Brain eQTLs of European, African American, and Asian ancestry improve interpretation of schizophrenia GWAS

Yu Chen[1,2,16], Sihan Liu[1,3,16], Zongyao Ren[1], Feiran Wang[1], Yi Jiang[1], Rujia Dai[4], Fangyuan Duan[1], Cong Han[1], Zhilin Ning[5], Yan Xia[6], Miao Li[1], Kai Yuan[2], Wenying Qiu [7], Xiao-Xin Yan [8], Jiapei Dai[9], Richard F. Kopp[4], Jufang Huang[8], Shuhua Xu[10], Beisha Tang[14], Eric R. Gamazon[11], Tim Bigdeli[12], Elliot Gershon[13], Hailiang Huang[2], Chao Ma[7*], Chunyu Liu[1,4*], and Chao Chen [1,14,15*]

[1] MOE Key Laboratory of Rare Pediatric Diseases & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, and Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410000, China.

[2] Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[3] Institute of Rare Diseases, West China Hospital, Sichuan University, Chengdu, China.

[4] Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA.

[5] Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China.

[6] Hunan Key Laboratory of Molecular Precision Medicine, Department of Neurosurgery, Xiangya Hospital, and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China.

[7] Institute of Basic Medical Sciences, Neuroscience Center, National Human Brain Bank for Development and Function, Chinese Academy of Medical Sciences; Department of Human

21 Anatomy, Histology and Embryology, School of Basic Medicine, Peking Union Medical College,
22 Beijing, China.

23 [8] Department of Human Anatomy and Neurobiology, Xiangya School of Medicine, Central South
24 University, Changsha, China.

25 [9] Wuhan Institute for Neuroscience and Engineering, South-Central University for Nationalities,
26 Wuhan, China.

27 [10] State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, Collaborative
28 Innovation Center of Genetics and Development, School of Life Sciences, Fudan University,
29 Shanghai, China.

30 [11] Division of Genetic Medicine Vanderbilt University School of Medicine Nashville, USA

31 [12] Institute for Genomics in Health, SUNY Downstate Health Sciences University, Brooklyn, New
32 York, USA

33 [13] Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL,
34 USA

35 [14] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South
36 University, Changsha, China.

37 [15] Hunan Key Laboratory of Animal Models for Human Diseases, Central South University,
38 Changsha, China.

39 [16] These authors contributed equally: Yu Chen, Sihan Liu.

40 [*]Corresponding author. Email: machao@ibms.cams.cn (C.M.); LiuCh@upstate.edu (C.L.);
41 chenchao@sklmg.edu.cn (C.C.)

Abstract

Research on brain expression quantitative trait loci (eQTLs) has illuminated the genetic underpinnings of schizophrenia (SCZ). Yet, the majority of these studies have been centered on European populations, leading to a constrained understanding of population diversities and disease risks. To address this gap, we examined genotype and RNA-seq data from African Americans (AA, n=158), Europeans (EUR, n=408), and East Asians (EAS, n=217). When comparing eQTLs between EUR and non-EUR populations, we observed concordant patterns of genetic regulatory effect, particularly in terms of the effect sizes of the eQTLs. However, 343,737 cis-eQTLs (representing ~17% of all eQTLs pairs) linked to 1,276 genes (about 10% of all eGenes) and 198,769 SNPs (approximately 16% of all eSNPs) were identified only in the non-EUR populations. Over 90% of observed population differences in eQTLs could be traced back to differences in allele frequency. Furthermore, 35% of these eQTLs were notably rare (MAF < 0.05) in the EUR population. Integrating brain eQTLs with SCZ signals from diverse populations, we observed a higher disease heritability enrichment of brain eQTLs in matched populations compared to mismatched ones. Prioritization analysis identified seven new risk genes (*SFXN2*, *RP11-282018.3*, *CYP17A1*, *VPS37B*, *DENR*, *FTCDNL1*, and *NT5DC2*), and three potential novel regulatory variants in known risk genes (*CNNM2*, *C12orf65*, and *MPHOSPH9*) that were missed in the EUR dataset. Our findings underscore that increasing genetic ancestral diversity is more efficient for power improvement than merely increasing the sample size within single-ancestry eQTLs datasets. Such a strategy will not only improve our understanding of the biological underpinnings of population structures but also pave the way for the identification of novel risk genes in SCZ.

Keywords: eQTLs; Schizophrenia; GWAS signals; Population diversity

65 **Introduction**

66 Genome-wide association studies (GWAS) have identified 287 risk loci associated with
67 schizophrenia (SCZ)[1]. Yet, the underlying mechanisms of these loci in disease development and
68 progression remain poorly understood. Primarily, over 80% of GWAS risk loci reside in non-
69 coding regions, devoid of protein-coding sequences, making it challenging to attribute them to
70 specific genes. Moreover, predicting the regulatory effect of these loci proves challenging due to
71 their tendency for gene-specific and tissue-specific effects. One effective strategy for gaining
72 insights into their functions involves the integration of SCZ GWAS signals with expression
73 quantitative trait loci (eQTLs), utilizing genotype and expression data from post-mortem brains.
74 These brain eQTLs establish crucial links between risk genomic regions and gene expression levels,
75 prioritizing potential disease risk genes through methods such as colocalization and transcriptome-
76 wide association study (TWAS).

77 Past brain eQTL studies primarily focused on European (EUR) ancestry[2–6]. Global population
78 diversity has not been adequately represented. Cross-population studies have shown that these
79 European ancestry based models do not effectively predict gene expression in other ancestral
80 groups[7]. This limitation weakens the power to detect TWAS associations in genetically diverse
81 samples. While multi-ancestry eQTL meta-analyses in the human brain improve statistical power
82 in uncovering risk loci shared across populations, key genetic variants regulating expression in
83 specific underrepresented populations remain largely uncharted. The benefits of having brain
84 eQTLs in diverse populations have not been thoroughly documented. Identifying eQTLs specific
85 to biomedically underrepresented groups like African Americans (AA) and East Asians (EAS) can
86 better understand the genetic contributions to disease susceptibilities and outcomes in these
87 populations[7]. These populations have unique genetic variants and linkage disequilibrium (LD)

88 patterns. Additionally, previous studies have shown that combining eQTLs from different
89 ancestries can enable fine-mapping of causal variants and uncover potential novel mechanisms of
90 brain disorders[8,9]. Thus, the question of how to effectively leverage difference to uncover potential
91 novel mechanisms of brain disorders is a significant topic in the field.

