Supplementary Material

A meta-analysis of gene expression quantitative train loci in brain*, *Kim et al.

Table of Contents

Supplemental Methods

Cortical gene expression data

We included brain cortical samples from neuropathologically normal subjects of European ancestry and aged ≥ 20 years at death. Below is definition of "neuropathologically normal" found from each study.

- In Colantuoni et al., it was defined "without neuropathological or neuropsychiatric diagnosis".
- In Gibbs et al., it was defined "neurologically normal".
- In Myers et al., it was defined "no Lewy bodies or co-morbidity with any other known neurological disease".
- In Stanley Medical Research Institute (SMRI), we selected "control" samples.
- In the NIH GTEx project, it was defined as following: "donors of either sex from any ancestry group with aged 21-70 were collected. Medical exclusionary criteria include HIV infection, high-risk behaviors, viral hepatitis, metastatic cancer, chemotherapy or radiation therapy for any condition within the past 2 years, and whole-blood transfusion in the past 48 hours or BMI of >35 or <18.5." We used 24 pre-frontal cortex samples from 2012 December freeze data.

We obtained raw expression intensity files from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), the Stanley Medical Research Institute project (SMRI, http://www.stanleygenomics.org), and the Genotype-Tissue Expression (GTEx, http://www.gtexportal.org). We conducted extensive and consistent quality control procedures for all expression data. As in our prior work, we processed gene expression data as consistently as possible across studies.[1] First, for all gene expression arrays, we mapped probe sequences to the genome (UCSC hg19) using Bowtie[2] and removed probes that did not map, mapped to multiple locations, or intersected a polymorphic SNP (HapMap3[3] and 1000 Genomes Project data[4]) because such probes can result in inaccurate expression.[5] Second, we excluded outlier samples based on inter-sample correlations[6] or if phenotypic sex did not match mean expression of probesets on chrX and chrY. Third, we used hierarchical clustering (R function hclust, average link function) and principal component analysis (PCA) in order to evaluate overall performance of all arrays. Below is a summary table showing data sources, brain regions, and quality control steps for all studies.

Feature		Colantuoni [7]	Gibbs [8]	GTEx [9]	Myers [10]	SMRI[11,12]
Data source		GSE30272	GSE15745	GTEx (2012 Dec freeze)	GSE8919	Arrays from Altar group
Sample size		56	124	24	189	31
Cortical region		Frontal	Frontal	Frontal	Frontal, temporal	Frontal
Technology		IL Human 49K Oligo array	IL HumanRef-8	RNA-seq (IL HiSeq 2000)	IL HumanRef-8	AF HG-U133a
Normalization (study-specific)		Background correction, normalization using loess and then using quantile method in limma	Background correction, $\log_2$ transformation, normalization using quantile method in limma	Gene read counts, normalization, [a] gene exclusion [b]	Used processed data as raw data not available	Robust multiarray average (RMA) from Affymetrix Power Tools (v1.12.0)
QC (common)	Probe mapping [c]	✓	✓	✓		✓
	Outlier exclusion [d]	✓	✓	✓	✓	✓

[a] We started with read counts per gene and normalized samples by dividing scale factors that were calculated by the ratio of 75% quantile read count in each sample over the mean among all the samples. Read count data were inverse-quantile normalized within each sample. [b] Genes were excluded if > 90% of samples had read counts < 1. [c] After mapping probe sequences to hg19 using Bowtie, we removed probes with sequences not mapping, mapping to multiple genomic locations, or intersecting a polymorphic SNP. [d] We removed outlier samples based on clustering, principal component analysis, inter-array correlations[6], and sex inconsistency. Abbreviations: IL=Illumina, AF=Affymetrix.

Genotype data

We obtained genotype data in PLINK format from dbGaP (http://www.ncbi.nlm.nih.gov/gap), as raw intensity files from SMRI (http://www.stanleygenomics.org) and from an author's website (http://labs.med.miami.edu/myers). For raw intensity files, we used Affymetrix Power Tools to call genotypes. We applied extensive and consistent quality control procedures described in the table below.

Feature	Colantuoni [7]	Gibbs [8]	GTEx [9]	Myers [10]	SMRI[11,12]
Data source	dbGaP:pht002399.v2.p1	dbGaP:pht001879.v1.p1	dbGaP:phs000424	http://labs.med.miami.edu/myers	Arrays from Altar group
Sample size	113	148	183	193	93
Tissue	Cerebellum	Cerebellum	Blood	Cortex	Frontal cortex
Technology	IL Hap650Yv3.0	IL Hap550v3.0	IL Omni5-Quad	AF 500K	AF 5.0
SNP exclusion criteria	Call rate < 90% MAF < 1% HWE < 1e-8 Allele frequency inconsistency with 1000G reference panel				
Sample exclusion criteria	Call rate < 90% Sex mismatch Relatedness Relative high IBD value with too many other subjects Ancestry outlier				
Sample size after QC	108	148	155	191	93

Imputation of genotype data

We downloaded EUR data from the 1000G reference panel (phase1 release v3.20101123) from the MACH website. Prior to imputation, we performed initial QC in each study. We excluded samples that were related ($\hat{\pi} > 0.04$), moderate relatedness with many subjects (indicating contamination or poor quality), low call-rate (< 0.9), inconsistent phenotype-genotype sex, and ancestry outliers detected by principal component analysis. We removed SNPs with low call rate (< 0.9), low minor allele frequency (< 0.01), deviation from Hardy-Weinberg equilibrium ($P_{HWE} < 1x10^{-8}$), and allele frequencies inconsistent with the 1000 Genomes Project reference panel (minor allele frequency difference < 0.07).

For each study, we imputed genotype dosages using the 1000 Genomes Project reference panel using MaCH-admix.[13] MaCH-admix does not require phasing before imputation, which is suitable for small studies. ChrX was imputed in males using the option in MaCH-admix and females were imputed in the same way as autosomes. In order to evaluate the quality of imputed SNP, we computed average R^2 under varying MAF categories and R^2 cutoffs (see the summary tables below), which is similar to a procedure in a prior study.[14] We retained SNPs with imputation $R^2 \geq 0.3$ and minor allele frequency greater than max(0.05, $\frac{5}{2N}$).

Colantuoni et al.	R^2 cutoffs						
MAF	0.3	0.35	0.4	0.45	0.5	0.55	0.6
0.01 – 0.03	0.71	0.74	0.76	0.79	0.81	0.83	0.85
0.03 – 0.05	0.80	0.82	0.83	0.85	0.86	0.88	0.89
0.05 - 0.5	0.90	0.90	0.91	0.92	0.92	0.92	0.93

Gibbs et al.	R^2 cutoffs						
MAF	0.3	0.35	0.4	0.45	0.5	0.55	0.6
0.01 – 0.03	0.70	0.73	0.76	0.78	0.80	0.82	0.84
0.03 – 0.05	0.79	0.80	0.82	0.84	0.85	0.86	0.88
0.05 - 0.5	0.89	0.90	0.90	0.91	0.91	0.92	0.92

GTEx	R^2 cutoffs						
MAF	0.3	0.35	0.4	0.45	0.5	0.55	0.6
0.01 – 0.03	0.74	0.76	0.79	0.81	0.83	0.85	0.87
0.03 – 0.05	0.82	0.84	0.85	0.87	0.88	0.90	0.91
0.05 - 0.5	0.92	0.92	0.93	0.93	0.94	0.94	0.95

Myres et al.	R^2 cutoffs						
MAF	0.3	0.35	0.4	0.45	0.5	0.55	0.6
0.01 – 0.03	0.66	0.69	0.72	0.74	0.77	0.79	0.81
0.03 – 0.05	0.71	0.74	0.76	0.78	0.80	0.82	0.84
0.05 - 0.5	0.81	0.83	0.84	0.85	0.86	0.87	0.88

SMRI	R^2 cutoffs						
MAF	0.3	0.35	0.4	0.45	0.5	0.55	0.6
0.01 – 0.03	0.69	0.72	0.75	0.77	0.80	0.82	0.84
0.03 – 0.05	0.74	0.76	0.78	0.80	0.82	0.84	0.86
0.05 - 0.5	0.81	0.83	0.84	0.85	0.86	0.87	0.88

Study Name	Number of SNPs retained after QC
Colantuoni et al.	7,887,133
Gibbs et al.	7,874,224
GTEx	7,780,852
Myers et al.	6,391,052
SMRI	6,779,110

Evaluation of covariates for gene expression

We selected covariates for gene expression in two steps. First, we assessed all covariates by computing the type III sum of squares (SAS, v9.2) which compares a full model containing all covariates to a

reduced model that excludes a covariate under consideration. The impact of a covariate was quantified by determining the number of genes with FDR < 0.05. We included a covariate when > 1% of genes met this criterion. Second, we regressed out selected covariates and performed PCA on the residuals. The final covariate list included covariates from the first step and the significant PCs from the second step. For the genotype data from each study, we included the first PC. Below are summary tables showing number of significant genes from the first step of type III sum of squares test (q < 0.05).

Colantuoni et al. 14,914 genes	Type	df	Significant genes
Age	Numeric	1	10
pmi	Numeric	1	0
sex	Factor	1	184
pH	Numeric	1	754
RNA integrity number	Numeric	1	3104
batch	Factor	13	6298

Gibbs et al. 16,554 genes	Type	df	Significant genes
Age	Numeric	1	178
pmi	Numeric	1	1
sex	Factor	1	33
Tissue bank	Factor	2	1718
batch	Factor	13	3247

GTEX, 25,287 genes	Type	df	Significant genes
Age	Numeric	1	0
Height	Numeric	1	0
Weight	Numeric	1	0
BMI	Numeric	1	0
Donor ischemic time (min)	Numeric	1	0
race	Factor	1	2
sex	Factor	1	0
Dthhrdy (hardy scale)	Factor	3	0

Myers et al. 15,857 genes	Type	df	Significant genes
Age	Numeric	1	651
pmi	Numeric	1	0
sex	Factor	1	21
Average transcript detection rate (all 24354 probes)	Numeric	1	0
Average transcript detection rate (14078 brain expressed probes)	Numeric	1	0
Hybridization date	Factor	8	8442
Brain bank	factor	17	5416

SMRI, 10,038 genes	Type	df	Significant genes
Age	Numeric	1	0
pmi	Numeric	1	0
sex	Factor	1	8
pH	Numeric	1	157
Brain side	Factor	1	0

Below are summary tables showing number of significant genes from the second step of type III sum of squares test. The PCs below are computed from the expression data after residualizing out the effects of impactful covariates.

