstream) being annotated with arrows.

eFigure 10. Regional Plot of Chromosome 8q24.3 Locus (100Kb window) in Cross Population Meta-analysis Results With EUR LD



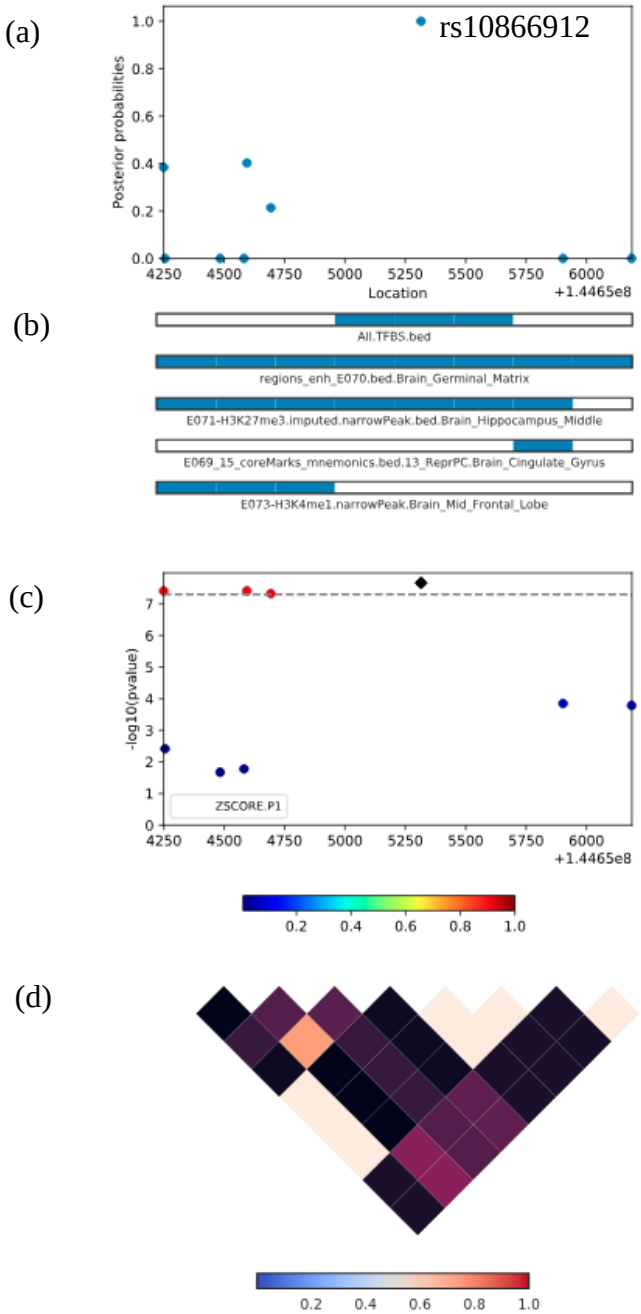
The index SNP, rs10866912 (coloured purple) was significant (P = 1.58 ×10^{-09}) in trans ethnic meta-analysis. The other SNPs are coloured according to the degree of linkage disequilibrium (measured by r^2) with the index SNP. The Thousand Genome Phase 3 EUR population was used to calculate LD. The x-axis shows the SNP locus position on chromosome 8 (GRCh37/hg19 build). The y-axis shows the significance of association (-log_{10}P). Nine genes are located within the 100Kb window, the direction of transcription (upstream/downstream) being annotated with arrows.

eFigure 11. Genomic Profile Risk (PRS) Scoring



Results of profile risk scores using the PGC2 schizophrenia results as the discovery set and the Indian data set as the target set. The x-axis shows the *P*-value thresholds (P_T) used to calculate the polygenic risk scores. The y-axis shows the Nagelkerke R^2 values (observed scale), representing the proportion of variance in case–control status explained by the risk profile scores. The p-value for the risk profile score's predictive ability of case–control status for a given P_T is shown above each bar. European genetic profile scores (maximum Nagelkerke's R^2=0.05; $P = 2.2 \times 10^{-37}$) at *P* value threshold P_T = 0.05 was most predictive of schizophrenia status in the Indian population.

eFigure 12. Fine Mapping of Indian Top Locus

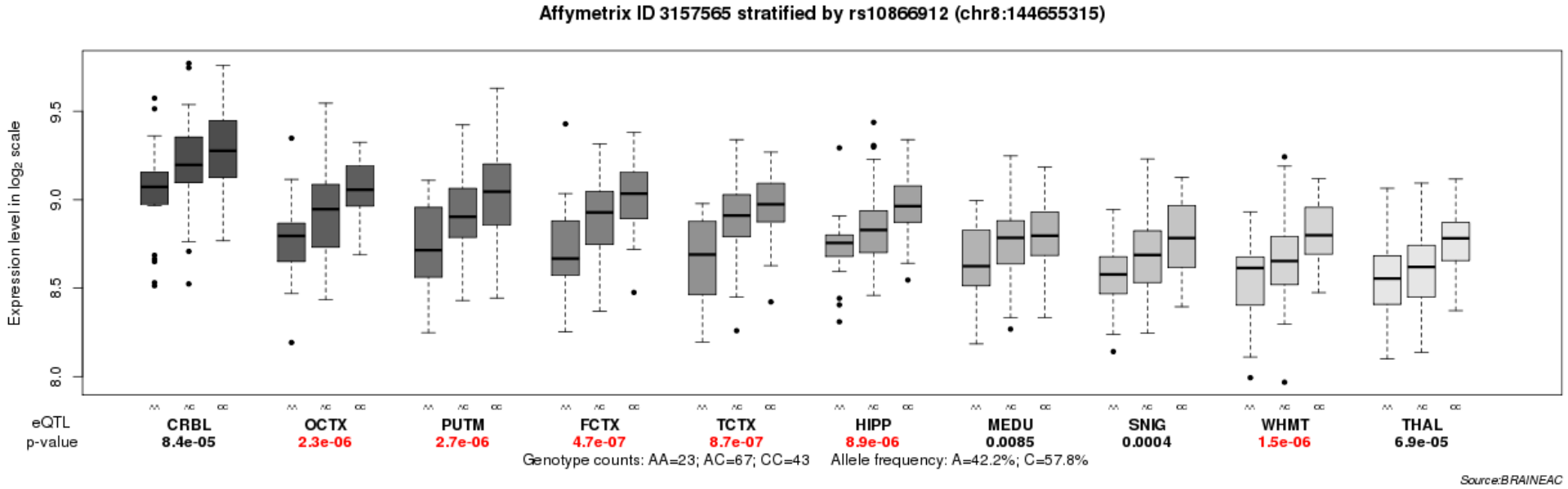


(a) Output from PAINTOR 3.1: demonstrating that our top SNP (rs10866912) has the highest posterior probability of all SNPs at the top Indian locus, quantifying the likelihood that this SNP is causal (i.e. should be prioritised for functional investigation); (b) – (d) depict inputs for PAINTOR; (b) Functional annotation data sets (from top to bottom: Transcription Factor Binding Sites; Enhancer regions; Methylation sites in brain (Hippocampus, Cingulate, Frontal Lobe); (c) Z-scores at the given locus; (d) LD matrix of locus defined by r^2>0.6.

eFigure 13. Summary of the eQTL Results of *NAPRT1*in GTEx Heat Map

The columns are SNPs, and by default are sorted by base-pair position (hg19) within our top locus (~17Kb window). The rows are tissues sorted in order of decreasing eQTL effect size based on our top SNP, rs10866912 (in bold) on *NAPRT1*. Only the tissues that have eQTLs associated with *NAPRT1* are reported. An eQTL appears as a circle, the colour and size of the circle being used to represent the effect size and *P* value of the eQTL result, respectively.

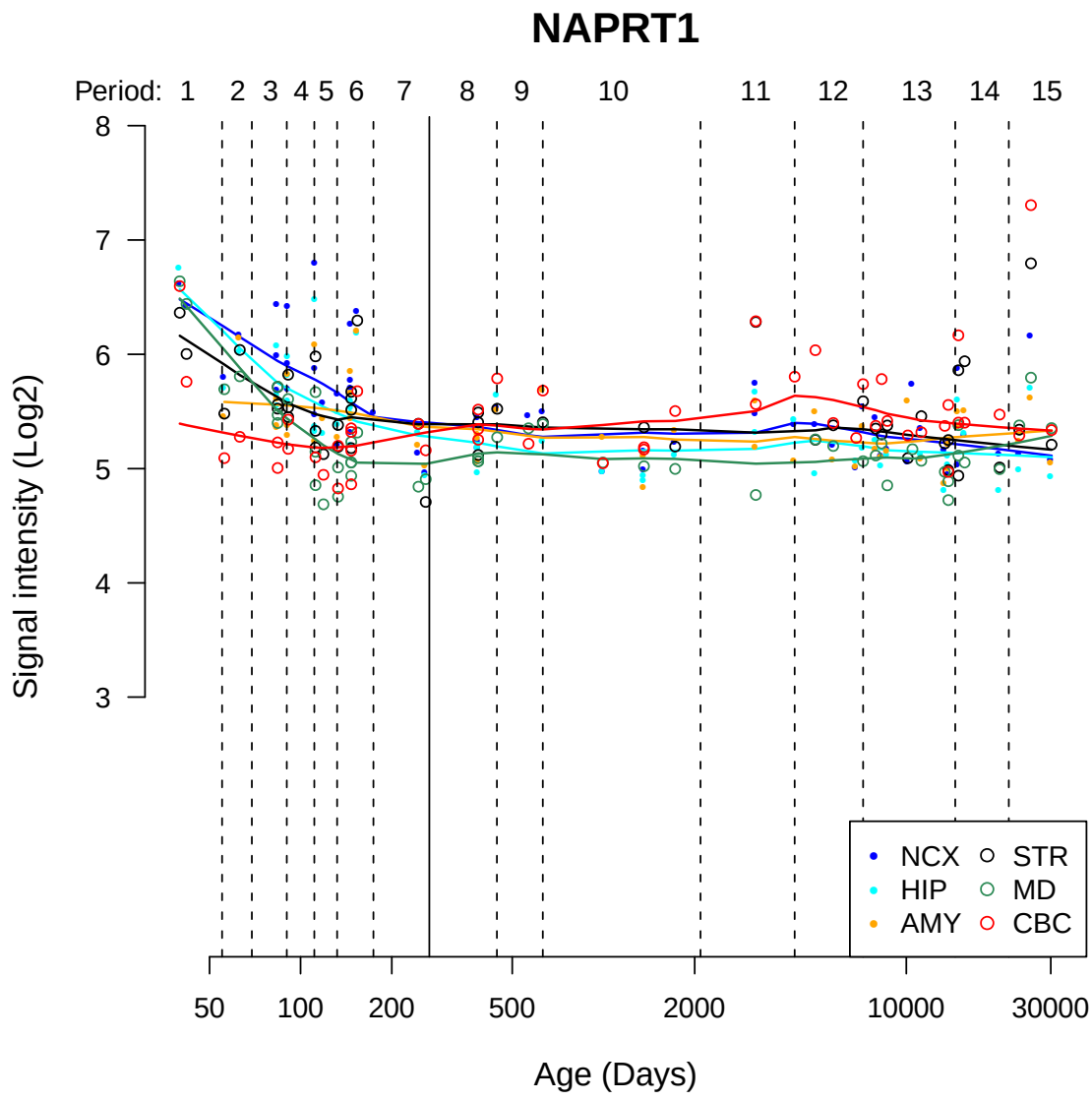
eFigure 14. Lead SNP, rs10866912, Is Associated (eQTL) With Brain Expression of *NAPRT1*