92 To enhance the diversity in brain eQTL mapping and improve the interpretation of SCZ GWAS
93 across populations, we performed brain eQTL mapping in three major ancestries. Our data pool
94 comprised genotype and RNA-seq data of AA (n= 158) and EUR (n= 408) from the
95 PsychENCODE Consortium and EAS (n= 217) from the Chinese Human Brain Bank
96 (Supplementary Table 1). We juxtaposed non-EUR results against EUR to systematically examine
97 differences and similarities in the brain eQTLs. Further, we investigated the contributing factors
98 for eQTL differences across populations. By applying diverse population brain eQTLs to TWAS
99 and colocalization analysis of SCZ GWAS, we identified new risk genes and pathways. Lastly, we
100 identified likely causal variants by multi-ancestry fine-mapping. The two key questions we sought
101 to answer are: 1) what drives the brain eQTL differences across populations? 2) what do we gain
102 by studying brain eQTLs in diverse populations?

103 **Results**

104 To capture brain eQTLs across diverse populations, we utilized high-density genotype data
105 alongside high-throughput RNA-sequencing from prefrontal cortices. We derived AA (n = 158)
106 and EUR (n = 408) data from the BrainGVEx project of the PsychENCODE Consortium. We
107 generated EAS (n = 217) data from the Chinese Human Brain Bank (Fig. 1). Following rigorous
108 quality checks and preprocessing (Extended Data Fig. 1 and Extended Data Fig. 2), we compiled
109 expression data for 18,939 genes and genotype data at 6.4 million autosomal single nucleotide

110 polymorphisms (SNPs) across the three groups. Aligning the samples with the 1000 Genomes
111 Project reference populations, principal component analysis (PCA) confirmed the ancestry origins
112 of donors (Fig. 2a). We ensured sample identity consistency by comparing the genotypes from the
113 DNA and RNA samples. See methods for additional details.

114 **Characterizing the cis-acting eQTLs in European, East Asian, and African American**
115 **population**

116 We separately conducted cis-eQTLs mapping in the EUR, EAS, and AA samples using a 5%
117 empirical gene-level false discovery rate (FDR) threshold. This yielded 1,966,209 significant
118 eQTL signals covering 11,622 genes (eGenes) and 1,226,769 SNPs (eSNPs) across the populations
119 (see Fig. 2b and Supplementary Table 2). Specifically, we identified 1,616,818 eQTL signals
120 spanning 10,236 eGenes and 1,025,004 eSNPs for EUR; 562,058 eQTLs incorporating 5,000
121 eGenes and 416,025 eSNPs for EAS; and 143,736 eQTLs covering 3,039 eGenes and 121,079
122 eSNPs for AA. To identify credible SNP sets harboring plausible causal variants in cis-eQTLs, we
123 applied a fine-mapping method named SuSiE[10] to each population's eQTL results. The results
124 showed 966 credible SNP sets for 757 eGenes in the EUR cohort, 826 sets for 726 eGenes in EAS,
125 and 847 sets for 746 eGenes in AA (Supplementary Tables 3-5).
126 To investigate the genomic features of these cis-eQTLs, we evaluated the SNP distributions and
127 locations relative to various functional regions. 20% of cis-eQTLs in both EUR and non-EUR
128 populations were located within 10kb of transcription start site (TSS) regions (Fig. 2c). According
129 to the chromatin states predicted by Genomic Regulatory Elements and Gwas Overlap algoRithm
130 (GREGOR)[11] for prefrontal cortical tissue, eSNPs from the non-EUR populations were
131 significantly enriched in TSSs, promoters, and transcribed regulatory promoters or enhancers

132 ($P_{Bonferroni} < 0.05$), identical to the observation in the EUR data. Moreover, using transcription
133 factor binding site (TFBS) annotation for 51 TFs, 46 and 49 TFs were significantly enriched with
134 cis-eQTLs in the AA and EAS populations respectively ($P_{Bonferroni} < 0.05$). All these TFBS were
135 also significantly enriched with cis-eQTLs in the EUR population.

136 To maximize the power of our population-based datasets, we employed METAL to amalgamate
137 the cis-eQTLs data from all three populations. This meta-analysis generated 598,193 eQTLs,
138 correlating 436,456 eSNPs with 5,209 eGenes (with a stringent P<2.5e-6) as depicted in Fig.
139 2b. However, meta-analysis does not consider the LD differences among the populations. We thus
140 used SuSiEx[16] to identify likely causal variants regulating expression by incorporating the LD
141 reference data from different populations. In total, SuSiEx identified 2,121 credible SNP sets for
142 1,801 eGenes in the 3-population combined data. Further details from the meta-analysis and fine-
143 mapping results can be found in Supplementary Table 6.