Colantuoni et al.	Type	df	Significant genes
sex	Factor	1	3555
pH	Numeric	1	5046
RNA integrity number	Numeric	1	5310
batch	Factor	13	11243
PC1	Numeric	1	6471
PC2	Numeric	1	6579
PC3	Numeric	1	3418
PC4	Numeric	1	2940
PC5	Numeric	1	1150

Gibbs et al.	Type	df	Significant genes
Age	Numeric	1	1816
Tissue bank	Factor	2	5890
batch	Factor	13	8186
PC1	Numeric	1	9549
PC2	Numeric	1	7426
PC3	Numeric	1	6940
PC4	Numeric	1	4436
PC5	Numeric	1	6119

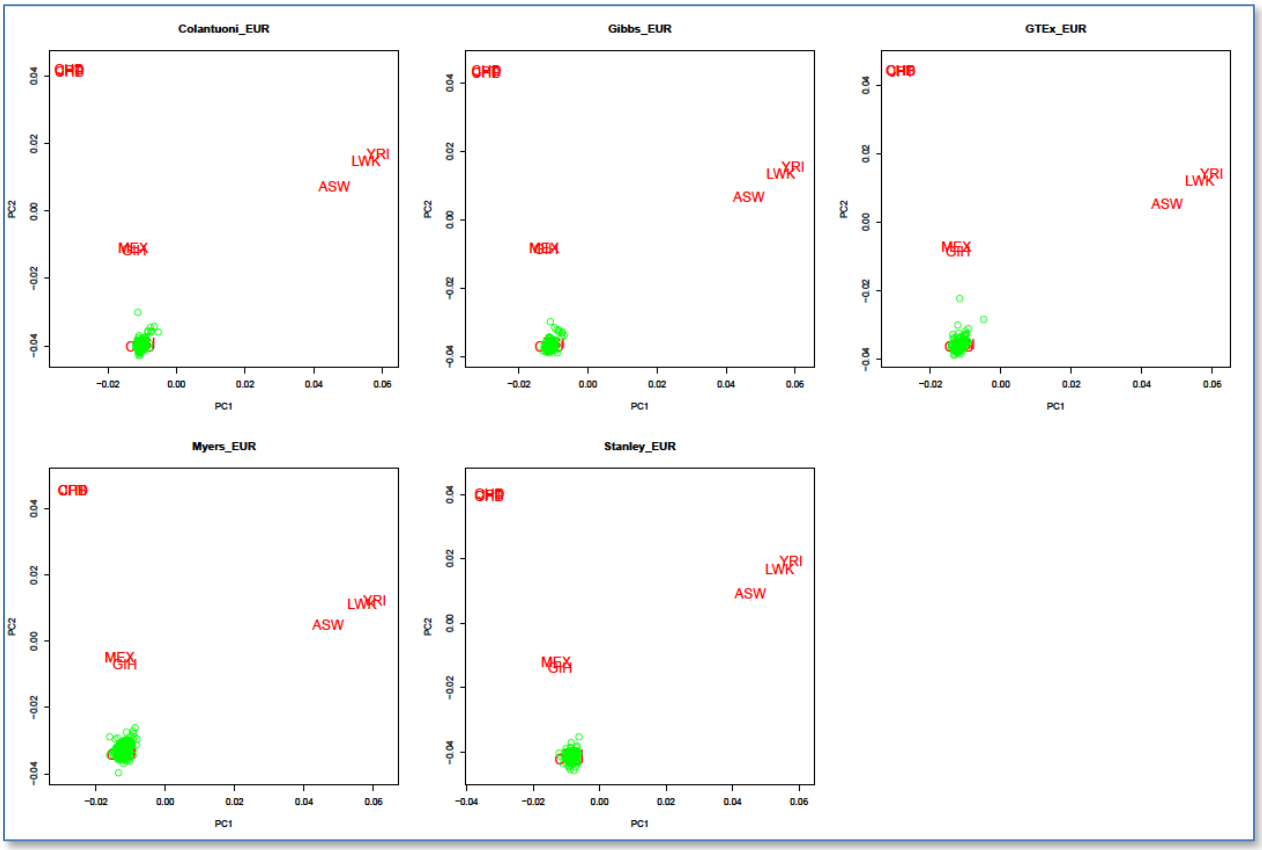
GTEx	Type	df	Significant genes
PC1	Numeric	1	12420
PC2	Numeric	1	9074
PC3	Numeric	1	7798
PC4	Numeric	1	3489
PC5	Numeric	1	1581

Myers et al.	Type	df	Significant genes
Age	Numeric	1	1157
Hybridization date	Factor	8	9826
Brain bank	factor	17	7302
PC1	Numeric	1	11328
PC2	Numeric	1	0
PC3	Numeric	1	0
PC4	Numeric	1	0
PC5	Numeric	1	0

SMRI	Type	df	Significant genes
pH	Numeric	1	5617
PC1	Numeric	1	7238
PC2	Numeric	1	5154
PC3	Numeric	1	2523
PC4	Numeric	1	988
PC5	Numeric	1	1118

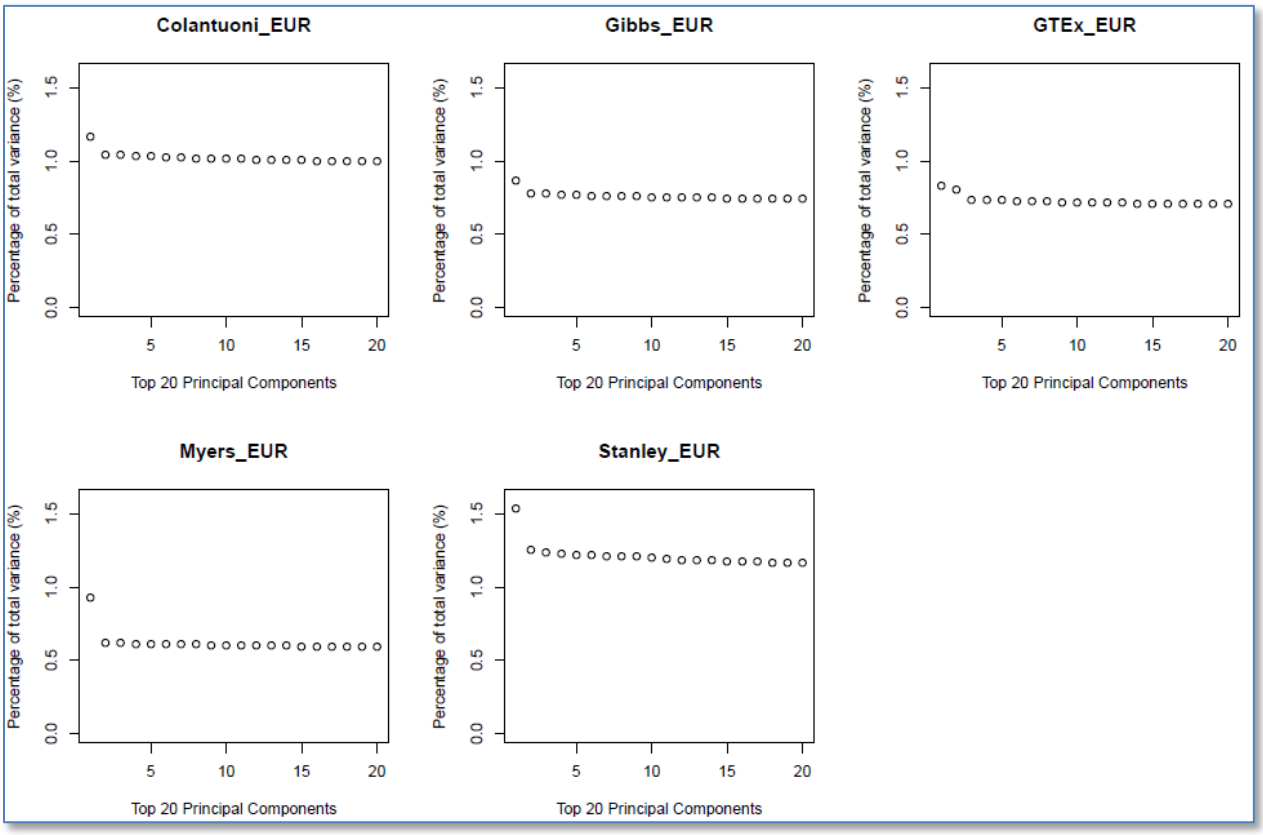
Supplemental Figures

Figure S1. PC1 vs PC2 with HapMap3 samples



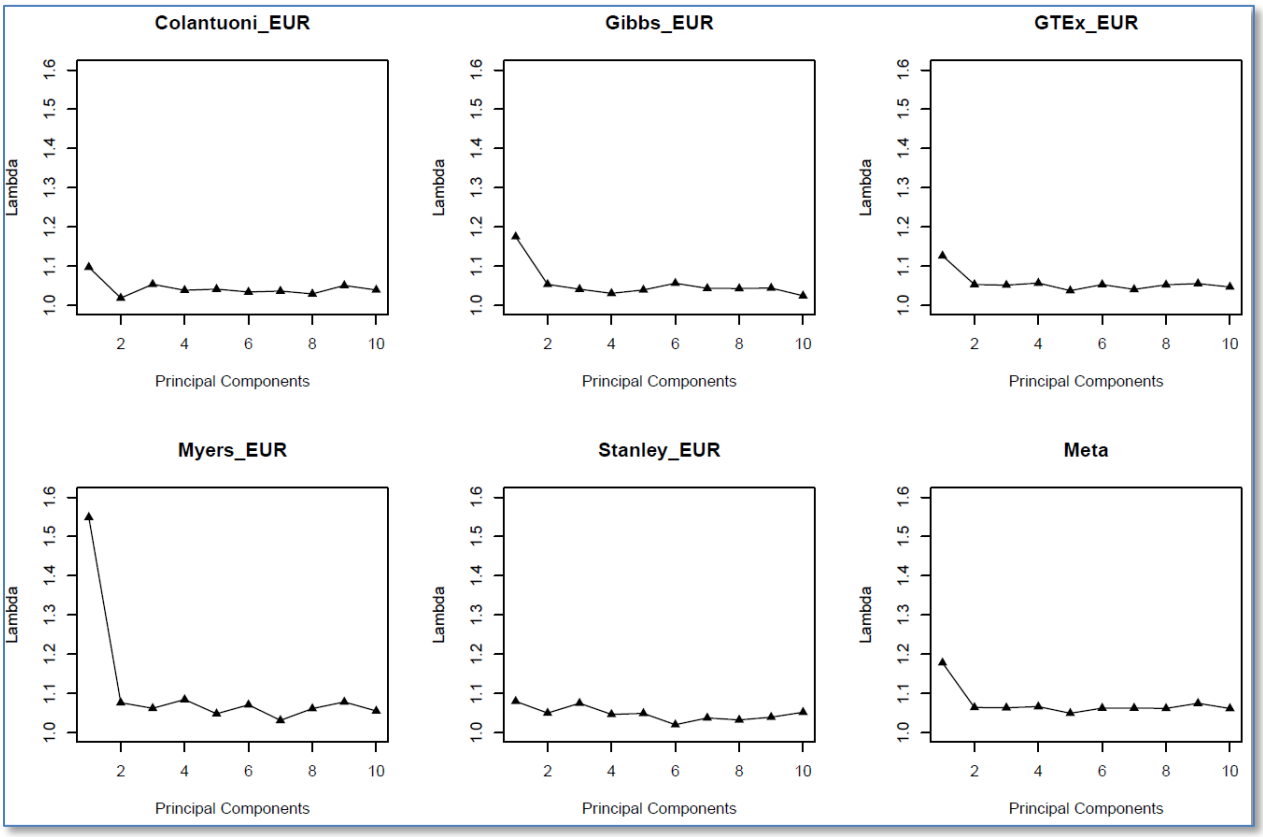
Green points indicate study samples remained after QC. Abbreviations: ASW (African ancestry in Southwest USA), CEU (Utah residents with Northern and Western European ancestry from CEPH collection), CHB (Han Chinese in Beijing, China), CHD (Chinese in Metropolitan Denver, Colorado), GIH (Gujarati Indians in Houston, Texas), JPT (Japanese in Tokyo, Japan), LWK (Luhya in Webuye, Kenya), MEX (Mexican ancestry in Los Angeles, California), MKK (Maasai in Kinyawa, Kenya), TSI (Toscans in Italy), YRI (Yoruba in Ibadan, Nigeria)