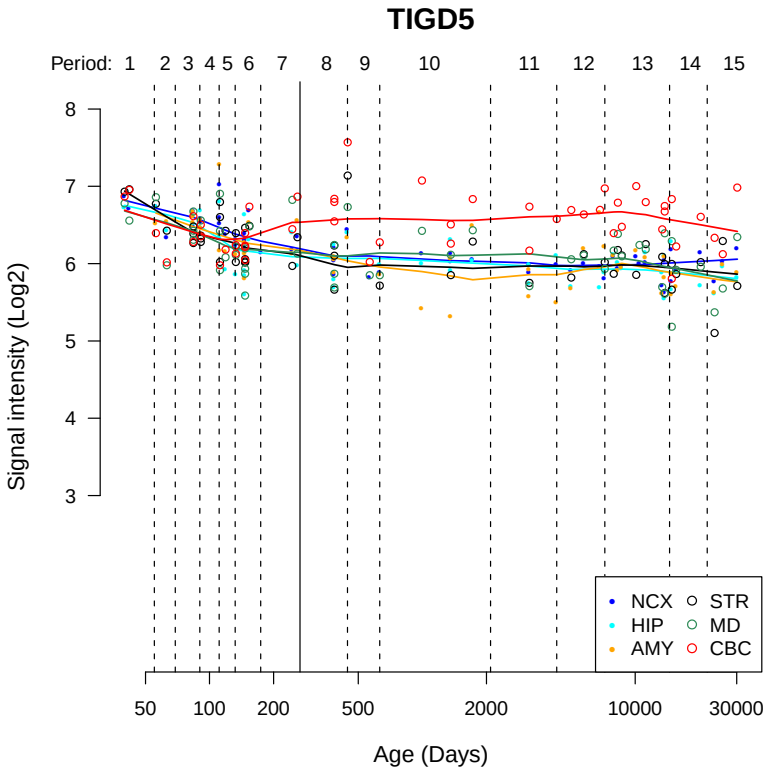
Allele specific expression and eQTL for the lead SNP across 10 brain regions (BRAINEAC; http://www.braineac.org/): cerebellar cortex (CRBL), frontal cortex (FCTX), hippocampus (HIPP), medulla (specifically inferior olivary nucleus, MEDU), occipital cortex (specifically primary visual cortex, OCTX), putamen (PUTM), substantia nigra (SNIG), thalamus (THAL), temporal cortex (TCTX) and intralobular white matter (WHMT). *P* values in red depict statistically significant results. Also see eTable 10.

eFigure 15. Spatiotemporal Brain Expression Profiling of Genes in Top Locus

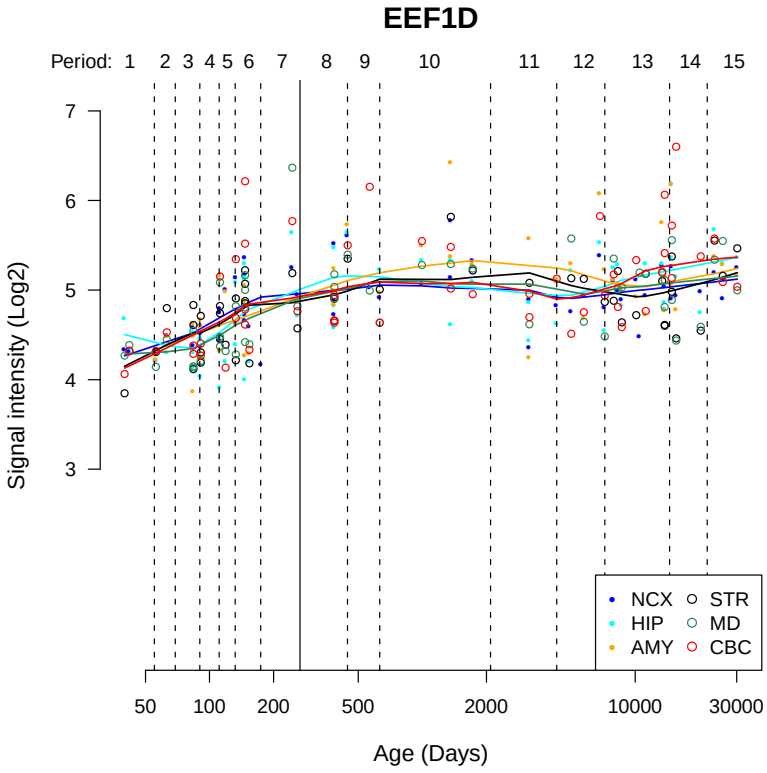
(A)



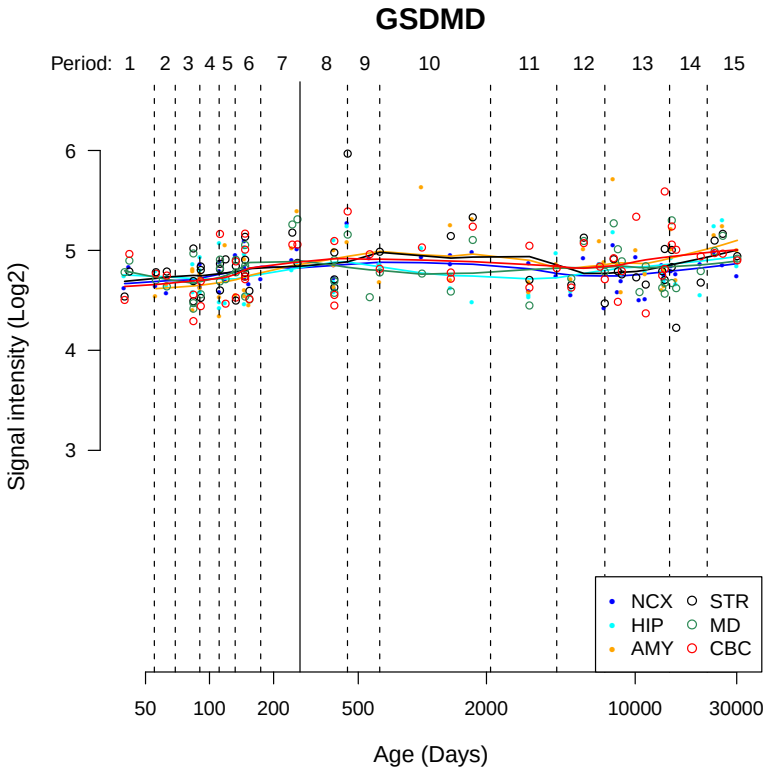
(B)



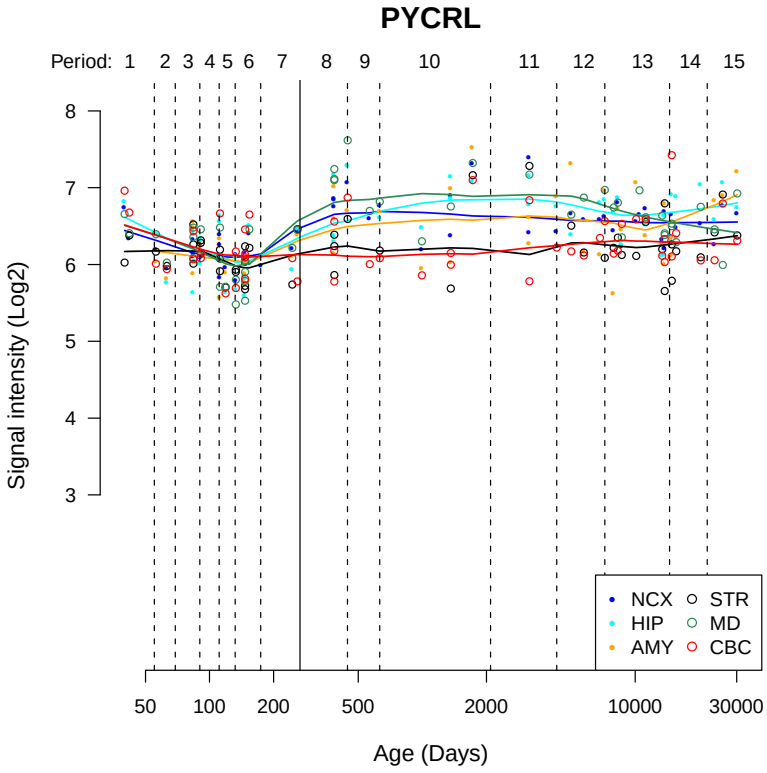
(C)