144 **Population shared eQTLs showed similar regulatory effect across population**

145 To evaluate the effect sizes across populations, we conducted a correlation test of effect size values
146 between the EUR and non-EUR populations. The effect sizes of the shared eQTLs between the
147 non-EUR and EUR were highly concordant (0.910 for eQTLs comparing EAS to EUR and 0.944
148 for AA to EUR; Fig 3a). To assess the robustness of the concordant effect size, we examined eQTL
149 slopes in the different populations. We looked at eQTLs obtained from the nominal, permutation,
150 and conditional tests, separately. eQTLs with smaller p-values or larger effect sizes showed greater
151 consistency across populations (Extended Data Fig. 3). Considering that sample size and
152 heterogeneity may influence the results, we randomly down-sampled the EUR data to match the
153 size of the non-EUR data. The results were similar to the results comparing all samples, showing

154 highly concordant effect sizes across populations (see Fig. 3c, Extended Data Fig. 3). In addition,
155 we compared the slope of the down-sampled EUR data with GTEx data (also of EUR). We
156 randomly sub-sampled 100 times and obtained a distribution of correlation values (R^2). The mean
157 R^2 was 0.94, which was not significantly different from the correlation between the EUR and non-
158 EUR population. We therefore concluded that the effect sizes of eQTLs in diverse populations
159 were mostly stable across human populations.

160 We evaluated the replicated rate (π_1), which gauges the true positive rate for the eQTLs identified
161 in the non-EUR populations that were also associated in the EUR population. The replicated rate
162 was π_1(EAS-EUR) = 0.86 and π_1(AA-EUR) = 0.91 (see Fig.3b). The π_1 for the non-EUR
163 populations in EUR was slightly but significantly lower than the π_1 between two EUR cohorts, as
164 represented by the GTEx cis-eQTL data (prefrontal cortex) in our EUR eQTL data (π_1(GTEx-
165 EUR) = 0.86, p-value = 0.023). To ensure a fair comparison of the replication rate of detected cis-
166 eQTLs in non- EUR data, we adjusted the EUR data to reflect the smaller sample size of the non-
167 EUR data. This adjustment enabled us to determine how many non-EUR cis-eQTLs were
168 confirmed in the adjusted EUR dataset. The adjusted results revealed a concordant trend: the
169 replicated rate between different populations was still slightly lower than that within the same
170 population assuming the same sample size (EUR-nonEUR average π_1 = 0.68, $EUR_{adjusted}$-EUR
171 average π_1 = 0.72, p-value = 0.037) (see Fig. 3d).

172 **Population differences in brain cis-eQTLs are mainly caused by differences in SNP allele**
173 **frequency while differences in effect size are small and uncertain**

174 Here we defined those eQTLs that were exclusively observed in a single population as population-
175 specific eQTLs. Upon analyzing the cis-eQTLs overlapping between populations, we identified
176 343,737 cis-eQTLs that were exclusively observed in the non-EUR populations, as detailed in
177 Supplementary Table 2. This number represents approximately 17% of all eQTL pairs. These
178 eQTLs involved 1,276 genes (about 10% of all eGenes) and 198,769 SNPs (around 16% of all
179 eSNPs, Fig. 2d-f.). Specifically, there were 292,254 cis-eQTLs involving 165,300 eSNPs and 937
180 eGenes that were observed only in the EAS population and 51,483 cis-eQTLs involving 33,469
181 eSNPs and 339 eGenes that were observed only in the AA population.

182 To further characterize these non-EUR-specific eQTLs, we analyzed the variance, taking into
183 account both the eQTL slope (effect size) and differences in allele frequency between populations.
184 We found that more than 90% of the population differences in variance were attributable to
185 differences in allele frequency. Moreover, to delve deeper into the distinctive characteristics of the
186 eQTLs exclusive to the non-EUR groups, we leveraged two statistics, the fixation index (FST) and
187 the minor allele frequency (MAF), retrieved from the 1000 Genomes Selection Browser[12]. A high
188 FST value indicates that the measured locus has diverged over time in the populations. As expected,
189 eSNPs detected only in the EAS or AA population displayed a significantly elevated FST when
190 juxtaposed against eSNPs shared across populations (Wilcoxon test $P < 2.2e-16$). Meanwhile, the
191 non-EUR-specific eSNPs showed higher MAF values in their respective source populations
192 (Wilcoxon test $P < 2.2e-16$;) than in EUR. Of the 343,737 eQTLs absent in the EUR data, 309,363

193 were likely due to inadequate statistical power because they have smaller MAF in the EUR than
194 non-EUR population. Furthermore, 70,980 were rare in the EUR population ($MAF_{1KG} < 0.05$).

195 For the rest the eQTLs for which population differences could not be explained by differences in
196 MAF, a test for differences in eQTL slopes (effect sizes) was also conducted between the EUR
197 and non-EUR populations. The z-score of each independent eQTLs from conditional analysis was
198 calculated based on effect size and its standard deviation. Here the null hypothesis was that the
199 difference in eQTL effect size between the populations equals zero. No eQTL pairs detected by
200 conditional analysis could reject the null hypothesis. We then investigated if any eQTLs exhibited
201 opposite effect directions across populations. None of the independent eQTLs from the conditional
202 analysis displayed such effects. We relaxed our eQTL threshold using a nominal p-value < 0.05.
203 Only 534 eQTLs involving eighteen genes exhibited opposing eQTL effects between the EUR and
204 non-EUR populations. For example, the bitter taste receptor gene *TAS2R31* showed opposite
205 directions of eQTLs in EAS and EUR (Fig. 3e), which could be replicated using the blood eQTLs
206 from a previous study[7,13].

207 In conclusion, the variance in population differences can be largely attributed to differences in
208 allele frequency. The influence of effect size differences, on the other hand, appears to be minimal
209 and inconclusive.

210 **Brain eQTLs from matched population can improve interpretation of SCZ GWAS**

211 To determine if eQTLs detected from a specific population could explain the disease GWAS
212 signals and SNP-based disease heritability better than eQTLs from non-matching populations, we
213 undertook a two-step analysis. Firstly, we gathered SCZ GWAS summary statistics for the EUR,
214 EAS, and AA populations from previously published studies[14–16]. We employed the partitioned

215 LD-score regression (LDSR)[17] approach to assess the GWAS signal enrichment of these eQTLs.
216 eQTLs identified in the EAS population demonstrated a markedly higher enrichment in EAS-based
217 GWAS signals than eQTLs identified in the EUR population. Conversely, eQTLs identified in the
218 EUR population showed a greater enrichment for EUR-based GWAS signals than the eQTLs from
219 the EAS cohort. Both of these enrichments were statistically significant (Welch Modified Two-
220 Sample t-Test P-value < 0.001, Fig. 4a and 4b).