Figure S2. PCA scree plots



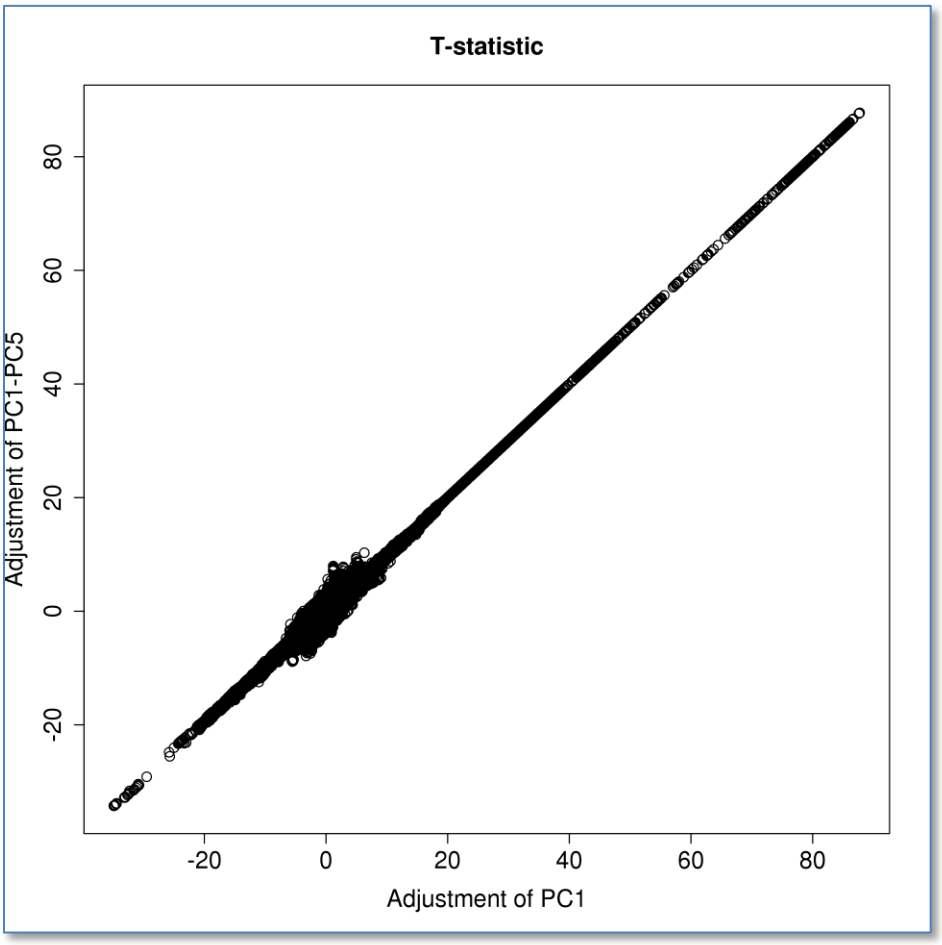
Scree plots show the proportion of total variance explained by each PC.

Figure S3. Scree plots of lambda versus PCs



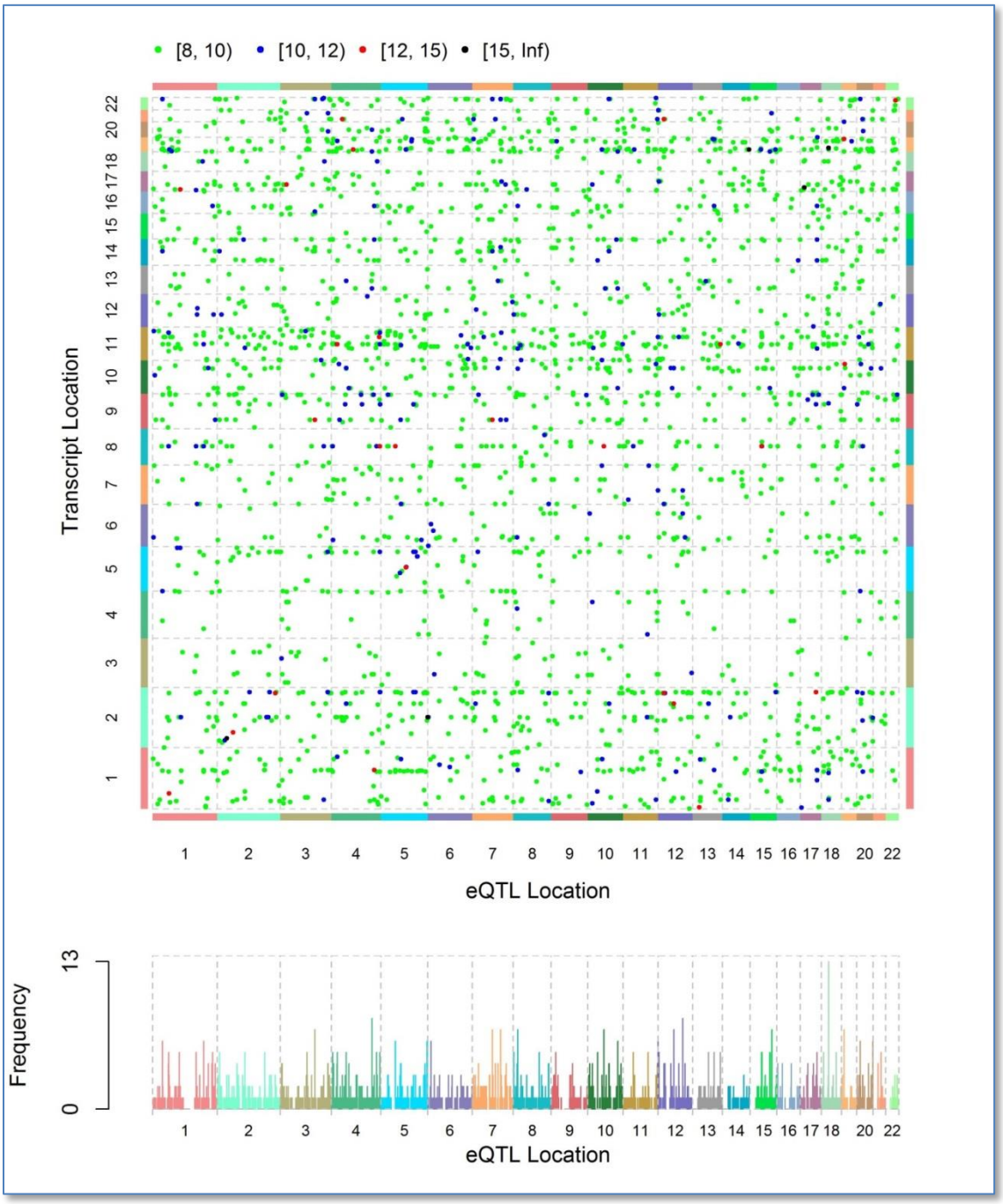
We tested genome-wide association between each genomic PC as a dependent variable and SNPs as independent variables and calculated the genomic inflation factor (λ) from each association test. It shows the contribution of each PC to λ.

Figure S4. T-statistics from adjustment of genomic PC1 vs PC1-5



We evaluated whether adjustment of genomic PC1 should be enough or several more PCs may be needed to control for population sub-stratification. For the latter approach, we included PC1-PC5 in eQTL analysis for each study and meta-analyzed eQTLs using fixed effect model. A scatter plot of t-statistics from the two approaches show that the magnitude and direction are almost identical (Pearson's correlation=0.98).

Figure S5. Local and distant eQTLs in Colantuoni et al.



Upper panel: each dot represents –log10(p-value) for SNP-gene eQTLs. Lower panel: eQTL hotspots, the number of significant eQTLs per SNP that pass the highest cutoff (eQTL with black color in upper panel).

Figure S6. Local and distant eQTLs in Gibbs et al.