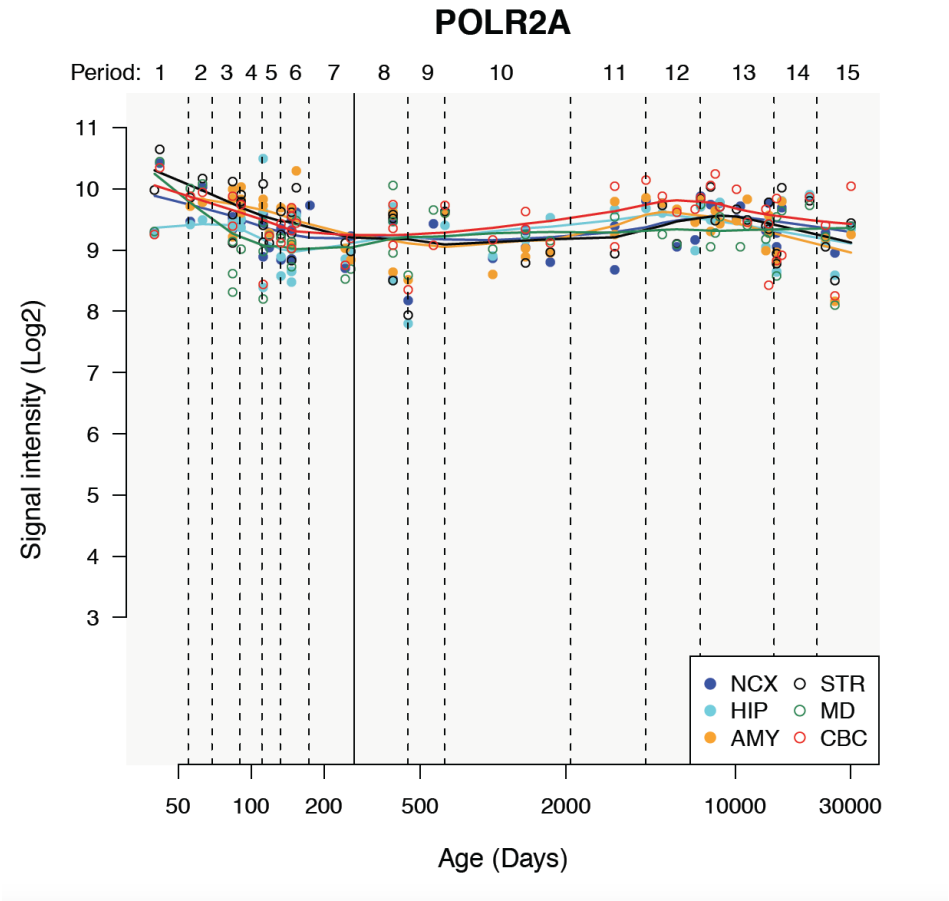
(D)



(E)

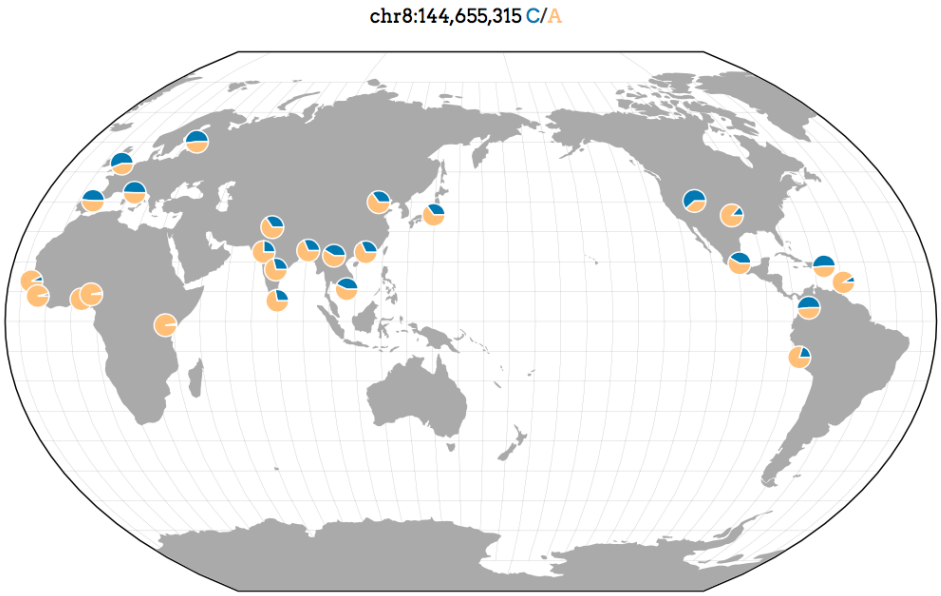


(F)



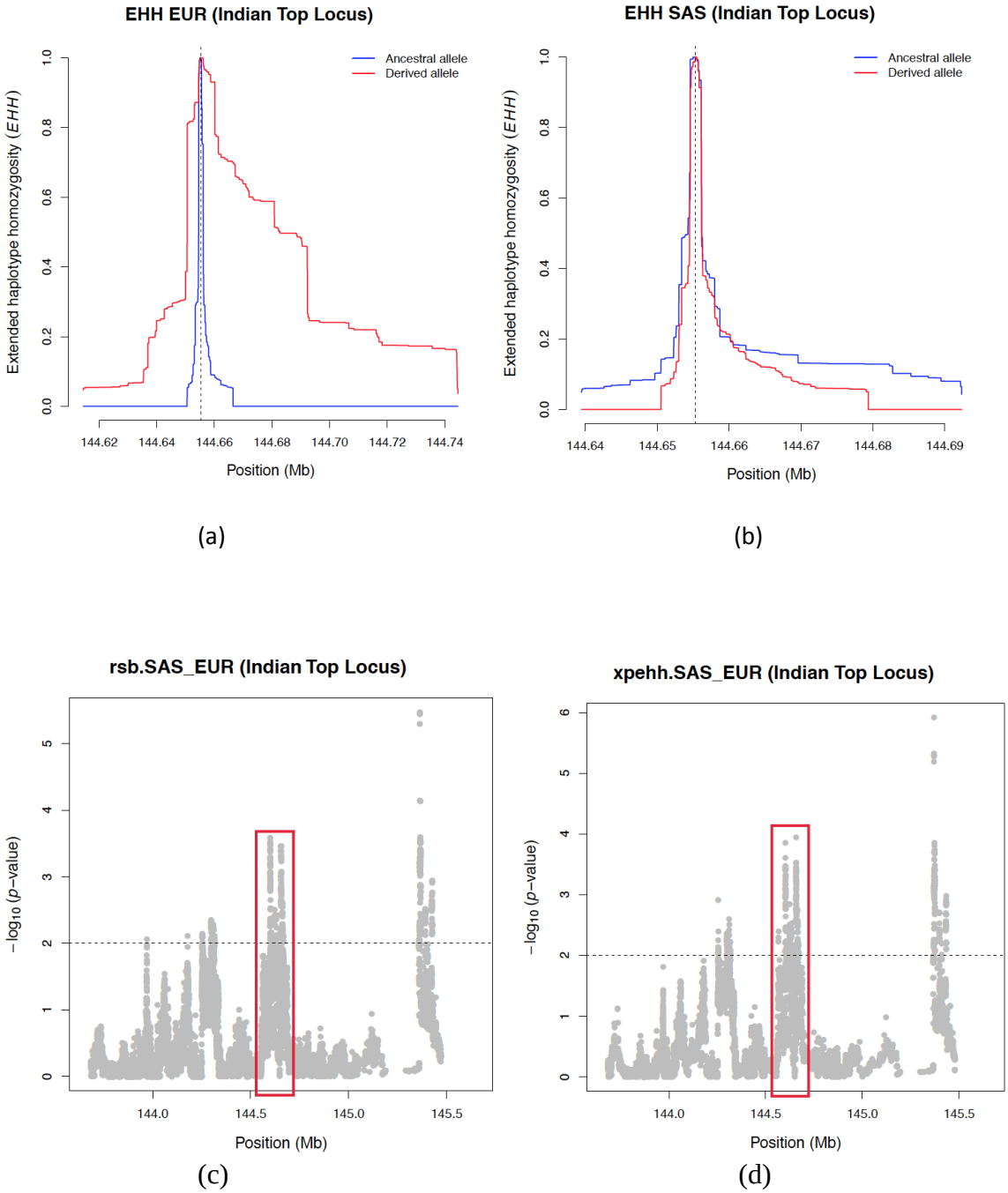
(A) shows the cis-eQTL gene, *NAPRT1*; (B-E) show the other genes in our top locus (defined by r^2>0.6): *TIGD5*, *EEF1D*, *GSDMD*, *PYCRL* respectively; note that the 6th gene in this locus, *MROH6*, is not represented in the Human Brain Transcriptome (HBT) project (www.hbatlas.org/); (F) depicts POLR2A, the most active transcription factor within our locus. Brain regions in these figures include: amygdala(AMY), cerebellar cortex(CBC), hippocampus(HIP), mediodorsal nucleus of the thalamus(MD), neocortex(NCX) and striatum(STR). Periods of human brain development and adulthood include: Embryonic (periods 1), Early fetal (periods 2–3), Early mid-fetal (periods 3–4), Late mid-fetal (periods 6), Late fetal (periods 7), Neonatal and early infancy (periods 8), Late infancy (periods 9), Early childhood (periods 10), Middle and late childhood (periods 11), Adolescence (periods 12), Young adulthood (periods 13), Middle adulthood (periods 14), Late adulthood (periods 15).

eFigure 16. Worldwide Pattern of Allele Frequencies for Indian Top SNP (rs10866912)



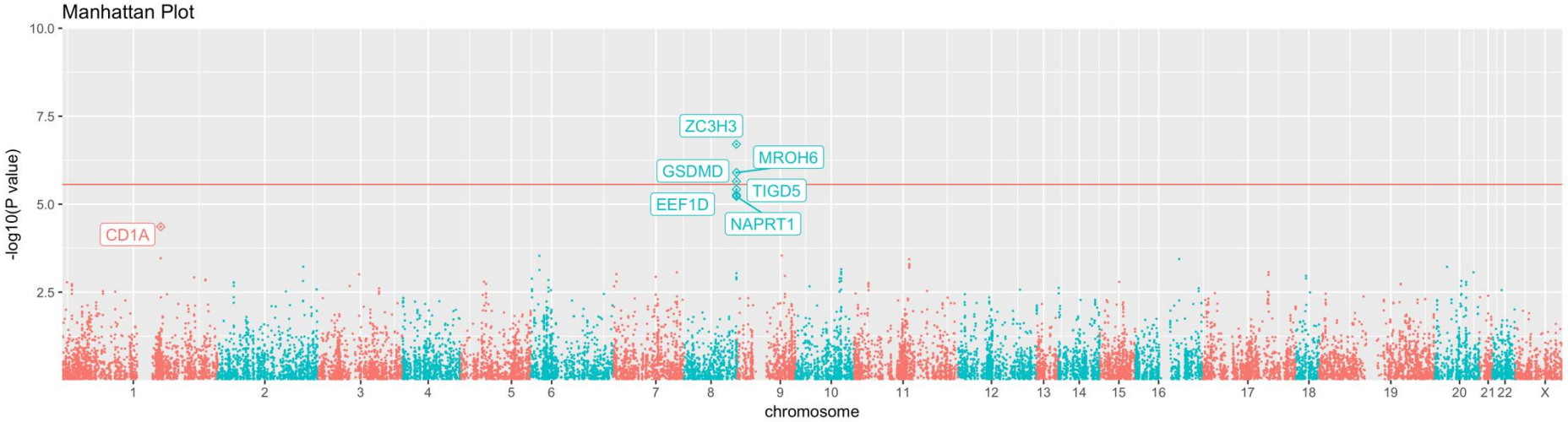
The frequency of the risk allele (“A”) is progressively declining from African (~97%) to Indian (~72%) and European (~43%), while the frequency of the protective allele (“C”) is increasing across these populations.