221 Besides the SNP heritability enrichment of all eQTLs, we also compared the significance of the
222 GWAS signals for population-unique eSNPs. We found that population-unique eSNPs tended to
223 have smaller p-values of disease association (i.e., stronger associations) in the corresponding
224 population than in the mismatched population (Welch Modified Two-Sample t-Test P < 0.001),
225 indicating the ability of population-unique eSNPs to explain the disease association and propose
226 the relevant gene, which will be missed when only one population is studied.

227 **SCZ risk genes identified using eQTLs and GWAS from non-EUR population**

228 To uncover risk genes and pathways for SCZ in non-EUR populations, we used TWAS, regulatory
229 Trait Concordance (RTC)[18], and summary data-based mendelian randomization (SMR)[19] to
230 prioritize SCZ candidate risk genes in non-EUR populations and compared them with risk genes
231 identified in the EUR (see Methods). In total, we prioritized ten risk genes in the EAS
232 (Supplementary Tables 7-9). It is worth noting that our TWAS analysis of AA data did not reveal
233 any significant associations. This lack of association in AA data might be attributed to the
234 relatively small sample size available from the AA SCZ GWAS.

235 Seven SCZ candidate risk genes (*SFXN2*, *RP11-282018.3*, *CYP17A1*, *VPS37B*, *DENR*, *FTCDNL1*,
236 and *NT5DC2*) uniquely discovered in the EAS population were assessed for allele frequency. From

237 these seven genes, two of them were located in chromosome 10 and five of them were located in
238 the chromosome 12. The eSNPs for these genes showed lower allele frequency in the EUR
239 population than in EAS. For instance, the GWAS signal chr12:123286491:A:G in the gene
240 *VPS37B* was found to be significant in the EAS population with a high AF of 0.48. In contrast,
241 this association was not significant in the EUR population with a markedly lower allele frequency
242 of 0.04. A parallel pattern emerged with the eSNP for *VPS37B*, with a markedly higher frequency
243 (chr12:123306558:G:A, MAF=0.24) in EAS than in the EUR (MAF = 0.04). These results further
244 confirm that allele frequency differences between populations can explain most of the
245 discrepancies between the EUR and EAS GWAS and the eQTL results (Supplementary Table 10).

246 **Novel potential SCZ regulatory variations were refined utilizing brain eQTLs from non-**
247 **EUR population**

248 Three of the ten risk genes identified in the EAS population were shared with the EUR population
249 (*CNNM2*, *C12orf65*, and *MPHOSPH9*), but differences in genetic architecture between
250 populations were still apparent. For example, two distinct significant SNPs in EAS and in EUR
251 were associated with SCZ on chromosome 10 ($GWAS_{EUR}$: chr10:104850632:G:A with GWAS p-
252 value = 6.4E-13; $GWAS_{EAS}$: chr10:104657300:T:C with GWAS p-value = 5.4E-12). Using
253 eQTLs with GWAS signals in EAS and EUR separately, colocalization and SMR analysis
254 prioritized these two distinct GWAS SNPs to the same risk gene *CNNM2* in the two populations,
255 respectively.

256 To further investigate if these signals are located within any regulatory elements, we utilized the
257 non-coding variant annotation database (NCAD)[20] to annotate their regulatory information. Our
258 findings revealed that all the EAS GWAS risk SNPs are situated in enhancer regions

259 (Supplementary Table 11). The EAS GWAS risk SNPs near the *CNNM2* showed the strongest
260 evidence. Furthermore, we used the Lineage-specific Brain Open Chromatin Atlas[21] to investigate
261 whether this enhancer region shows different effects in major brain cell types. The results did not
262 show any cell type differences, which indicates the enhancer effect exists universally in major
263 brain cell types (Extended Data Fig. 4a-c). Integrating these insights, we discovered strong
264 evidence for multiple regulatory regions among the EAS eSNPs-chr10:104654577:T:C, which
265 have a high LD with *CNNM2* GWAS SNPs (LD $R^2 = 1$, p-value < 0.00001). Additionally, our
266 dual luciferase reporter assay results confirmed that the EAS eSNPs C-allele at
267 chr10:104654577:T:C significantly enhances luciferase activity compared to the reference vector,
268 as detailed in the Extended Data Fig. 4d-e.

269 **High-confidence putative causal variants of SCZ using multi-ancestry brain eQTLs**

270 To identify high-confidence putative causal variants from multiple populations, we applied
271 colocalization to our fine-mapped eQTLs and SCZ GWAS signals. In total, we identified four
272 SNP-gene-disease triplets in which the SNP colocalized with both gene expression and SCZ
273 GWAS (Supplementary Table 12, PiPcoloc = $PiP_{GWAS} \times PiP_{cis\text{-}eQTLs} > 0.1$). The top genes with
274 $PIP_{coloc} > 0.1$ include *FURIN*, *ZNF823*, *RHOA*, and *VWA5A* (Fig. 4c). As an example, we identified
275 the strongest putative SCZ causal SNP for *FURIN*-chr15:91426560:G:A. This SNP is located in
276 the 3' untranslated regions (UTRs) of *FURIN*. Notably, this variant did not reach genome-wide
277 significance in the EAS population (P = 1.06E-3) likely due to limited statistical power. Our result
278 strongly supported that this causal variant is shared across populations, with causal probabilities
279 of 1. Previous study has also implicated the variant in both the EUR and EAS populations[13].