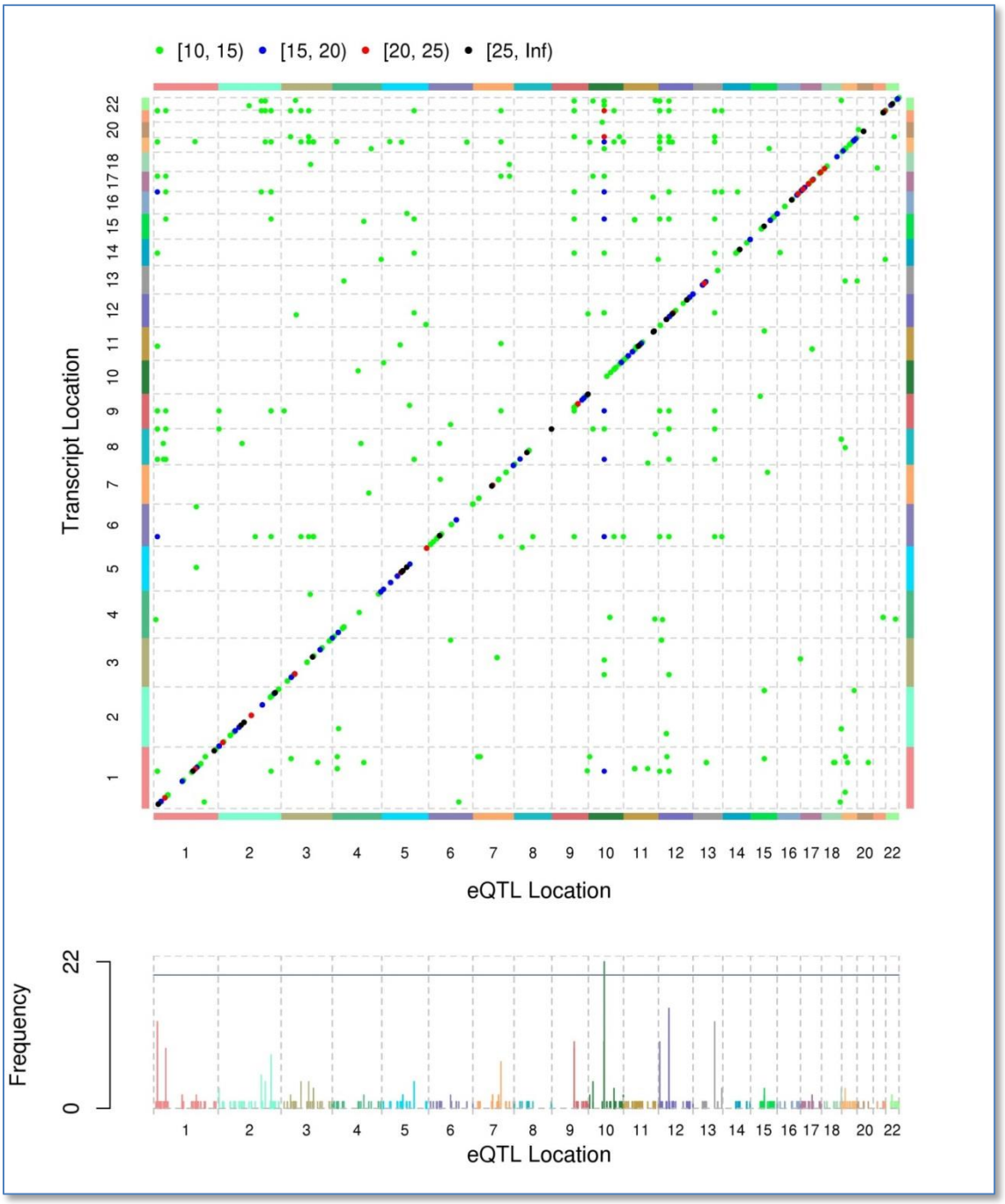


Figure legend is same as in Figure S6.

Figure S7. Local and distant eQTLs in GTEx

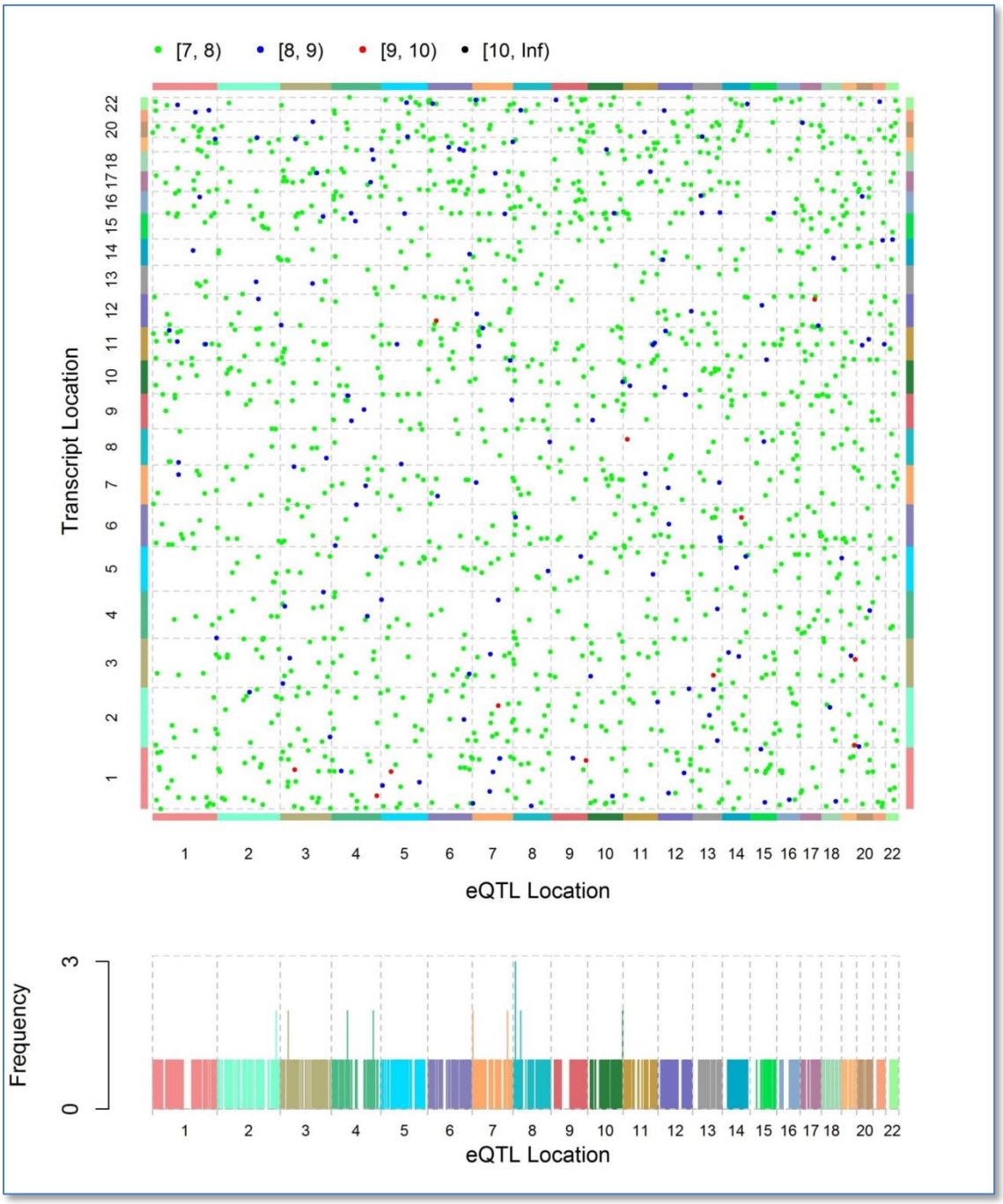


Figure legend is same as in Figure S6.

Figure S8. Local and distant eQTLs in Myers et al.

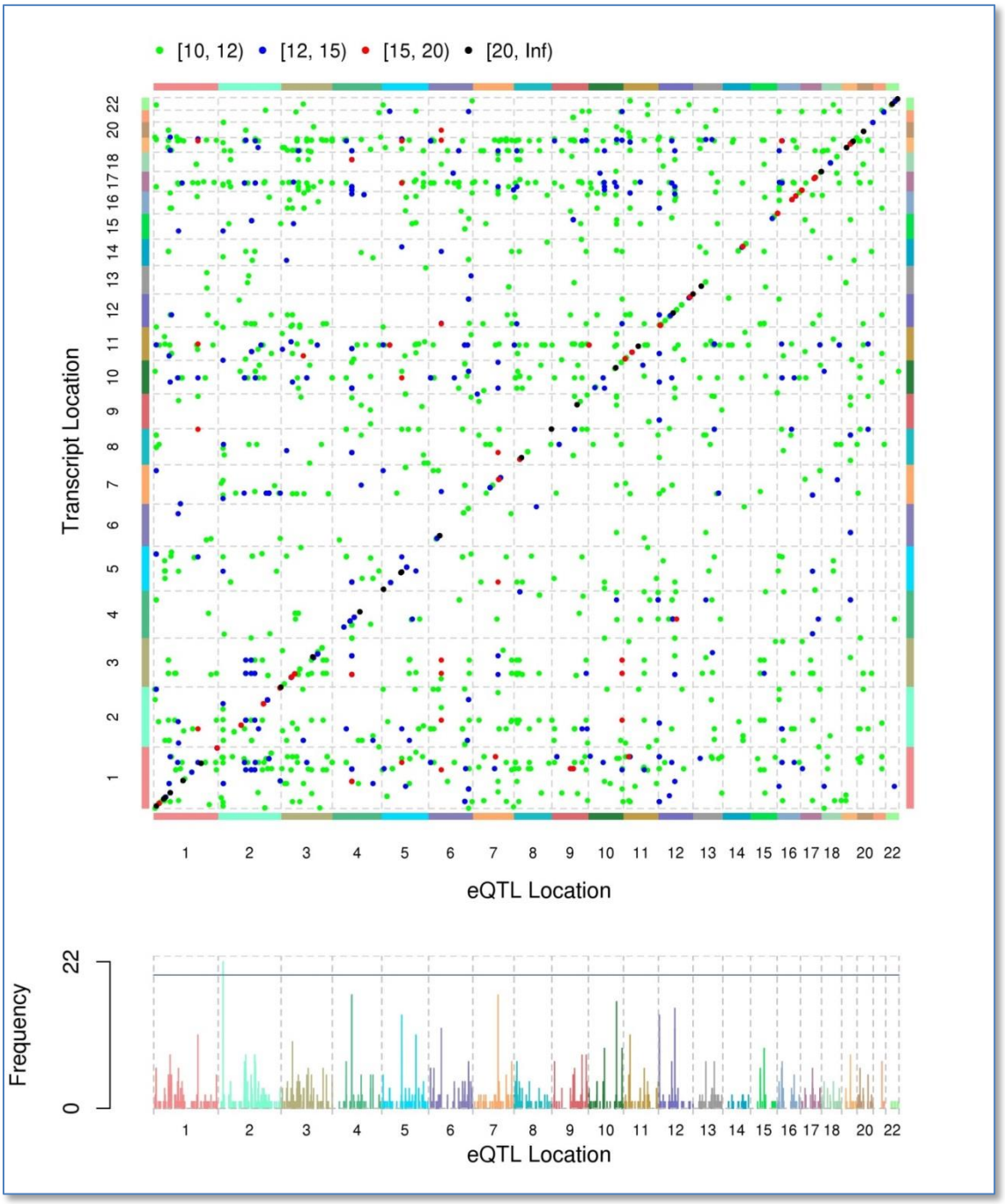


Figure legend is same as in Figure S6.