eFigure 17. Natural Selection Analyses Between South Asian and European Populations



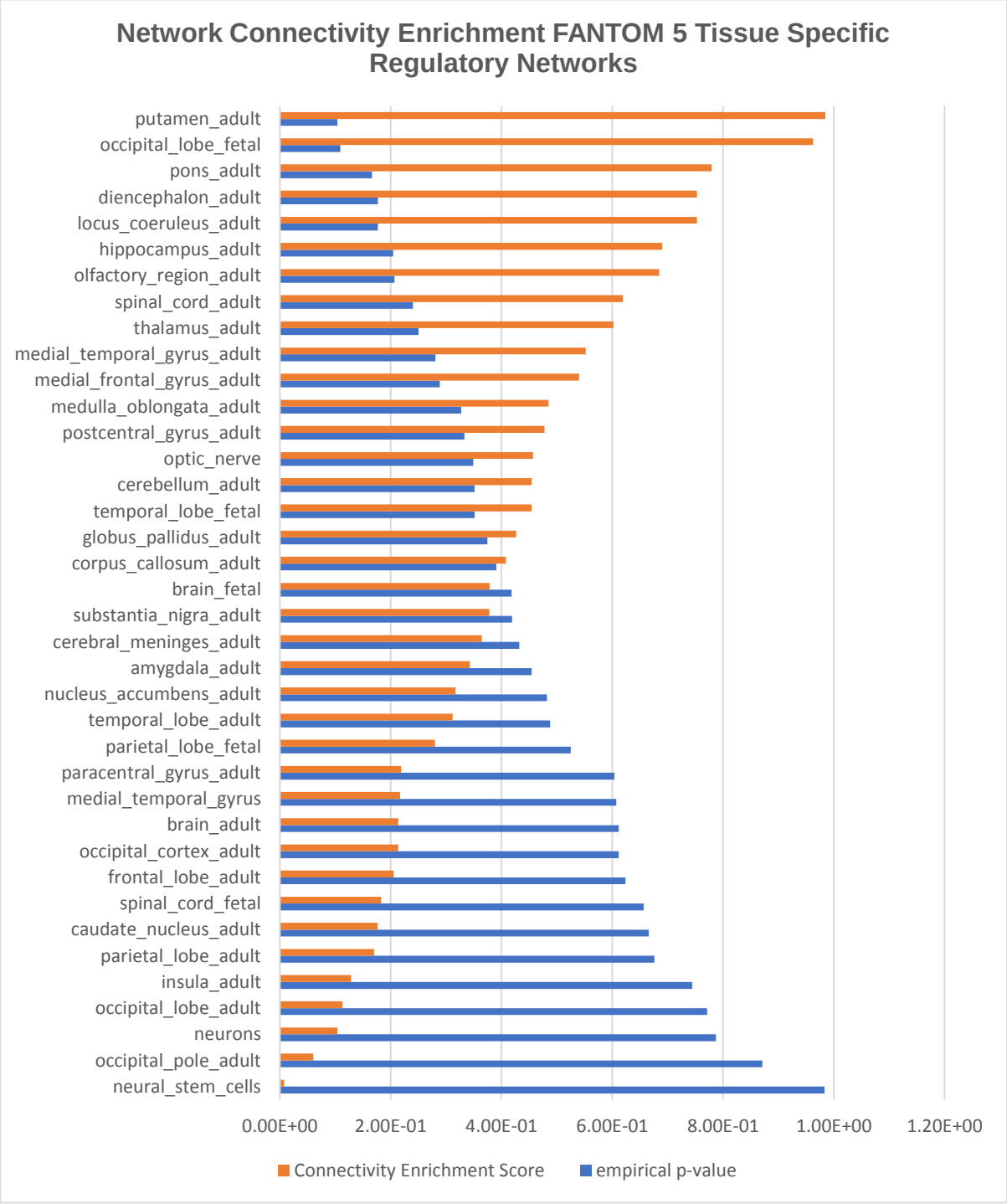
Extended Homozygosity Haplotype(EHH) based test of Indian top locus: (a) In 1000 Genome European populations showing relative decay of the ancestral (risk) and derived (protective) alleles, in which the highly homozygous derived allele is undergoing positive selection; (b) In 1000 Genome South Asian populations showing relative decay of the ancestral (risk) and derived (protective) alleles, in which neither allele is undergoing selection; (c) Rsb analyses (i.e. the observed standardised log-ratio of the integrated EHHS between SAS and European populations) showing a significant peak on chromosome 8q24.3 (144.6 - 144.85Mb) including our top locus indicating that that this region is undergoing selection in Europeans, but not in South Asians; (d) XP-EHH (i.e. cross-population extended haplotype homozygosity test for recent positive selection) analysis of the observed standardised log-ratio of the integrated EHHS between SAS and European populations) showing a significant peak on chromosome 8q24.3 (144.6 - 144.8Mb) including our top locus, indicating that the selective sweep has occurred in Europeans, but not in South Asian populations.

eFigure 18. Manhattan Plot for Gene-Based Association Using MAGMA Version 1.06



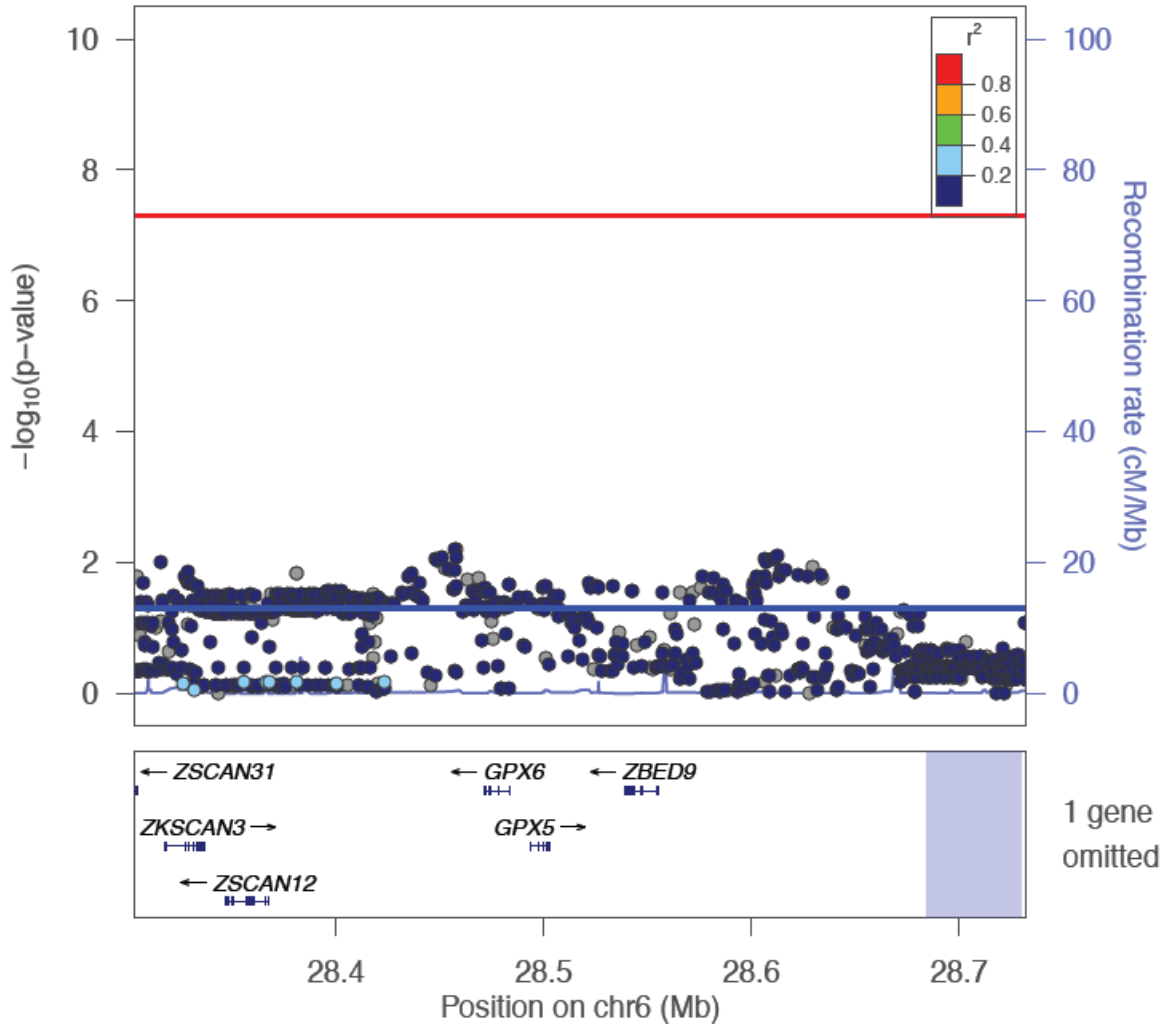
Manhattan plot showing gene-based genome-wide schizophrenia associations in 3092 Indian individuals (1321 cases, 1771 controls). The x-axis shows the chromosomal position and the y-axis the significance of association ($-\log_{10}P$). The horizontal red line represents the level of genome-wide significance (5×10^{-06}). Our top genome-wide significant locus is situated on chromosome 8q24.3.