280 **Discussion**

281 In this study, we have created a brain transcriptome resource and identified eQTLs in the prefrontal
282 cortex, specifically focusing on non-European populations. Our findings address the initial
283 inquiries raised in the introduction. Firstly, we investigated the driver behind the variation in brain
284 eQTLs across different populations. We found that differences in allele frequency and LD are
285 instrumental in connecting disease susceptibility to gene expression regulation. This finding
286 greatly augments our comprehension of genetic influences on gene expression in the human brain.
287 Secondly, when examining brain eQTLs from diverse populations, we gained power to explain the
288 GWAS heritability, uncover novel risk genes, and fine-map risk variants. We observed a
289 pronounced enrichment of disease heritability among eQTLs in matched populations. In the non-
290 EUR cohort, the allele frequencies and LD configurations facilitated the identification of seven
291 novel SCZ risk genes. Additionally, we identified four high-confidence putative causal SCZ
292 variants. These results highlight the utility of studying non-European cohorts.

293 Population differences appear to be more pronounced at the allele frequency level but are less so
294 at the effect size level. In general, the estimated π_1 of eQTLs from non-EUR populations in EUR
295 is lower compared to the rate observed between down-sampling-EUR and the EUR population
296 cohort. Despite the relatively small sample size and statistical power, we still detected a significant
297 number of 343,737 cis-eQTLs including 232,254 EAS eQTLs and 51,483 AA eQTLs that were
298 significant only in the non-EUR populations. While over half of eSNPs in our non-EUR dataset
299 were population-unique, 80% of eGenes identified in the non-EUR were also eGenes in the EUR
300 data, but associated with different SNPs. The consistency of our observations with prior research
301 involving diverse populations, including studies on gene expression[22,23], methylation[24], and
302 chromatin accessibility[25], confirms the shared regulatory patterns across different populations.

303 Interestingly, some eQTLs showed contrasting effects across populations. ~0.1% of the non-EUR-
304 specific eQTLs displayed opposing directions in effect size. A notable example of this is the eQTL
305 rs2599400-*TAS2R31*, which showed opposite effects in different populations. Blood eQTLs from
306 EAS[26], EUR, and AFR[7] also support this observation. Prior studies have underscored the
307 population-specific variations in *TAS2R31*, linking these variations to differing sensitivity to the
308 bitter taste[27]. It is important to further investigate these discrepancies in gene expression regulation
309 among populations as they influence phenotypes. However, it should be noted that lack of
310 replication across all these studies limited our power to detect opposite-effect eQTLs.
311 Our findings underscore that enhancing genetic ancestral diversity is more efficient for power gain
312 than increasing the sample size within large-scale eQTLs datasets. Through our benchmarking of
313 eQTLs across three populations, we have established robust capabilities for identifying eQTLs
314 with a MAF greater than 0.2 and an effect size of 0.6 (Extended Data Fig. 5). Our power analysis
315 indicates that more than 30,000 individuals of European ancestry is needed to uncover all eQTLs
316 with MAF of 0.01 in this population, based on the estimated effect size of eQTLs exclusively
317 observed in non-EUR populations (Extended Data Fig. 5). For example, one the eQTL pair
318 (chr12:123306558:G:A-VPS37B) would require 3,215 EUR samples based on the power estimate
319 because of the low frequency in the EUR population (MAF = 0.04). However, the MAF of this
320 eSNP is 0.22 in EAS, which reduces the required sample size from 3,215 to 246. Thus,
321 incorporating a more diverse population would not only reveal numerous regulatory variants that
322 are rarer in EUR but more prevalent in non-EUR groups. Advancing towards a broader, more
323 diverse human reference dataset will facilitate more comprehensive investigations into the impact
324 of human demography on eQTL detection, thereby deepening our understanding of the distribution
325 and influence of genetic regulation in the human brain.

326 Differences in the genetic architecture underlying gene expression can help us to prioritize new
327 risk genes. Notably, prior research has reported that disease-associated loci tend to be skewed
328 towards variants with higher allele frequency in the discovery population, indicating that limited
329 statistical power may result in "missing" disease-association signals. Incorporating diverse
330 samples can enhance our ability to uncover the etiology of the disease. In our study, we identified
331 seven new SCZ risk genes using the non-EUR population, including *VPS37B* in the EAS
332 population. *VPS37B* is associated with calcium-dependent protein binding, providing new
333 evidence to support the involvement of the calcium-related pathway in SCZ risk in the EAS
334 population[28]. Another interesting candidate highlighted in our study was *CYP17A1* (RTC = 0.99).
335 The corresponding GWAS signal was significant in the EAS and EUR populations (P_{EAS} = 4.5E-
336 8; MAF_{EAS} = 0.48; P_{EUR} = 2.6E-13; MAF_{EUR} =0.30), while the corresponding eSNP in EAS
337 population (MAF=0.48) showed extremely low frequency in EUR population (MAF<0.001).
338 *CYP17A1* notably serves as an enzyme important for the production of glucocorticoids and sex
339 hormones, such as estrogen, which have been linked to schizophrenia[29–31] .

340 Besides enhancing the power for detecting risk genes, the inclusion of brain eQTLs from diverse
341 populations improves the ability to fine-map SCZ GWAS loci, identifying new regulatory variants
342 which have the potential to regulate downstream gene expression. This approach aids in
343 interpretation, thereby facilitating subsequent computational and experimental functional
344 investigations. Our result revealed a potential novel regulatory region near the population-shared
345 risk gene *CNNM2*. This discovery showcases the power of leveraging diverse populations.

346 By leveraging the multi-ancestry information, trans-ancestry fine-mapping also helped us identify
347 high-confidence putative causal variants. In addition to previously validated genes, our study
348 uncovered another significant finding at chr3:50297330:A:G-*RHOA*-SCZ through trans-ancestry

349 colocalization. *RHOA* encodes a member of the Rho family of small GTPases, pivotal in signal
350 transduction cascades by toggling between inactive GDP-bound and active GTP-bound states.

351 Some limitations of our study merit attention. Our sample size is still relatively small. Our analysis
352 suggests that expanding the sample size would capture a larger set of eQTLs. The small sample
353 size could have significantly impacted the comprehensiveness of our findings. The modest sample
354 size of the AA eQTL dataset and the SCZ GWAS cohort likely led to the failure of TWAS in the
355 AA population.