Figure S9. Local and distant eQTLs in SMRI

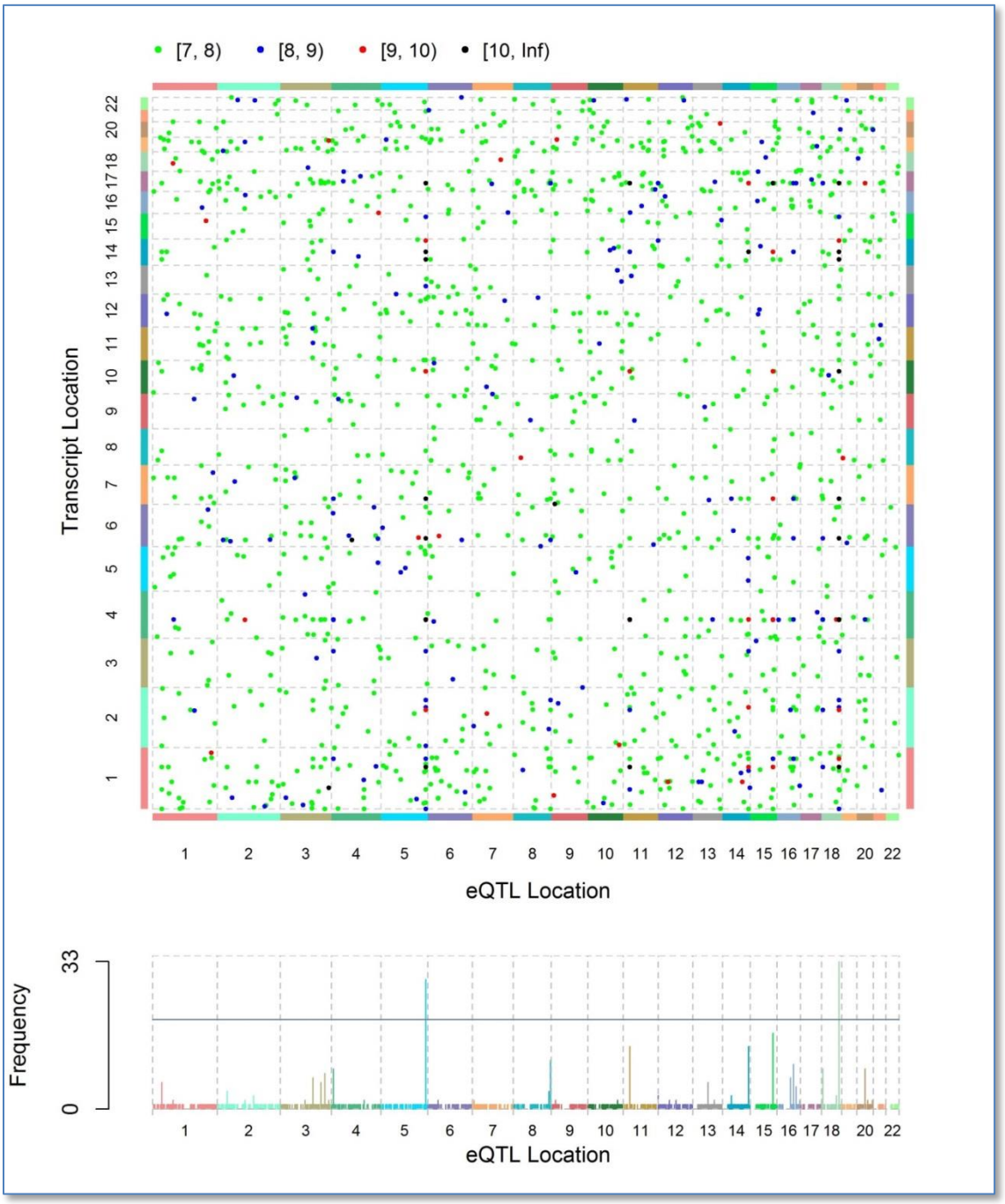
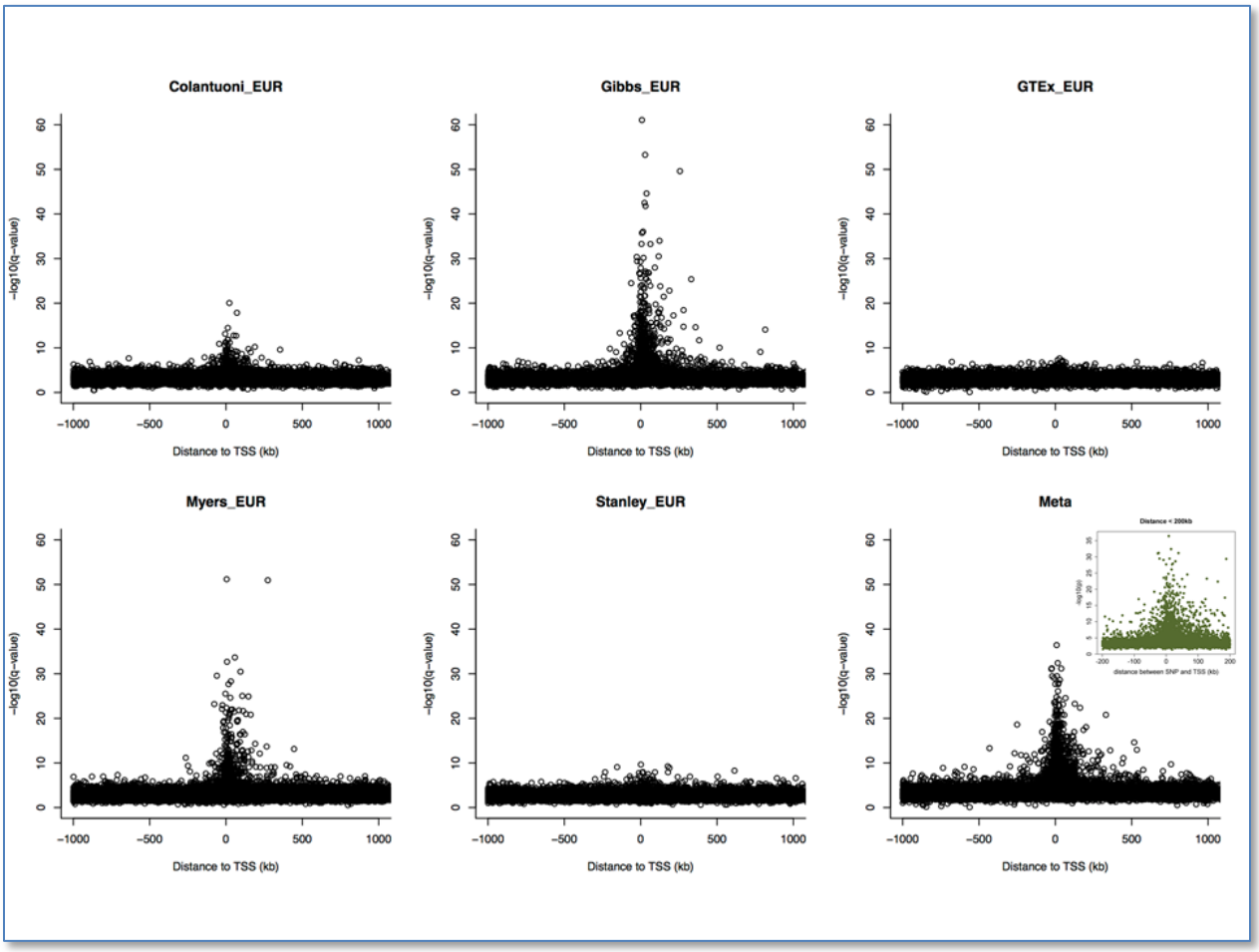


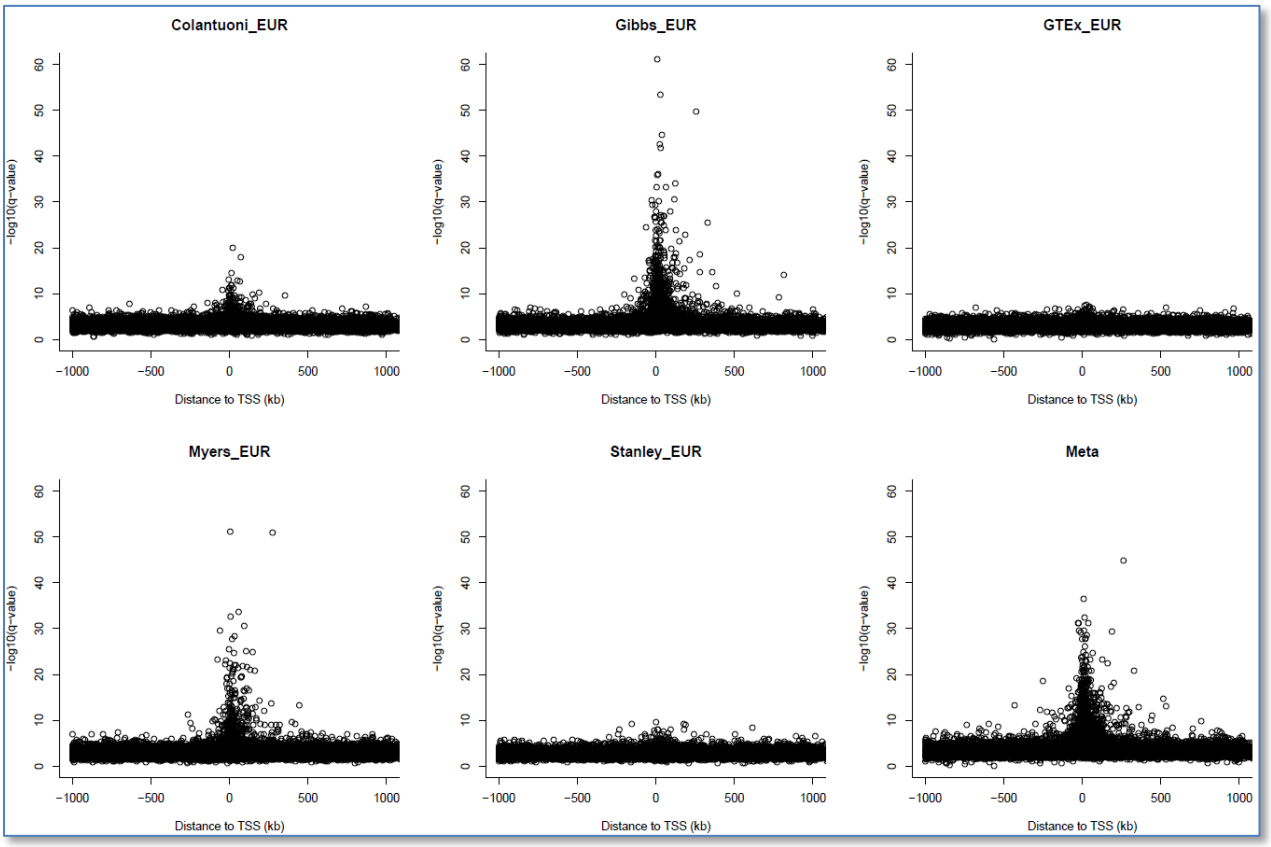
Figure legend is same as in Figure S6.