eFigure 19. Network Connectivity Results for Indian Data Set Using MAGNUM



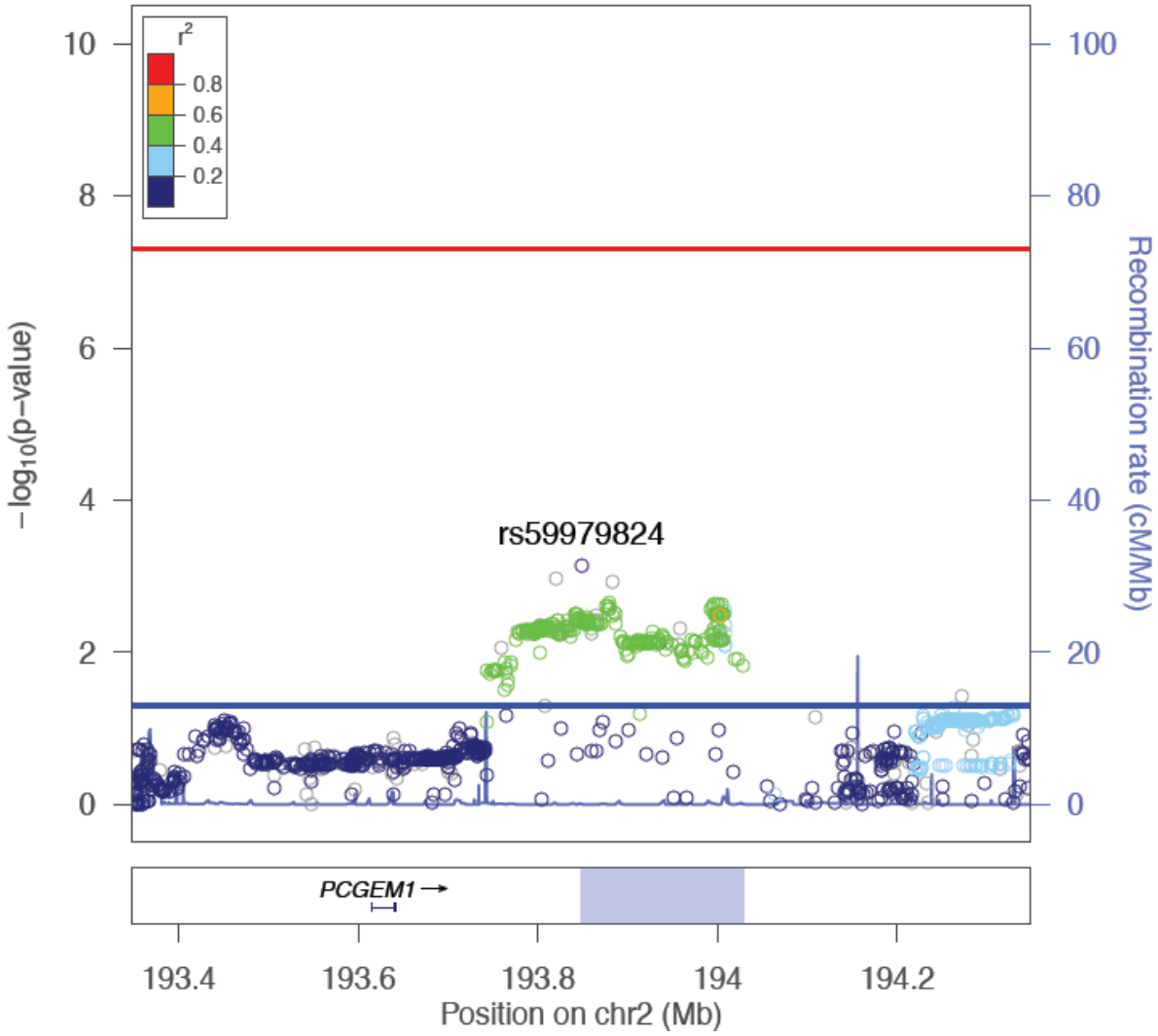
Plot showing brain regions (y-axis) with connectivity enrichment scores and empirical p-values (x-axis). The putamen generated the highest enrichment score, similar to what was observed in the PGC2 schizophrenia data set.

eFigure 20. Regional Plot of Indian SNP Data Set at PGC2 Top Locus (Chromosome 6p22:28303247-287712247; GRCh37/hg19 Build)



The PGC2 index SNP, chr6:28712247 is located within the 28.7Mb purple band representing the Sri Lankan Tamils in the UK (STU) high LD block (r^2>0.9; ~48Kb); all other SNPs, coloured in dark blue/light blue are in low LD with the index SNP. The Thousand Genome Phase 3 SAS population was used to calculate LD. The x-axis shows the SNP locus position on chromosome 6. The y-axis shows the significance of association ($-\log_{10}P$). Six genes are located within the PGC2 top locus window, the direction of transcription (upstream/downstream) being annotated with arrows.

eFigure 21. Regional Plot of Top Replicated PGC2 Locus (Chromosome 2q32.3:193848340-194028340; GRCh37/hg19 Build) in Indian SNP Data Set



The PGC2 index SNP, rs59979824 is located within the194Mb purple band representing the Sri Lankan Tamils in the UK (STU) high LD block (r^2>0.9); all other SNPs are coloured according to the degree of linkage disequilibrium with the index SNP. The Thousand Genome Phase 3 SAS population was used to calculate LD. The x-axis shows the SNP locus position on chromosome 2 (GRCh37/hg19 build). The y-axis shows the significance of association ($-\log_{10}P$). One gene is located upstream of the PGC2/STU locus (purple band), the direction of transcription (upstream/downstream) being annotated with arrows.

eTable 1. Indian Clinical Data

Ethnicity (n=3092)			
Caste/Relationship	**Tamil Ethnicity**		
Caste	**Brahmin**	**Non Brahmin**	**total**
Cases	299	977	1276
Family Controls	275	570	845
Unrelated Controls	47	834	881
	621	2381	3002
	Other South Indian Ethnicity		
Caste	**Brahmin**	**Non-Brahmin**	**total**
Cases	23	22	45
Family Controls	17	23	40
Unrelated Controls	5	0	5
	45	45	90
Total	**666**	**2426**	**3092**
Degree of consanguinity (Probands n=1231)			
None	NA	NA	823
Uncle-niece	NA	NA	35
First Cousins	NA	NA	27
Second Cousins	NA	NA	40
Others/Unknown consanguinity	NA	NA	68
Information Not Available	NA	NA	238

eTable 2. Clinical Characteristics

Clinical Characteristics (n=1321 Cases)		
	Mean	**SD**
Age at assessment (years)	39.08	11.42
Age at disease onset (years)	25.01	8.73
Age at first treatment (years)	27.34	9.44
Disease duration (years)	12.9	9.15
Affected individuals		
Male (Age)	38.4	11.1
Female (Age)	39.8	11.7

eTable 3. Symptoms and Treatment

DSM-IV Symptoms (n=1321)								
Symptoms (%)	**None**	**Questionable**	**Mild**	**Moderate**	**Marked**	**Severe**	**Unknown**	**Total (%)**
Affective Flattening or Blunting (Neg)	44	3.40	11.80	21.90	13.00	2.90	3.40	100
Alogia (Neg)	63	2.10	9.10	10.90	11.80	0.8	2.40	100
Avolition/Apathy (Neg)	24	1.70	13.60	22.20	28.00	7.30	3.00	100
Anhedonia/Asociality (Neg)	22	1.90	12.20	27.80	25.20	7.30	3.50	100
Hallucinations (Pos)	46	6.00	9.60	18.00	10.90	4.00	5.30	100
Delusions (Pos)	25	6.00	8.50	23.30	21.70	6.90	8.20	100
Bizarre Behavior (Pos)	69	2.20	8.80	10.20	4.60	1.60	3.90	100
Formal Thought Disorder (Pos)	78	3.30	6.50	5.20	3.00	0.70	3.00	100
Treatment status at recruitment (n=1321)								
On Treatment	NA	NA	NA	NA	NA	NA	NA	88.40
Once Treated (discontinued Treatment)	NA	NA	NA	NA	NA	NA	NA	8.30
Never Treated	NA	NA	NA	NA	NA	NA	NA	3.30

Neg: Negative symptoms; Pos: Positive symptoms

eTable 4. Forward and Reverse Primer Sequences for DRE-*NAPRT1* mRNA

Oligos Name	Oligos sequence	Mature
NAPRT1-01a	TGCTGTAAAGGAGATGCAGGAATTTC GTTTTGGCCACTGACTGACGAAATTC CCATCTCCTTTA	TAAAGGAGATGCAG GAATTTC
NAPRT1-01b	CCTGTAAAGGAGATGGGAATTTCGTC AGTCAGTGGCCAAAACGAAATTCCTG CATCTCCTTTAC	
NAPRT1-02a	TGCTGTCTAGATTCATGGTTGAGACT GTTTTGGCCACTGACTGACAGTCTCA AATGAATCTAGA	TCTAGATTCATGGTT GAGACT
NAPRT1-02b	CCTGTCTAGATTCATTTGAGACTGTC AGTCAGTGGCCAAAACAGTCTCAAC CATGAATCTAGAC	
NAPRT1-03a	TGCTGAATCCAGCATCTCTGCATCTG GTTTTGGCCACTGACTGACCAGATG CAGATGCTGGATT	GAATCCAGCATCTC TGCATCTG
NAPRT1-03b	CCTGAATCCAGCATCTGCATCTGGTC AGTCAGTGGCCAAAACCAGATGCAG AGATGCTGGATTC	
NAPRT1-04a	TGCTGAGCACAGTGAATCCAGCATCT GTTTTGGCCACTGACTGACAGATGCT GTTCACTGTGCT	AGCACAGTGAATCC AGCATCT
NAPRT1-04b	CCTGAGCACAGTGAACAGCATCTGT CAGTCAGTGGCCAAAACAGATGCTG GATTCACTGTGCTC	
dreNAPRT1-Forward	GGGGacaagtttgtacAAAAAAgcaggctAT GGCGACAAGTAACGAAGC	
dreNAPRT1-Reverse	GGGGaccacTTTGTAcaagaaagctgggtT CAGTTGGACAGCAGGAAGT	

eTable 5. Study Population Haplotypes Present in Our Top Locus

SNP	Haplotype A	Haplotype B	Haplotype C	Haplotype D
rs58774517	T	C	C	T
rs1869434	G	G	A	G
rs4606077	G	A	A	G
rs4873805	G	A	G	G
rs12546272	A	G	G	A
rs367672647	-	+	+	-
rs10866912	A	A	A	C
Population frequency (%)	31.1	22.0	18.9	26.8