356 In conclusion, we presented a novel genome-wide map of human brain gene expression regulation.
357 Importantly, this resource bridges the gap between neuropsychiatric GWAS and brain gene
358 expression profiling in non-European populations. Our study emphasizes the significance of this
359 new atlas of brain gene expression regulation in non-European populations for advancing our
360 understanding of human diversity, addressing health disparities, and developing precision
361 medicine.

362 Materials and methods

363 Sample collection and sequencing

364 We collected 217 prefrontal cortical samples of Han Chinese ancestry from the National Human
365 Brain Bank for Development and Function[32,33]; the samples were handled according to the
366 standardized operational protocol of the China Human Brain Banking Consortium, under the
367 approval of the Institutional Review Board of the Institute of Basic Medical Sciences, Chinese
368 Academy of Medical Sciences, Beijing, China (Approval Number: 009-2014).

369 We sequenced 217 samples following the BGISEQ-500 protocol outsourced to BGI. 1μg genomic
370 DNA was randomly fragmented by Covaris, the fragmented DNA was selected by Agencourt
371 AMPure XP-Medium kit to an average size of 200-400bp, followed by adapter ligation, PCR
372 amplification, and the products were recovered by the AxyPrep Mag PCR clean up kit. The double-
373 stranded PCR products were heat-denatured and circularized by the splint oligo sequence. The
374 single-strand circle DNA (ssCir DNA) was formatted as the final library and qualified by QC.
375 Sequencing was performed on BGISEQ-500 platform with an average depth of 10X.

376 Total RNA was extracted from the brain tissue using Trizol (Invitrogen, Carlsbad, CA, USA)
377 according to manufacturers' instructions. Then, total RNA was qualified and quantified using a
378 Nano Drop and Agilent 2100 bioanalyzer (Thermo Fisher Scientific, MA, USA). Ribo-zero
379 method was used to remove the rRNA. Purified mRNA was fragmented into small pieces with
380 fragment buffer at an appropriate temperature. The cDNAs were purified by magnetic beads. After
381 purification, A-Tailing Mix and RNA Index Adapters were added by incubating to carry out end
382 repair. The cDNA fragments with adapters were amplified by PCR, and the products were purified
383 by Ampure XP Beads. The library was validated on the Agilent Technologies 2100 bioanalyzer
384 for quality control. The final library was amplified with phi29 (Thermo Fisher Scientific, MA,
385 USA) to make DNA nanoball (DNB), DNBs were loaded into the patterned nanoarray and single
386 end 50 base reads were generated on BGISEQ500 platform.

387 **Data quality control**

388 Raw sequencing reads were filtered to get clean reads by using SOAPnuke (v1.5.6)[34], and
389 FastQC[35] was used to evaluate the quality of sequencing data via several metrics, including
390 sequence quality per base, sequence duplication levels, and quality score distribution for each

391 sample. The average quality score for overall DNA and RNA sequences was above 30, indicating
392 that a high percentage of the sequences had high quality.

393 **Variant identification**

394 Clean DNA sequencing reads were mapped to the human reference genome hg19 (GRCh37) using
395 BWA-MEM algorithm (BWA v. 0.7.128)[36]. Ambiguously mapped reads (MAPQ <10) and
396 duplicated reads were removed using SAMtools v. 1.29[37] and PicardTools v. 1.1 respectively.
397 Genomic variants were called following the Genome Analysis Toolkit software (GATK v. 3.4.4.6)
398 best practices.

399 **Population validation, imputation, and filtering**

400 We used PLINK to infer the genomic ancestry of each sample in this study by combining our
401 genotype data and the genotype data from the 1000 Genomes Project[38]; no sample was excluded.
402 Using Michigan Imputation Server[41], EAS genotypes were imputed into the 1000 Genomes Project
403 phase 3 EAS reference panel by chromosome and subsequently merged. Imputed genotypes were
404 filtered for LD $R^2 < 0.3$, Hardy-Weinberg equilibrium p-value < 10e-6 and MAF < 0.05, resulting
405 in ~ 6 million autosomal SNPs.

406 For AA population, genotypes were imputed into the 1000 Genomes Project phase 3 AA reference
407 panel by chromosome and subsequently merged. To further confirm the ancestry of the African
408 American samples, all AA samples were evaluated for their ancestry with three broad population
409 groups with PC1 ≥ 25% AFR and < 25% AMR, < 25% EAS, < 25% SAS; clustering of individuals
410 in each broad population group with the 1000 Genomes Project reference populations are shown
411 in Fig.2a.

412 **Sex check and sample swap identification**

413 The sex of each sample was inferred with SNPs using PLINK. In the EAS cohort, two samples
414 were identified as sex-mismatched and were subsequently removed in downstream analysis.
415 Quality control was performed on genotypes using sample Binary Alignment Map (BAM) files to
416 detect any sample identity swaps between the RNA and DNA experiments. The QTLtools match
417 function[39] confirmed that all samples were appropriately matched.

418 **Gene expression quantification and quality control**

419 The RNA-sequencing reads were mapped using STAR (2.4.2a)[40] and the genes and transcripts
420 quantification was performed using RSEM (1.3.0)[41]. Raw read counts were log-transformed using
421 R package VOOM[42], thereafter filtering those with log2(CPM) < 0 in more than 75% of the
422 samples. Mitochondrial DNA and X and Y chromosome-derived transcripts were excluded.
423 Samples with a Z-score (measured for inter-sample connectivity) less than -3 were also discarded.
424 Finally, quantile normalization was utilized to equalize distributions across samples.