Figure S10. eQTL distance to transcription start site (genomic PC1)



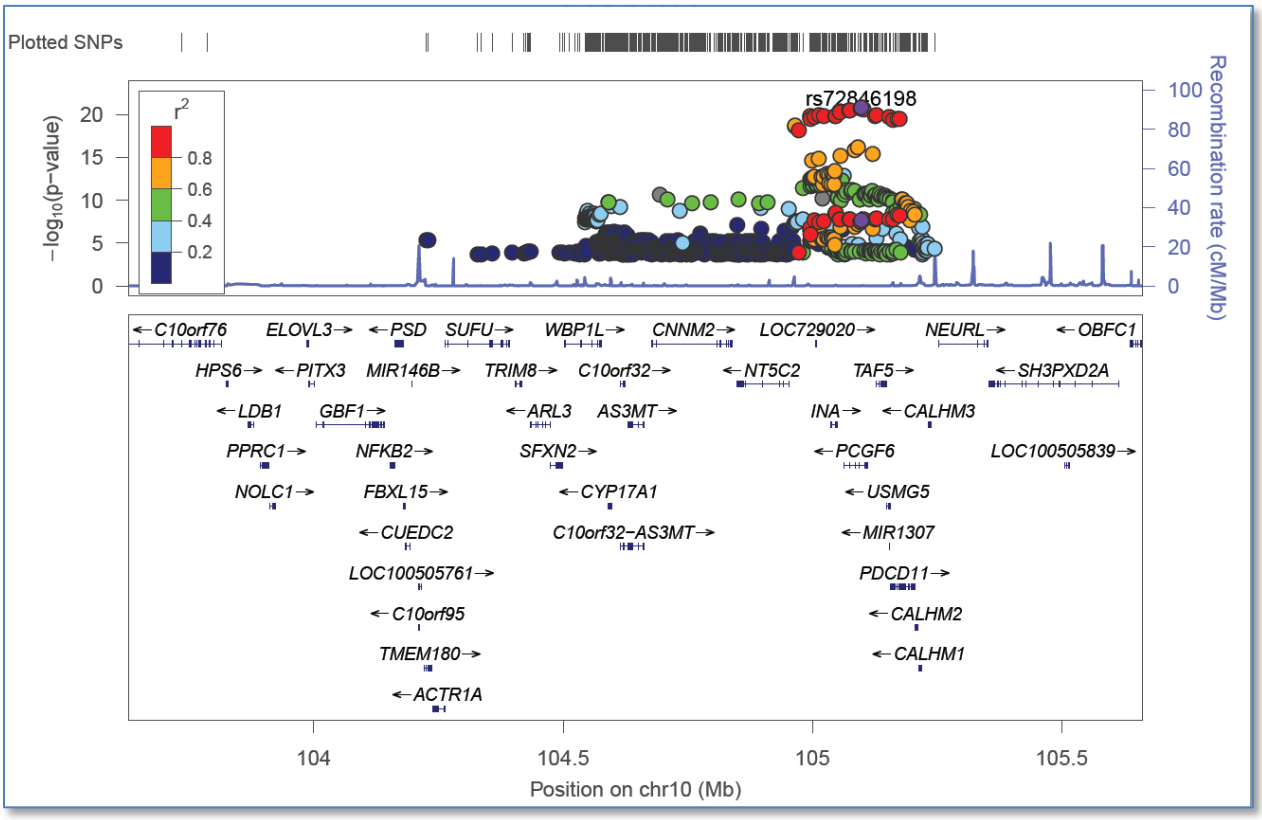
Shown are the distribution of distances and significance between genes and their most associated eQTL SNPs when genomic PC1 was included in eQTL analysis. Each point represents the most significant eQTL SNP per gene. X-axis shows the distance between the transcription start site of gene and the SNP location. Y-axis indicates -$\log_{10}$(q-value). The inset depicts -$\log_{10}$(q-value) of meta-analysis results when distances to TSS are < 200kb.

Figure S11. eQTL distance to transcription start site (genomic PC1-5)



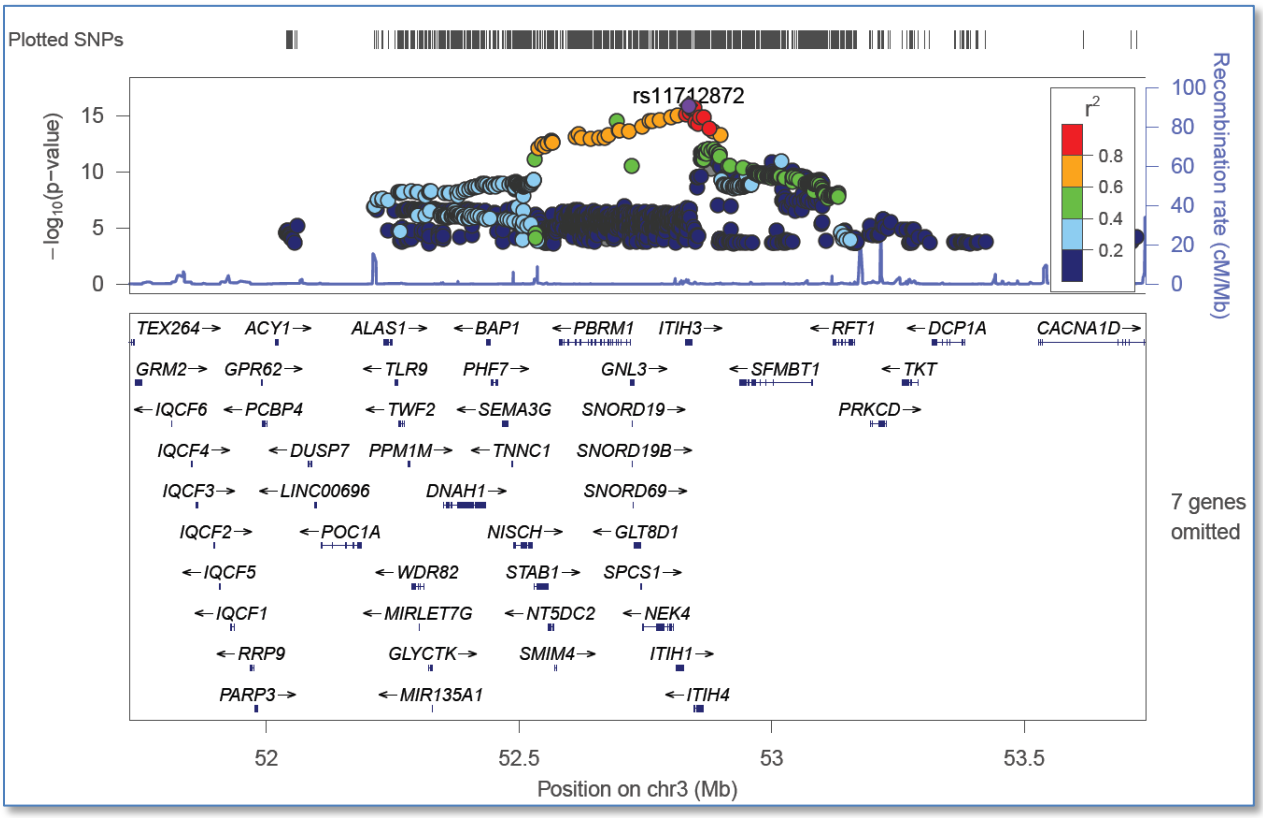
Shown are the distribution of distances and significance between genes and their most associated eQTL SNPs when genomic PC1-PC5 were included in eQTL analysis. Each point represents the most significant eQTL SNP per gene. X-axis shows the distance between the transcription start site of gene and the SNP location. Y-axis indicates $-\log_{10}$(q-value). We note that this is almost identical to Figure S11.

Figure S12. eQTLs and schizophrenia-associated *CNNM2* region



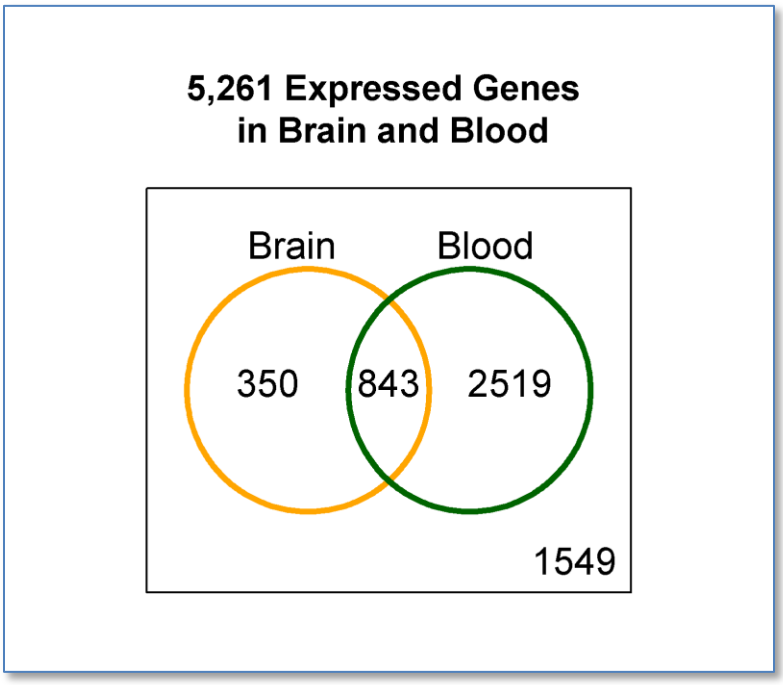
Upper panel shows -log10(p-value) of eQTLs.