Four distinct haplotypes in our top locus situated between the *MROH6* and *NAPRT1* genes on chromosome 8, covering 98.8% of the alleles present in the study population. Haplotypes differed at 6 SNPs and one indel. The risk allele (A) of rs10866912 was present in haplotypes A, B, and C, while the protective allele (C) of our index SNP was present in Haplotype D.

eTable 6. Imputation Batch-Based Meta-analysis for Top Indian Locus

SNP	A1	A2	Freq	Effect	SE	P-value	OR
rs10866912	a	c	0.7231	0.0585	0.0117	3.53E-07	1.27
rs10866911	a	g	0.2697	-0.0586	0.0119	5.04E-07	0.79
rs4873803	t	g	0.2693	-0.0586	0.0118	4.90E-07	0.79
rs4873804	c	g	0.2703	-0.0578	0.0118	6.82E-07	0.79

eTable 7. Transethnic Meta-analysis Results for Indian Top Locus With PGC2 Using Metasoft-RE2C

SNP ID	LS *P*-value	Heterogeneity statistic	RE2C *P*-value
rs4873803	1.70E-06	94	1.49E-09
rs10866911	2.35E-06	94	1.93E-09
rs10866912	4.28E-06	95	2.09E-09
rs4873804	2.34E-06	94	2.26E-09
rs2305492	8.33E-05	93	6.20E-07
rs4292703	0.00011498	93	7.03E-07
rs7034361	0.66104456	96	4.60E-06
rs16998895	0.00315589	93	1.05E-05
rs35043390	0.00347417	93	1.60E-05
rs12199670	0.61251295	95	2.77E-05
rs4776288	0.37430489	95	8.17E-05

LS *P*-value: Lin-Sullivan p-value; FE: Fixed Effect; RE2C *P*-value: Optimized Random Effect p-value; Notably all SNPs in the Indian top locus have high heterogeneity statistic (>65), indicating that these SNPs are heterogeneous between PGC2 and Indian data sets (i.e. having different allelic effects, allele frequencies, and differences in linkage disequilibrium)

eTable 8. Transethnic Meta-analysis Results for 108 PGC2 Significant Regions Using Metasoft-RE2C

Rank	*P value*	Position (hg19)	SNP ID	Chr	BP	Meta *P value*	Is more significant
1	3.48×10^{-31}	chr6:28303247-28712247	rs7766356	6	28400538	3.48×10^{-21}	FALSE
2	3.36×10^{-19}	chr1:97792625-98559084	rs1702294	1	98501984	1.10×10^{-17}	FALSE
3	6.20×10^{-19}	chr10:104423800-105165583	rs12244388	10	104640052	2.04×10^{-18}	FALSE
4	3.22×10^{-18}	chr12:2321860-2523731	rs2007044	12	2344960	9.10×10^{-18}	FALSE
5	1.74×10^{-15}	chr8:143309503-143330533	rs4129585	8	143312933	7.36×10^{-13}	FALSE
6	7.98×10^{-15}	chr4:103146888-103198090	rs13150344	4	103170124	4.83×10^{-05}	FALSE
7	8.20×10^{-15}	chr7:1896096-2190096	rs58120505	7	2029867	7.61×10^{-14}	FALSE
8	1.10×10^{-14}	chr5:60499143-60843543	rs4391122	5	60598543	1.36×10^{-13}	FALSE
9	1.86×10^{-14}	chr12:123448113-123909113	rs7299943	12	123593485	1.98×10^{-08}	FALSE
10	5.65×10^{-14}	chr2:200715237-200848037	rs11693528	2	200736507	2.53×10^{-14}	TRUE
11	8.30×10^{-14}	chr15:91416560-91429040	rs4702	15	91426560	1.86×10^{-11}	FALSE
12	1.05×10^{-13}	chr3:36843183-36945783	rs75968099	3	36858583	2.33×10^{-13}	FALSE
13	1.36×10^{-13}	chr14:103996234-104184834	rs12887734	14	104046834	1.86×10^{-13}	FALSE
14	2.44×10^{-13}	chr15:78803032-78926732	rs8043009	15	78908154	3.06×10^{-12}	FALSE
15	3.03×10^{-13}	chr7:110843815-111205915	rs13240464	7	110898915	2.01×10^{-12}	FALSE
16	1.09×10^{-12}	chr11:130714610-130749330	rs10750450	11	130719061	7.20×10^{-13}	TRUE
17	1.53×10^{-12}	chr2:185601420-185785420	rs11693094	2	185601420	3.41×10^{-11}	FALSE
18	1.61×10^{-12}	chrX:21193266-21570266	rs6528018	X	21413018	3.84×10^{-01}	FALSE
19	1.97×10^{-12}	chr10:18681005-18770105	rs17691888	10	18734528	7.44×10^{-11}	FALSE
20	2.02×10^{-12}	chr12:57428314-57682971	rs11172142	12	57680101	1.21×10^{-03}	FALSE
21	2.03×10^{-12}	chr1:73766426-73991366	rs12129573	1	73768366	7.38×10^{-10}	FALSE
22	2.32×10^{-12}	chr2:233559301-233753501	rs778371	2	233743109	4.53×10^{-12}	FALSE
23	2.80×10^{-12}	chr11:124610007-124620147	rs55661361	11	124613957	4.05×10^{-12}	FALSE
24	3.34×10^{-12}	chr18:52747686-53200117	rs9636107	18	53200117	1.22×10^{-12}	TRUE
25	1.26×10^{-11}	chr11:46342943-46751213	rs7951870	11	46373311	9.35×10^{-11}	FALSE
26	1.30×10^{-11}	chr3:180588843-181205585	rs34796896	3	180623255	1.17×10^{-10}	FALSE
27	1.46×10^{-11}	chr20:37361494-37485994	rs6065094	20	37453194	5.25×10^{-12}	TRUE

Rank	*P value*	Position (hg19)	SNP ID	Chr	BP	Meta *P value*	Is more significant
28	1.47×10^{-11}	chr2:57943593-58502192	rs11682175	2	57987593	3.60×10^{-12}	TRUE
29	1.62×10^{-11}	chr15:84661161-85153461	rs66486766	15	84806060	5.03×10^{-11}	FALSE
30	1.97×10^{-11}	chr18:53453389-53804154	rs7235891	18	53454774	1.29×10^{-08}	FALSE
31	2.06×10^{-11}	chr2:198148577-198835577	rs2565164	2	198337520	1.79×10^{-11}	TRUE
32	2.07×10^{-11}	chr22:41408556-41675156	rs9607782	22	41587556	4.43×10^{-13}	TRUE
33	2.61×10^{-11}	chr8:111460061-111630761	rs4642619	8	111487468	7.50×10^{-11}	FALSE
34	2.69×10^{-11}	chr3:2532786-2561686	rs17194490	3	2547786	5.25×10^{-11}	FALSE
35	2.75×10^{-11}	chr11:113317794-113423994	rs2514218	11	113392994	8.50×10^{-10}	FALSE
36	3.87×10^{-11}	chr11:133808069-133852969	rs502834	11	133817333	7.22×10^{-07}	FALSE
37	4.26×10^{-11}	chr3:52541105-52903405	rs2535627	3	52845105	7.52×10^{-11}	FALSE
38	4.55×10^{-11}	chr16:29924377-30144877	rs12691307	16	29939877	3.67×10^{-10}	FALSE
39	4.73×10^{-11}	chr22:39975317-40016817	rs926231	22	40006986	2.22×10^{-11}	TRUE
40	7.26×10^{-11}	chr3:135807405-136615405	rs7427564	3	136274435	1.63×10^{-11}	TRUE
41	1.06×10^{-10}	chr5:151941104-152797656	rs2910032	5	152540354	1.57×10^{-09}	FALSE
42	1.98×10^{-10}	chrX:68377126-68379036	rs5936660	X	68377937	8.67×10^{-07}	FALSE
43	2.86×10^{-10}	chr17:2095899-2220799	rs4523957	17	2208899	2.64×10^{-09}	FALSE
44	3.33×10^{-10}	chr7:86403226-86459326	rs12704290	7	86427626	1.64×10^{-11}	TRUE
45	3.38×10^{-10}	chr15:61831663-61909663	rs35225048	15	61856263	7.08×10^{-10}	FALSE
46	3.39×10^{-10}	chr1:44029384-44128084	rs11210892	1	44100084	2.03×10^{-11}	TRUE
47	3.63×10^{-10}	chr19:19374022-19658022	rs2905426	19	19478022	9.49×10^{-10}	FALSE
48	4.49×10^{-10}	chr1:149998890-150242490	rs11578204	1	150001721	3.46×10^{-03}	FALSE
49	8.15×10^{-10}	chr6:84279922-84407274	rs3798869	6	84328660	1.16×10^{-10}	TRUE
50	8.70×10^{-10}	chr1:2372401-2402501	rs4648845	1	2387101	5.78×10^{-09}	FALSE
51	1.01×10^{-09}	chr16:13728459-13761359	rs7405404	16	13749859	4.71×10^{-10}	TRUE
52	1.13×10^{-09}	chr7:104598064-105063064	rs6466055	7	104929064	1.48×10^{-09}	FALSE
53	1.17×10^{-09}	chr1:8411184-8638984	rs301797	1	8487323	2.09×10^{-12}	TRUE
54	1.40×10^{-09}	chr12:110723245-110723245	NA	NA	NA	NA	NA
55	1.47×10^{-09}	chr4:170357552-170646052	rs12647126	4	170376705	3.17×10^{-08}	FALSE
56	1.64×10^{-09}	chr6:96459651-96459651	NA	NA	NA	NA	NA