425 **Covariate selection**

426 To measure technical covariates, quality control metrics were collected using STAR, PicardTools
427 v1.139 and RNASeQC. Principal components of the metrics data were calculated and included as
428 SeqPCs for covariate selection. Hidden covariates were measured using probabilistic estimation
429 of expression residuals (PEER)[43] and found to be significantly correlated with technical and
430 biological covariates such as experimental batch, RNA Integrity Number (RIN), sex, and age of
431 death. Based on the Bayesian information criterion (BIC) score, redundant covariates were
432 removed to avoid overfitting. A forward and backward selection procedure was followed, and
433 PEER hidden factors and known covariates were added. The covariate with the higher BIC score
434 was selected for subsequent QTL mapping. PEER was run with varying numbers of inferred
435 hidden factors: 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50, independently.

436 **cis-eQTL mapping**

437 Cis-eQTL mapping was performed using QTLtools, accounting for PEER factors, with a defined
438 cis window spanning one megabase upstream and downstream of the gene/intron cluster body.
439 Using the PEER from gene expression matrix, twenty hidden covariates were identified and
440 adjusted. To detect all available QTLs, QTLtools was conducted in nominal pass mode. To identify
441 the best nominal associated SNP per phenotype, QTLtools was executed in the permutation pass
442 mode. Additionally, to identify SNPs with independent effects on regulating gene expression,
443 QTLtools was run in the conditional pass mode. SNPs with P-values below the threshold of the
444 permutation threshold is classified as significant QTLs.

445 To address potential sample size disparities that could impact the results, the EUR data were
446 randomly sampled with various sample size (150, 200, 250, 300, 350, and 400) and applied the
447 same analytical pipeline while exploring the relationship between sample size and number of QTLs.

448 **eQTL meta-analysis**

449 The SNP-gene pairs with a genome-wide P value reaching the threshold of 10e-6 for each
450 population were collected and performed the meta-analysis. Standard fixed-effects meta-analysis
451 were used to combine all data into a single regression model by METAL[44]. The meta-analysis
452 assumes a fixed-effects size, as well as constant error variance, across all data.

453 **eQTL fine-mapping**

454 The initial step of fine-mapping involved using the in-sample LD of the three populations. We
455 extracted common variants with MAF > 5% from each group and used PLINK to determine the
456 LD regions of these common variants for each population. To eliminate strand flipping and
457 alignment issues, multi-allelic variations and indels were removed. Next, SuSiEx was applied to

458 merge the eQTLs summary statistics from the three groups. Credible set is defined as a set of
459 putative causal variants. A credible set was discarded if it lacked genetic variants reaching genome-
460 wide significance ($p < 1e-6$) in either the population-unique eQTLs or cross-population meta-
461 eQTLs. By considering prior knowledge and the observed data, this method provides a posterior
462 probability (PIP) for each variant being the causal one in the associated region. Variants with high
463 PIPs are then considered strong candidates for functional follow-up studies.

464 **Functional enrichment**

465 To determine functional enrichment, GREGOR[48] was performed to test the eQTL enrichment.
466 GREGOR calculated the enrichment value based on the observed and expected overlap within
467 each annotation. To conduct our analysis, the 15-state ChromHMM model BED (Browser
468 Extensible Data) files from the Roadmap Epigenetics Project[45], and 78 consensus transcription
469 factor and DNA-protein binding site BED files existing in multiple cells were downloaded. Fifty
470 binding proteins showed cortical brain expression in EAS and AA populations data[46].

471 **The fraction of shared eQTLs between non-EUR and EUR populations**

472 Sharing rate was assessed based on significant eQTLs in the discovery dataset by estimating the
473 proportion of true associations (π_1) on the distribution of corresponding p-values of the overlapping
474 eQTLs in the replication dataset[47].

475 **F_{ST} and MAF analysis**

476 Fixation index (F_{ST}) was estimated using vcftools following the Weir and Cockerham approach
477 for each eSNP[48]. The population-divergent SNPs were defined as those with $F_{ST} >= 0.05$ and
478 population-shared SNPs as those with $F_{ST} < 0.05$. To generate the list of population-unique QTLs

479 and population-shared QTLs, we collected the overlap of eQTLs from the pairwise comparisons
480 of the list of AA eQTLs, EAS eQTLs, and EUR eQTLs. Finally, Fisher's exact test was performed
481 between population-unique QTLs and population-shared QTLs to test the contribution of MAF in
482 the QTL comparison.

483 **Variance explained**

484 Variance explained, which combines the effect size (beta) and frequency of the allele (f), can be
485 considered an approximate measure of a causal variant's importance within a population. Variance
486 is approximated using the formula $2f(1 - f)\log(\text{beta})^2/(\pi^2/3)$[49]. Although these variants often exhibit
487 similar odds ratios across populations, their allele frequencies may differ. By considering both the
488 effect size (OR) and the frequency of the risk allele (f), the variance explained offers a valuable
489 approximation of a causal variant's significance within a given population.

490 **Power estimation**

491 We used R to calculate the sample size needed to achieve a given power level in a chi-square test,
492 based on an assumed effect size and a significance threshold. This analysis starts by setting initial
493 values for power, effect size, and p-value threshold. Then, the critical chi-square statistic required
494 to meet the power level is computed. A function, *calculate_ncp*, is defined to calculate the non-
495 centrality parameter from the p-value and degrees of freedom, adjusting for the critical chi-square
496 statistic. Subsequently, the non-centrality parameter is computed for the given power and p-value
497 threshold. Another function, *af_n_relation*, is created to determine the relationship between allele
498 frequency and sample size, incorporating the effect size and the non-centrality parameter. Finally,
499 the code iteratively solves for the sample size corresponding to a range of allele frequencies, thus

500 enabling the determination of the necessary sample size for different allele frequencies to maintain
501 the specified power level in the chi-square test.