Figure S13. eQTLs and schizophrenia-associated *ITIH3* region



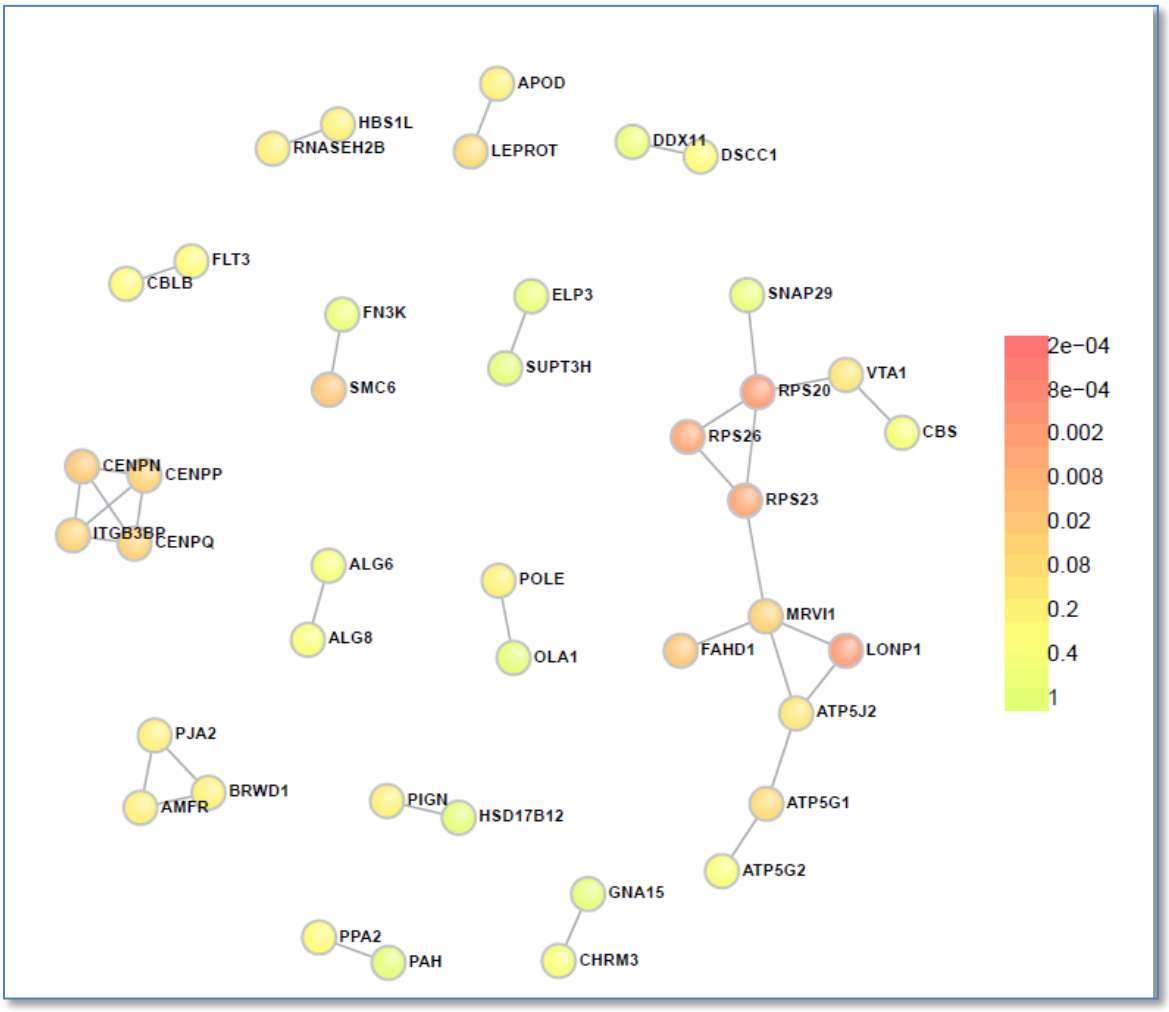
Upper panel shows -log10(p-value) of eQTLs.

Figure S14. Significant eQTL genes expressed in brain and blood



Of 5,261 genes expressed in brain and blood, 70.7% of 1,193 eQTL genes in brain were eQTL genes in blood.

Figure S15. DAPPLE protein-protein interaction graph



Using significant genes with evidence of local association, we performed protein-protein interaction analysis using DAPPLE. This figure shows network connectivity for directly interacting genes.

Supplemental Tables

Table S1. Sign test between all pairs of studies

From	To				
	Colantuoni	Gibbs	GTEx	Myers	SMRI
Colantuoni (n=56)	NA	186/255 (73%) P=6.9e-14	154/224 (69%) P=1.0e-8	183/264 (69%) P=1.5e-10	107/151 (71%) P=1.5e-7
Gibbs (n=125)	561/734 (76%) P=2.2e-16	NA	470/620 (76%) P=2.2e-16	569/714 (80%) P=2.2e-16	320/440 (73%) P=2.2e-16
GTEx (n=24)	61/109 (56%) P=0.13	61/97 (63%) P=0.007	NA	61/99 (62%) P=0.01	42/68 (62%) P=0.03
Myers (n=190)	346/474 (73%) P=2.2e-16	358/465 (77%) P=2.2e-16	258/362 (71%) P=2.2e-16	NA	227/329 (69%) P=2.3e-12
SMRI (n=31)	55/100 (55%) P=0.18	56/95 (59%) P=0.05	44/80 (55%) P=0.22	58/101 (57%) P=0.08	NA

SNP-gene eQTL pairs with $p < 1x10^{-5}$ were selected from the study in the "from" column, and the sign of the same association in the "To" column was assessed. For example, of 734 significant SNPs in Gibbs et al. that also passed quality control in Colantuouni et al., 561 (76%) had the same direction of association, a highly significant deviation from the null hypothesis of 50% ($P=2.2x10^{-16}$). In each cell, the ratio represents number of SNP-gene pairs whose t-statistic signs were the same between two studies. *P*-values were calculated using the R binom.test function. The asymmetry of the table is by design.

Table S2. eQTL SNP-gene pairs at varying FDR cutoffs (genomic PC1)

FDR Threshold	SNP-gene eQTL pairs	Unique SNPs	Unique genes
q < 0.1	227,621	202,926	5,374
q < 0.05	**176,794**	**159,151**	**3,520**
q < 0.01	115,839	105,447	1,797
q < 0.001	73,814	68,508	986
q < 0.0001	47,540	44,898	650

This table shows number of SNP-gene eQL pairs at varying FDR cutoffs when genomic PC1 was included as a covariate in eQTL analyses.

Table S3. eQTL SNP-gene pairs at varying FDR cutoffs (genomic PC1-5)

FDR Threshold	SNP-gene eQTL pairs	Unique SNPs	Unique genes
q < 0.1	233,978	209,864	6,179
q < 0.05	**176,566**	**159,613**	**4,015**
q < 0.01	112,258	102,984	1,925
q < 0.001	68,788	64,117	998
q < 0.0001	44,587	42,039	642

This table shows number of SNP-gene eQL pairs at varying FDR cutoffs when genomic PC1 – PC5 were included as covariates in eQTL analyses.

Table S4. Functional consequences of eQTL SNPs versus random SNPs

Functional effect	eQTL SNPs (%)	Random SNPs (%)	OR (CI)	P-value
5' UTR	0.98	0.29	9.2 (8.2-10.3)	$< 1x10^{-4}$
Non-synonymous coding	1.04	0.31	9.0 (8.1-10.1)	$< 1x10^{-4}$
3' UTR	2.66	0.85	8.4 (7.9-9.0)	$< 1x10^{-4}$
Synonymous coding	0.93	0.33	7.4 (6.7-8.2)	$< 1x10^{-4}$
Splice site related	0.30	0.13	6.2 (5.2-7.4)	$< 1x10^{-4}$
Upstream	7.11	4.45	4.3 (4.2-4.5)	$< 1x10^{-4}$
Downstream	5.84	3.77	4.2 (4.0-4.3)	$< 1x10^{-4}$
Intronic	65.91	50.12	3.5 (3.5-3.6)	$< 1x10^{-4}$
Within non-coding gene	0.61	0.47	3.5 (3.1-3.9)	$< 1x10^{-4}$
Intergenic	14.62	39.29	1.0	NA

OR represents odds ratio (95% confidence interval) of each functional effect versus intergenic.

Table S5. Top predicted genes for schizophrenia using SHERLOCK

Gene	SNP	chr	bp	eQTL *P*	GWAS *P*	LBF[a]	Associated diseases (PubMed ID)
ALMS1	rs11899902	2	73485072	4.66E-07	2.98E-05	5.11	Chronic kidney disease (20383146) Creatinine levels (20383145) Glomerular filtration rate (23535967) Metabolic traits (21886157) Metabolite levels (21931564)
PCGEM1	rs17439755	2	194238487	1.41E-04	5.24E-05	4.60	5 major psychiatric disorders (23453885) HIV-1 control (20041166) Schizophrenia (21926974)
NGEF	rs11535	2	233451776	1.02E-05	5.14E-06	5.95	SCZ (23974872)
NEK4	rs2590838	3	52597126	8.54E-05	3.53E-05	4.23	Bipolar disorder (19416921; 21926972)
GNL3	rs1108842	3	52695120	4.60E-05	3.82E-05	4.42	Adiponectin levels (22479202) Bipolar disorder (21926972; 23092984) Schizophrenia (23974872) Osteoarthritis (22763110)
GLT8D1	rs736408	3	52810394	1.52E-06	1.19E-06	5.68	5 major psychiatric disorders (23453885) Bipolar disorder (21926972) Osteoarthritis (22763110)
ITIH4	rs736408	3	52810394	4.28E-07	1.19E-06	5.41	Bipolar (21926972) Immune response to smallpox (22610502) Schizophrenia (21926974) Ulcerative colitis (23128233)
MEF2C	rs410216	5	88039857	1.97E-05	2.09E-05	4.22	Bone mineral density (21533022; 22504420; 23572186; 19801982) Height (20881960) Mean platelet volume (22139419) Obesity-related traits (23251661) Platelet counts (22139419) Retinal vascular caliber (21060863) Sex hormone-binding globulin levels (22675492) Thiazide-induced adverse metabolic effects (23400010) Tonometry (17903302)
GPR146	rs11762834	7	1862167	3.79E-05	1.25E-08	5.62	Schizophrenia (23974872)
TFAMP1	rs3800924	7	2154775	1.12E-13	3.56E-08	6.50	Schizophrenia (23974872)
FTSJ2	Rs6971396	7	2310729	4.58E-05	4.21E-05	4.71	5 major psychiatric disorders (23453885)
MED30	rs10105834	8	118684632	3.32E-07	5.68E-05	4.23	Kawasaki disease (22446961)
MASTL	rs4749219	10	27383238	7.29E-07	1.14E-05	4.55	
RHOBTB1	rs16914993	10	61842440	1.58E-05	1.31E-05	6.33	Diastolic blood pressure (19430483)
CALHM1	rs12219246	10	104603345	1.37E-04	9.40E-11	5.61	
C10orf32	rs7096169	10	104608685	3.76E-05	1.83E-11	4.58	5 major psychiatric disorders (23453885) Schizophrenia (23974872) Parkinson's disease (19915575)

AS3MT	rs7085104	10	104618863	3.18E-06	1.07E-11	6.10	5 major psychiatric disorders (23453885) Schizophrenia (23894747) Systolic blood pressure (19430483)
USMG5	rs10786736	10	104839106	7.08E-08	4.01E-08	5.64	Schizophrenia (23974872)
SNX19	rs3751039	11	130278286	5.94E-07	5.02E-06	4.21	Schizophrenia (22688191; 21926974)
VSIG10	rs3825102	12	116204568	1.47E-04	1.39E-5	4.80	Obesity-related traits (23251661)
MPHOSPH9	rs7299943	12	122159438	5.59E-05	1.48E-05	5.11	Multiple sclerosis (19525953)
OGFOD2	rs1969354	12	122307729	2.76E-05	5.99E-06	5.61	Platelet counts (21507922)
CORO7	rs6500596	16	4410028	6.90E-05	2.77E-06	4.94	
NMRAL1	rs1362626	16	4489227	1.96E-05	2.63E-05	5.09	
CMTM1	rs2290182	16	65171509	9.23E-08	0.0001398	4.28	
SRR	rs4523957	17	2155649	4.98E-07	3.01E-07	6.10	Aortic root size (19584346) Coronary heart disease (21378990) Metabolite levels (23319000) Type 2 diabetes (20174558)
NAGA	rs5751191	22	40700937	7.95E-08	9.02E-05	4.23	

The extended MHC region (chr6: 25Mb-334Mb) was removed due to high LD. Five major psychiatric disorders refers to ADHD, autism, bipolar disorder, major depressive disorder, and schizophrenia.