Rank	*P value*	Position (hg19)	SNP ID	Chr	BP	Meta *P value*	Is more significant
57	1.71×10^{-09}	chr22:42315744-42689414	rs6002655	22	42603814	1.87×10^{-09}	FALSE
58	1.81×10^{-09}	chr2:146416922-146441832	rs6430095	2	146439945	2.55×10^{-09}	FALSE
59	2.24×10^{-09}	chr11:57386294-57682294	rs9420	11	57510294	7.48×10^{-08}	FALSE
60	2.55×10^{-09}	chr11:24367320-24412990	rs12363019	11	24374545	3.21×10^{-09}	FALSE
61	2.86×10^{-09}	chr1:30412551-30437271	rs1498232	1	30433951	1.29×10^{-09}	TRUE
62	3.28×10^{-09}	chr7:137039644-137085244	rs3735025	7	137074844	9.73×10^{-09}	FALSE
63	3.61×10^{-09}	chr9:84630941-84813641	rs11139497	9	84739941	3.00×10^{-09}	TRUE
64	3.73×10^{-09}	chr1:243503719-244002945	rs10803138	1	243555219	2.38×10^{-08}	FALSE
65	4.18×10^{-09}	chr15:40566759-40602237	rs56282503	15	40566759	3.12×10^{-07}	FALSE
66	4.49×10^{-09}	chr19:30981643-31039023	rs2053079	19	30987423	2.29×10^{-09}	TRUE
67	4.61×10^{-09}	chr5:88581331-88854331	rs62378245	5	88743962	1.22×10^{-08}	FALSE
68	4.64×10^{-09}	chr3:17221366-17888266	rs4908972	3	17796192	8.38×10^{-09}	FALSE
69	4.67×10^{-09}	chr5:137598121-137948092	rs3849046	5	137851192	2.12×10^{-09}	TRUE
70	4.80×10^{-09}	chr14:99707919-99719219	rs2693698	14	99719219	5.13×10^{-08}	FALSE
71	4.86×10^{-09}	chr14:72417326-72450526	rs2332700	14	72417326	1.35×10^{-09}	TRUE
72	5.05×10^{-09}	chr5:45291475-45393775	rs9292918	5	45301035	6.46×10^{-08}	FALSE
73	5.97×10^{-09}	chr8:60475469-60954469	rs6984242	8	60700469	1.27×10^{-09}	TRUE
74	7.39×10^{-09}	chr2:72357335-72368185	rs55707322	2	72358959	2.84×10^{-03}	FALSE
75	7.54×10^{-09}	chr11:123394636-123395986	rs7927176	11	123395864	1.58×10^{-08}	FALSE
76	8.33×10^{-09}	chr2:200161422-200309252	rs6704641	2	200164252	3.64×10^{-08}	FALSE
77	8.41×10^{-09}	chr2:193848340-194028340	rs59979824	2	193848340	2.69×10^{-10}	TRUE
78	9.47×10^{-09}	chr4:176851001-176875801	rs1106568	4	176861301	1.19×10^{-08}	FALSE
79	1.06×10^{-08}	chr8:4177794-4192544	rs10046758	8	4184170	8.60×10^{-08}	FALSE
80	1.12×10^{-08}	chr2:225334096-225467796	rs11685299	2	225391296	1.38×10^{-08}	FALSE
81	1.22×10^{-08}	chr8:89340626-89753626	rs6990941	8	89644431	1.78×10^{-08}	FALSE
82	1.28×10^{-08}	chr16:9875519-9970219	rs7191183	16	9900057	1.57×10^{-08}	FALSE
83	1.41×10^{-08}	chr14:30189985-30190316	rs2068012	14	30190316	3.39×10^{-07}	FALSE
84	1.43×10^{-08}	chr3:63792650-64004050	rs704364	3	63874734	7.48×10^{-08}	FALSE
85	1.51×10^{-08}	chr16:67709340-68311340	rs9928653	16	68252079	6.83×10^{-08}	FALSE

Rank	*P value*	Position (hg19)	SNP ID	Chr	BP	Meta *P value*	Is more significant
86	1.59×10^{-08}	chr2:149390778-149520178	rs12614977	2	149510282	1.72×10^{-04}	FALSE
87	1.77×10^{-08}	chr17:17722402-18030202	rs8082590	17	17958402	3.13×10^{-08}	FALSE
88	1.79×10^{-08}	chr15:70573672-70628872	rs1355585	15	70586617	6.72×10^{-08}	FALSE
89	1.87×10^{-08}	chr16:58669293-58682833	rs12325245	16	58681393	4.75×10^{-09}	TRUE
90	2.10×10^{-08}	chr8:27412627-27453627	rs59724122	8	27424696	8.15×10^{-08}	FALSE
91	2.21×10^{-08}	chrX:5916533-6032733	rs73627050	X	6012628	7.02×10^{-06}	FALSE
92	2.69×10^{-08}	chr6:73132701-73171901	rs1339227	6	73155701	1.43×10^{-07}	FALSE
93	2.85×10^{-08}	chr7:24619494-24832094	rs12154597	7	24785882	6.66×10^{-07}	FALSE
94	3.05×10^{-08}	chr5:109030036-109209066	rs71575306	5	109046210	1.28×10^{-07}	FALSE
95	3.06×10^{-08}	chr4:23366403-23443403	rs215411	4	23423603	6.25×10^{-08}	FALSE
96	3.15×10^{-08}	chr5:153671057-153688217	rs6864084	5	153685526	1.27×10^{-07}	FALSE
97	3.70×10^{-08}	chr11:109285471-109610071	rs3858402	11	109555838	1.08×10^{-07}	FALSE
98	3.71×10^{-08}	chr7:110034393-110106693	rs9656169	7	110074276	7.86×10^{-07}	FALSE
99	3.91×10^{-08}	chr12:29905265-29940365	rs679087	12	29917265	3.88×10^{-08}	TRUE
100	4.42×10^{-08}	chr7:131539263-131567263	rs7801375	7	131567263	5.82×10^{-08}	FALSE
101	4.45×10^{-08}	chr1:177247821-177300821	rs16851048	1	177276006	1.88×10^{-07}	FALSE
102	4.47×10^{-08}	chr1:207912183-208024083	rs7523273	1	207977083	6.91×10^{-08}	FALSE
103	4.56×10^{-08}	chr20:48114136-48131649	rs6019879	20	48130628	1.05×10^{-07}	FALSE
104	4.59×10^{-08}	chr12:92243186-92258286	NA	NA	NA	NA	NA
105	4.62×10^{-08}	chr2:162798555-162910255	rs2909457	2	162845855	1.11×10^{-07}	FALSE
106	4.69×10^{-08}	chr19:50067499-50135399	rs56873913	19	50091199	4.34×10^{-07}	FALSE
107	4.84×10^{-08}	chr12:103559855-103616655	rs10860965	12	103605337	4.60×10^{-06}	FALSE
108	4.85×10^{-08}	chr5:140023664-140222664	rs13168670	5	140146899	6.25×10^{-08}	FALSE

RANK: PGC2 risk loci ranked according to their significance; *P*-value: PGC2 *P values*; SNP ID: SNP with most significant *P value* in the PGC region; Chr: Chromosome; BP: Base pair (hg19); Meta *P value*: Trans-ethnic meta-analysis *P-values*; is more significant: True if Meta-analysis *P* value is more significant than PGC *P value; NA in some of the rows indicate that meta-analysis results did not have any SNPs in those regions.*

.

eTable 9. Risk Profile Scoring–Sign Test *P*-values for SNPs (P_T < .05) Sharing the Same Direction of Effects

MAF (CEU)	Total SNPs	Homogeneity	Sign Test *P* Value
<0.1	2640	1383	0.007483
0.1-0.2	3052	1601	0.003493
0.2-0.3	2623	1406	0.0001202
0.3-0.4	2465	1341	6.71×10^{-6}
0.4-0.5	2183	1201	1.51×10^{-6}
All	12963	6932	1.31×10^{-15}

The table shows the total number of SNPs, the SNPs sharing the same direction of effects, and their respective sign test for all the SNPs included below the *P* value threshold (P_T = 0.05) and MAF ranges <0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.5. For all SNPs below P_T = 0.05, there was an excess of same-direction effects between PGC2 and Indian studies (53.48%, binomial *P* = 1.31×10^{-15}).

eTable 10. BRAINEAC *cis*-eQTL Results for the Top SNP and Genes in Indian Top Locus (100-Kb Window)

Gene	Gene Position (hg19)	SNP ID	SNP Position	Brain eQTL *P* Value
NAPRT1	chr8:144656958-144660799	rs10866912	chr8:144655315	2.90×10^{-13}
ZC3H3	chr8:144519831-144623975	rs10866912	chr8:144655315	3.0×10^{-04}
EEF1D	chr8:144660506-144665462	rs10866912	chr8:144655315	5.10×10^{-03}
TIGD5	chr8:144680015-144683133	rs10866912	chr8:144655315	2.00×10^{-02}
GSDMD	chr8:144635868-144645225	rs10866912	chr8:144655315	2.40×10^{-02}
MROH6	chr8:144648365-144655121	rs10866912	chr8:144655315	2.70×10^{-02}
PYCRL	chr8:144686083-144691764	rs10866912	chr8:144655315	5.40×10^{-02}
TSTA3	chr8:144694794-144718237	rs10866912	chr8:144655315	7.20×10^{-02}
ZNF623	chr8:144718104-144738583	rs10866912	chr8:144655315	9.00×10^{-02}

eTable 11. SMR Results Using the Brain-eMeta eQTL Data Set for Our Top SNP, rs10866912

Gene	topSNP	SNP_Pos	*P*_GWAS	*P*_eQTL	b_SMR	se_SMR	*P*_SMR	*P*_SMR_multi	*P*_HEIDI
GSDMD	rs10866912	8:144655315	4.35×10^{-8}	8.69×10^{-10}	0.237438	0.0598935	7.36×10^{-5}	2.26×10^{-3}	9.68×10^{-5}
RP11-661A12.9	rs10866912	8:144655315	4.35×10^{-8}	1.81×10^{-12}	0.0830525	0.0198564	2.88×10^{-5}	3.65×10^{-3}	2.06×10^{-5}
NAPRT1	rs10866912	8:144655315	4.35×10^{-8}	0.00E+00	0.055346	0.0107506	2.63×10^{-7}	2.96×10^{-3}	3.4×10^{-5}
TIGD5	rs10866912	8:144655315	4.35×10^{-8}	8.89×10^{-16}	0.119777	0.0274438	1.27×10^{-5}	2.85×10^{-4}	3.61×10^{-11}

P: *P*-value; *P*_GWAS: the *P*-value for top SNP in the Indian GWAS; *P*_eQTL; Brain-eMeta eQTL *P*-value for top SNP; b_SMR: Beta value for SMR association; se_SMR: Standard Error for SMR association; *P*_SMR: *P*-value for SMR association; *P*_SMR_multi: *P*-value for SMR multi association *P*_HEIDI: *P*-value for HEIDI test

eTable 12. Indian Replication Results for PGC2 Top Loci (96/108)

PGC2 Top Regions (hg19)	Indian Top SNP	Chr	BP (hg19)	Z-Score	*P* Value
6_28684183_28729994	rs5875167	6	28703608	1.403	0.1607
1_98491248_98512127	rs11404556	1	98494388	-1.829	0.06736
10_104612335_104628873	10-104619327	10	104619327	2.023	0.04304
12_2344960_2344960	12-2344960	12	2344960	-1.535	0.1248
8_143312933_143316970	rs67498679	8	143316849	-0.232	0.8167
7_2017445_2048220	rs3996329	7	2018870	1.326	0.1848
5_60589739_60614879	rs113807000	5	60600123	-2.109	0.03491
12_123665113_123746961	12-123725688	12	123725688	-2.1	0.03575
15_91426560_91426560	rs4702	15	91426560	0.744	0.457
3_36858583_36858583	rs75968099	3	36858583	2.316	0.02056
14_104021141_104053764	rs12889403	14	104034746	-2.094	0.03627
15_78883813_78912943	rs62010330	15	78905854	0.67	0.5031
7_110851553_110957151	rs37752	7	110921994	-1.085	0.2778
11_130717153_130729430	rs61435928	11	130719075	1.873	0.061
2_185601420_185663304	rs10172501	2	185630341	2.808	0.004987
10_18734528_18751891	rs10764512	10	18745202	-2.603	0.009238
1_73766431_73884838	rs1923232	1	73853042	-1.387	0.1655
2_233592501_233707889	rs112967383	2	233606144	1.292	0.1964
11_124612932_124613957	rs12293670	11	124612932	0.387	0.699
18_53200117_53200117	rs9636107	18	53200117	1.964	0.0495
11_46350213_46350213	rs61126341	11	46350213	-0.43	0.6673
3_180592360_180700150	rs11411529	3	180594593	-2.142	0.03219
20_37422829_37457184	rs10645870	20	37450625	-2.392	0.01673
2_57961602_57998040	rs72804548	2	57974646	1.608	0.1079
15_84703470_84706461	rs950169	15	84706461	-0.803	0.4217
18_53533189_53568458	18-53558587	18	53558587	-1.677	0.0936
2_198248128_198304577	2-198295828	2	198295828	-1.86	0.06284
22_41587556_41587556	rs9607782	22	41587556	2.318	0.02046
8_111471166_111511043	rs143086783	8	111508076	1.891	0.05861
3_2547786_2554612	rs749613	3	2552565	-0.906	0.365
11_113317745_113392994	rs77655590	11	113337446	-2.231	0.02568
11_133822569_133830067	rs694424	11	133824539	1.297	0.1948
3_52833805_52847601	rs146326248	3	52837282	0.808	0.4191
16_29924422_29978827	rs11150575	16	29935066	1.3	0.1934
3_136165695_136425514	3-136352543	3	136352543	1.748	0.08049
5_152162776_152177121	rs4616882	5	152173139	3.068	0.002155
17_2207236_2209888	rs57130114	17	2209002	2.138	0.03255
7_86420126_86427626	rs13230421	7	86420126	-2.072	0.03826
15_61831680_61879179	15-61872804	15	61872804	1.437	0.1507
1_44078384_44107428	rs603542	1	44079411	2.873	0.004062
19_19478022_19484295	rs2905426	19	19478022	-2.011	0.0443
1_149999059_150134221	rs12022725	1	150000165	1.684	0.09211
1_2383039_2402499	rs4648845	1	2387101	-1.841	0.06555
16_13746548_13761333	rs34646006	16	13758253	1.326	0.185
7_104812129_105031108	rs13226540	7	104999901	2.348	0.01889
1_8421203_8424984	rs13596	1	8421203	2.836	0.004575
12_110723245_110723245	12-110723245	12	110723245	1.527	0.1267
4_170551373_170646003	rs6834147	4	170563153	1.204	0.2286
22_42603814_42603814	rs6002655	22	42603814	0.744	0.4572
2_146419047_146440672	rs6430095	2	146439945	1.759	0.0786
11_57409795_57644334	rs515939	11	57432629	-1.285	0.1989
11_24384697_24410865	rs142359455	11	24406704	-1.31	0.19
1_30428943_30437268	rs34418819	1	30432824	1.322	0.1862
7_137048159_137085250	rs1731947	7	137081525	-1.476	0.14

PGC2 Top Regions (hg19)	Indian Top SNP	Chr	BP (hg19)	Z-Score	*P* Value
9_84736303_84744273	rs6559650	9	84741407	0.985	0.3245
15_40566759_40569884	rs4924443	15	40567367	2.023	0.04307
19_30987423_30987423	rs2053079	19	30987423	-1.21	0.2264
5_88744550_88746331	rs1820170	5	88746203	-1.824	0.06811
3_17849789_17886678	rs2033379	3	17851238	-1.839	0.06591
5_137851192_137948140	rs1864980	5	137947101	-2.439	0.01472
14_99719219_99719219	rs2693698	14	99719219	1.048	0.2949
14_72417326_72417326	rs2332700	14	72417326	0.74	0.4594
5_45298611_45393754	rs10064595	5	45332612	-1.066	0.2862
8_60691526_60750473	rs34979222	8	60723658	1.533	0.1253
2_200162425_200164252	rs7605371	2	200162766	1.017	0.3091
2_193848340_193848340	rs59979824	2	193848340	-3.378	0.0007298
4_176861301_176866459	rs1106568	4	176861301	-0.859	0.3906
8_4178791_4183057	rs17069921	8	4182101	1.258	0.2085
2_225334070_225467840	rs72617131	2	225419922	1.728	0.08393
8_89535302_89601210	rs117535734	8	89552192	-2.105	0.03528
16_9937980_9959121	rs6497549	16	9950374	-1.791	0.07324
14_30190316_30190316	rs2068012	14	30190316	1.035	0.3008
3_63792668_63903759	rs10715557	3	63847469	-1.484	0.1378
16_68061171_68293320	rs77884900	16	68190239	2.223	0.02624
2_149049630_149873154	rs71406028	2	149458259	2.897	0.003762
17_17883848_17972973	rs2955368	17	17961349	0.294	0.7689
15_70573896_70595477	rs11306388	15	70580499	-2.254	0.02421
16_58669908_58682833	rs12325245	16	58681393	-1.569	0.1167
8_27412605_27442127	rs2741348	8	27420782	1.34	0.1801
6_73154800_73157926	rs4147060	6	73156139	-1.278	0.2013
7_24747494_24798639	rs2521769	7	24784485	-2.119	0.03409
5_109030041_109141527	rs3753174	5	109107937	1.052	0.2927
4_23423586_23423603	rs215411	4	23423603	-0.639	0.5229
5_153675891_153686366	5-153677362	5	153677362	-0.855	0.3927
11_109378071_109392839	rs79307239	11	109386518	-1.004	0.3152
7_110039196_110072128	rs67820402	7	110058388	-1.446	0.1483
12_29916839_29928388	12-29918582	12	29918582	-2.083	0.03724
7_131567263_131567263	rs7801375	7	131567263	1.198	0.2311
1_177280121_177300809	rs3066140	1	177288341	1.974	0.04836
1_207977083_207977083	rs7523273	1	207977083	1.28	0.2007
20_48114678_48131649	rs1393343	20	48115247	-1.456	0.1454
12_92246786_92257509	12-92252781	12	92252781	-1.512	0.1304
2_162802184_162845855	rs146357347	2	162809963	1.873	0.06107
19_50067508_50103252	rs5023763	19	50072067	1.988	0.04678
12_103561799_103608071	12-103563372	12	103563372	-0.625	0.5319
5_140136468_140221139	rs3806843	5	140212538	2.51	0.01206

Chr: Chromosome; BP: Base pair (hg19)

eReferences

1. Thara R, Srinivasan T, John S, et al. Design and Clinical Characteristics of a Homogeneous Schizophrenia Pedigree Sample from Tamil Nadu, India. *Aust N Z J Psychiatry.* 2009;43(6):561-570.
2. Holliday EG, Nyholt DR, Tirupati S, et al. Strong evidence for a novel schizophrenia risk locus on chromosome 1p31.1 in homogeneous pedigrees from Tamil Nadu, India. *Am J Psychiatry.* 2009;166(2):206-215.
3. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82.
4. Lee SH, DeCandia TR, Ripke S, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet.* 2012;44(3):247-250.
5. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet.* 2016;98(1):127-148.
6. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
7. Gao F, Chang D, Biddanda A, et al. XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome. *Journal of Heredity.* 2015;106(5):666-671.
8. Schork AJ, Won H, Appadurai V, et al. A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci.* 2019;22(3):353-361.
9. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.
10. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet.* 2013;14(7):507-515.
11. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2015;31(9):1466-1468.
12. Falconer DS. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet.* 1967;31(1):1-20.
13. Kichaev G, Roytman M, Johnson R, et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics.* 2017;33(2):248-255.
14. Kichaev G, Yang WY, Lindstrom S, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics.* 2014;10(10):e1004722.
15. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics.* 2017;33(4):594-595.
16. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics.* 2018;208(4):1397-1408.
17. Trabzuni D, Ryten M, Walker R, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem.* 2011;119(2):275-282.
18. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

19. Fromer M, Roussos P, Sieberts SK, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci.* 2016;19(11):1442-1453.
20. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature.* 2011;478(7370):483-489.
21. Brown BC, Asian Genetic Epidemiology Network Type 2 Diabetes C, Ye CJ, Price AL, Zaitlen N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am J Hum Genet.* 2016;99(1):76-88.
22. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48(5):481-487.
23. Qi T, Wu Y, Zeng J, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun.* 2018;9(1):2282.
24. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology.* 2015;11(4):e1004219.
25. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat Meth.* 2016;13(4):366-370.
26. Giacomotto J, Rinkwitz S, Becker TS. Effective heritable gene knockdown in zebrafish using synthetic microRNAs. *Nat Commun.* 2015;6:7378.
27. Laird AS, Mackovski N, Rinkwitz S, Becker TS, Giacomotto J. Tissue-specific models of spinal muscular atrophy confirm a critical role of SMN in motor neurons from embryonic to adult stages. *Hum Mol Genet.* 2016;25(9):1728-1738.
28. Machiela MJ, Chanock SJ. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics.* 2017.
29. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool.* 2005;2:16.
30. Murray RM, Bhavsar V, Tripoli G, Howes O. 30 Years on: How the Neurodevelopmental Hypothesis of Schizophrenia Morphed Into the Developmental Risk Factor Model of Psychosis. *Schizophr Bull.* 2017;43(6):1190-1196.