[a] : Log of Bayes factor.

Table S6. Top annotation clusters identified by DAVID

Category	Term	Count	Fold Enrichment	Benjamini
Annotation Cluster 1	Enrichment Score: 8.49			
GOTERM_CC_FAT	GO:0005739~mitochondrion	255	1.56	1.09E-11
GOTERM_CC_FAT	GO:0044429~mitochondrial part	135	1.51	1.39E-04
GOTERM_CC_FAT	GO:0005759~mitochondrial matrix	62	1.82	5.17E-04
GOTERM_CC_FAT	GO:0031980~mitochondrial lumen	62	1.82	5.17E-04
Annotation Cluster 2	Enrichment Score: 4.03			
GOTERM_BP_FAT	GO:0051186~cofactor metabolic process	56	1.88	0.010015
GOTERM_BP_FAT	GO:0006732~coenzyme metabolic process	47	2.01	0.005036
GOTERM_BP_FAT	GO:0051188~cofactor biosynthetic process	27	1.83	0.279835
GOTERM_BP_FAT	GO:0009108~coenzyme biosynthetic process	20	1.90	0.453828
Annotation Cluster 3	Enrichment Score: 3.77			
GOTERM_MF_FAT	GO:0050662~coenzyme binding	52	1.86	0.011634
GOTERM_MF_FAT	GO:0048037~cofactor binding	64	1.67	0.023209
GOTERM_MF_FAT	GO:0050660~FAD binding	23	2.07	0.142117
GOTERM_MF_FAT	GO:0009055~electron carrier activity	47	1.38	0.802524
Annotation Cluster 4	Enrichment Score: 3.40			
GOTERM_CC_FAT	GO:0000323~lytic vacuole	52	1.64	0.047607
GOTERM_CC_FAT	GO:0005764~lysosome	52	1.64	0.047607
GOTERM_CC_FAT	GO:0005773~vacuole	55	1.45	0.131342
Annotation Cluster 5	Enrichment Score: 3.21			
GOTERM_MF_FAT	GO:0016814~hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	15	3.14	0.041012
GOTERM_MF_FAT	GO:0019239~deaminase activity	13	3.37	0.048278
GOTERM_MF_FAT	GO:0004000~adenosine deaminase activity	5	4.06	0.773977
Annotation Cluster 6	Enrichment Score: 2.90			
GOTERM_CC_FAT	GO:0044429~mitochondrial part	135	1.51	1.39E-04
GOTERM_CC_FAT	GO:0031966~mitochondrial membrane	83	1.40	0.132058
GOTERM_CC_FAT	GO:0031967~organelle envelope	122	1.31	0.116647
GOTERM_CC_FAT	GO:0031090~organelle membrane	201	1.22	0.110334
GOTERM_CC_FAT	GO:0031975~envelope	122	1.31	0.097985
GOTERM_CC_FAT	GO:0005740~mitochondrial envelope	86	1.37	0.093481
GOTERM_CC_FAT	GO:0019866~organelle inner membrane	69	1.40	0.119952
GOTERM_CC_FAT	GO:0005743~mitochondrial inner membrane	63	1.37	0.205259
GOTERM_CC_FAT	GO:0044455~mitochondrial membrane part	30	1.60	0.23615

Table S7. Top 50 Gene Ontology pathways identified by InRich

GO ID	Type	Name	size	N.int	Empical P	Adjusted P
GO:0005739	CC	**mitochondrion**	1191	350	1.00E-05	0.039296
GO:0005759	CC	**mitochondrial matrix**	185	74	1.00E-05	0.039296
GO:0003735	MF	structural constituent of ribosome	151	62	1.00E-05	0.039296
GO:0005840	CC	ribosome	166	63	1.00E-05	0.039296
GO:0006412	BP	translation	225	84	1.00E-05	0.039296
GO:0010467	BP	gene expression	364	127	1.00E-05	0.039296
GO:0005743	CC	**mitochondrial inner membrane**	263	90	2E-05	0.051395
GO:0000278	BP	mitotic cell cycle	286	97	2E-05	0.051395
GO:0016491	MF	oxidoreductase activity	386	120	3E-05	0.066993
GO:0000166	MF	nucleotide binding	1889	483	3E-05	0.066993
GO:0044267	BP	cellular protein metabolic process	263	90	3E-05	0.066993
GO:0016740	MF	transferase activity	551	166	4E-05	0.081092
GO:0003676	MF	nucleic acid binding	681	178	5E-05	0.09589
GO:0003723	MF	RNA binding	553	160	5E-05	0.09589
GO:0022904	BP	respiratory electron transport chain	81	35	7E-05	0.121788
GO:0005654	CC	nucleoplasm	822	219	0.00011	0.175782
GO:0055114	BP	oxidation-reduction process	194	66	0.00013	0.20058
GO:0005975	BP	carbohydrate metabolic process	272	87	0.00014	0.212179
GO:0004497	MF	monooxygenase activity	58	23	0.00015	0.225977
GO:0006418	BP	tRNA aminoacylation for protein translation	40	21	0.00016	0.237076
GO:0015030	CC	Cajal body	42	22	0.00018	0.259674
GO:0006414	BP	translational elongation	88	33	0.0002	0.284872
GO:0005761	CC	mitochondrial ribosome	22	14	0.00025	0.340166
GO:0005861	CC	troponin complex	8	6	0.00027	0.362264
GO:0000084	BP	S phase of mitotic cell cycle	108	41	0.0003	0.393461
GO:0000087	BP	M phase of mitotic cell cycle	89	37	0.00032	0.411759
GO:0016071	BP	mRNA metabolic process	208	65	0.00032	0.411759
GO:0005764	CC	lysosome	160	58	0.0005	0.561244
GO:0005681	CC	spliceosomal complex	73	30	0.00051	0.567443
GO:0016070	BP	RNA metabolic process	238	74	0.00052	0.572143
GO:0005783	CC	endoplasmic reticulum	871	234	0.00053	0.578842
GO:0003674	MF	molecular_function	523	147	0.0006	0.627037
GO:0005829	CC	cytosol	1940	459	0.0006	0.627037
GO:0006415	BP	translational termination	80	30	0.00061	0.633337
GO:0016254	BP	preassembly of GPI anchor in ER membrane	15	10	0.00066	0.661234
GO:0005789	CC	endoplasmic reticulum membrane	551	153	0.00067	0.665533
GO:0019058	BP	viral infectious cycle	84	31	0.00067	0.665533
GO:0031080	CC	Nup107-160 complex	10	8	0.00087	0.758224
GO:0005524	MF	ATP binding	1362	361	0.00124	0.871713
GO:0006783	BP	heme biosynthetic process	17	11	0.0013	0.884812
GO:0000236	BP	mitotic prometaphase	79	32	0.00136	0.894011
GO:0006457	BP	protein folding	170	56	0.00139	0.89821

GO:0051301	BP	cell division	258	80	0.00141	0.90091
GO:0006778	BP	porphyrin-containing compound metabolic process	15	9	0.00156	0.921208
GO:0008137	MF	NADH dehydrogenase (ubiquinone) activity	36	18	0.00158	0.924308
GO:0000389	BP	nuclear mRNA 3'-splice site recognition	5	5	0.00161	0.928207
GO:0006165	BP	nucleoside diphosphate phosphorylation	5	4	0.002	0.960604
GO:0019885	BP	antigen processing and presentation of endogenous peptide antigen via MHC class I	6	5	0.00204	0.962604
GO:0006098	BP	pentose-phosphate shunt	10	7	0.00205	0.963604
GO:0006120	BP	mitochondrial electron transport, NADH to ubiquinone	35	17	0.00209	0.965203

To adjust for common biases due to gene size, LD within and between genes, and pathway sizes, we generated a set of independent, nominally associated genomic intervals using clumping implemented in PLINK and then tested it for enrichment in GO gene sets using InRich. Analysis was restricted to 4,074 GO categories containing genes between 5 and 3000 to account for pathway sizes. We used 100K times of permutation to get empirical p-values per pathway and 10K times of permutation to obtain multiple-testing corrected p-values. Multiple GO pathways related to mitochondrial structure and function were ranked as top pathways. This result is consistent with gene-based DAVID results, indicating that mitochondrial pathways are robust findings regardless of different gene-set enrichment methods. Size: number of gene in each GO category, N.int: number of intervals overlapped with genes in each GO category.

References

1. Wright, F.A. et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet* **46**, 430-7 (2014).
2. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
3. Altshuler, D.M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
4. Durbin, R.M. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
5. Ramasamy, A. et al. Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Res* **41**, e88 (2013).
6. Wright, F. et al. Heritability and genomics of gene expression in peripheral blood. (Submitted).
7. Colantuoni, C. et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519-23 (2011).
8. Gibbs, J.R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genetics* **6**, e1000952 (2010).
9. Lonsdale, J. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
10. Myers, A.J. et al. A survey of genetic human cortical gene expression. *Nat Genet* **39**, 1494-9 (2007).
11. Torrey, E.F. et al. Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains. *Biol Psychiatry* **57**, 252-60 (2005).
12. Torrey, E.F., Webster, M., Knable, M., Johnston, N. & Yolken, R.H. The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* **44**, 151-5 (2000).
13. Liu, E.Y., Li, M., Wang, W. & Li, Y. MaCH-admix: genotype imputation for admixed populations. *Genet Epidemiol* **37**, 25-37 (2013).
14. Liu, E.Y. et al. Genotype imputation of Metabochip SNPs using a study-specific reference panel of ~4,000 haplotypes in African Americans from the Women's Health Initiative. *Genet Epidemiol* **36**, 107-17 (2012).