502 **Partitioned LDSR**

503 Partitioned LD score regression v1.0.1[50] was used to measure the enrichment of GWAS summary
504 statistics in each functional category by accounting for LD. Brain QTL annotations were created
505 by eSNP, mapped to the corresponding 1000 Genome reference panel. LD scores were calculated
506 for each SNP in the QTL annotation using an LD window of 1cM in 1000 Genomes European
507 Phase 3 and 1000 Genomes Asian Phase 3 separately. Enrichment for each annotation was
508 calculated by the proportion of heritability explained by each annotation divided by the proportion
509 of SNPs in the genome falling in that annotation category. We then applied Welch Modified Two-
510 Sample t-Test on enrichment values generated from QTLs in the two populations.

511 **Colocalization**

512 Conditional association was used to test for evidence of colocalization. This method compares the
513 p-value of association for the lead SNP of an eQTL before and after conditioning on the GWAS
514 hit. The equation for the regulatory trait concordance (RTC) Score is as follows: RTC= (N_{SNPs} in
515 an LD block/$Rank_{GWAS_SNP}$)/ N_{SNPs} in an LD block. The rank denoted the number of SNPs, which
516 when used to correct the expression data, has a higher impact on the QTL than the GWAS SNPs.
517 RTC values close to 1.0 indicated causal regulatory effects. A threshold of 0.9 was used to select
518 causal regulatory elements.

519 Colocalization of fine-mapped variations from complex traits and cis-eQTL correlations were
520 performed. Based on complex trait and cis-eQTL fine-mapping data, a posterior inclusion

521 probability of colocalization for a variant was calculated as a product of PIP for GWAS and PIP
522 for the cis-eQTLs (PIPcoloc = PIP_{GWAS} * $PIP_{cis\text{-}eQTLs}$).

523 **Summary-data-based Mendelian randomization**

524 SMR[19] was applied on SCZ GWAS summary data to prioritize candidate genes. Significant QTLs
525 identified in the previous analysis (FDR < 0.05), were combined with filtered GWAS summary
526 data (p < 5e-8) to perform the SMR test. In general, we used the default parameters suggested by
527 the developers of the SMR software. These included the application of heterogeneity independent
528 instruments (HEIDI) testing, filtering out hits that arose from significant linkage with
529 pleiotropically associated variants (LD cutoff of P = 0.05 in the HEIDI test, as suggested by SMR).
530 Genes with an empirical P that passed Bonferroni correction in the SMR test and a P > 0.05 in the
531 HEIDI test were considered as risk genes.

532 **Prioritizing genes underlying GWAS hits using TWAS**

533 In this research, we initially developed gene expression prediction models for distinct populations
534 using MetaXcan software[51]. Following this, we integrated these models with GWAS (Genome-
535 Wide Association Studies) summary statistics specifically focused on schizophrenia. This
536 integration aimed to generate gene-level z-scores representing the association of the genetically-
537 determined expression for a gene from its prediction model with the phenotype. TWAS enabled
538 us to compute P-values and subsequently prioritize genes in relation to their association with
539 schizophrenia risk.

540 **Plasmid construction**

541 We obtained the 55 bp SNP-centered DNA sequence from UCSC Genome Browser
542 (GRCh38/hg38), then added the sticky end of restriction enzymes KpnI and NheI at both ends of

543 the 55bp sequence to synthesis primers. Primer annealing to obtained double-strand sequence and
544 then inserted into pGL3-Promoter Vector (Promega) using FastDigest enzymes (ThermoFisher)
545 and T4 DNA Ligase (Invitrogen). We valid the vector sequence using sanger sequence by primer
546 GLP2 and RVP3.

547 **Dual Luciferase Reporter Assay**

548 For transfection, we used SH-SY5Y and HS-683 cells to perform the experiments. Transfecting
549 cells at 50-60% confluency, cells were co-transfected with 500ng reconstruction vector and 10ng
550 pRL-TK using Lipofectamine 3000 Transfection Reagent (ThermoFisher) in 24 well plates. After
551 48h transfection, using Dual Luciferase Reporter Assay Kit (Promega) to measure the firefly
552 luciferase activity and renilla luciferase activity, the luminescence was detected using Tube
553 Luminometer (Berthold Sirius).

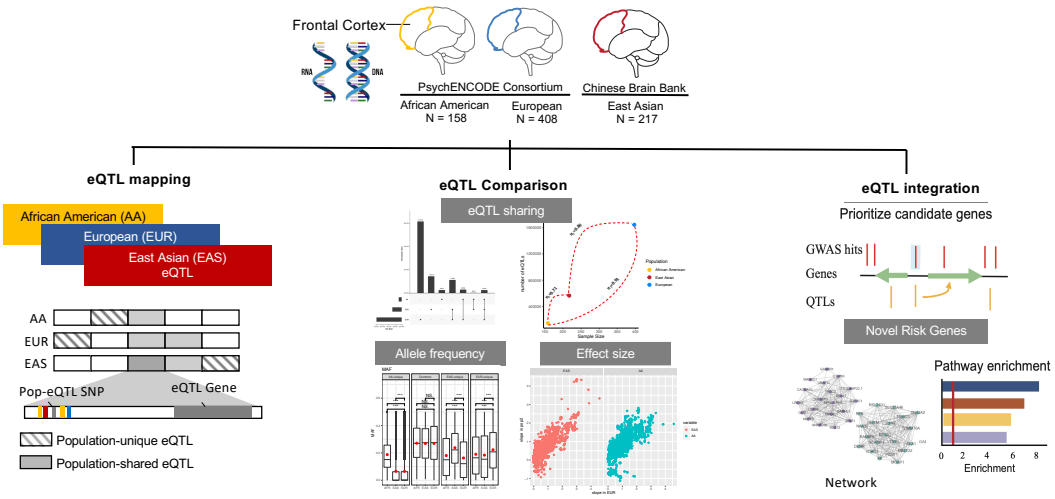


Fig. 1. Study design. We examined genotype and RNA-seq data from African Americans (AA), Europeans (EUR), and East Asians (EAS) to identify expression quantitative trait loci (eQTLs) unique to non-European populations and their role in schizophrenia risk.

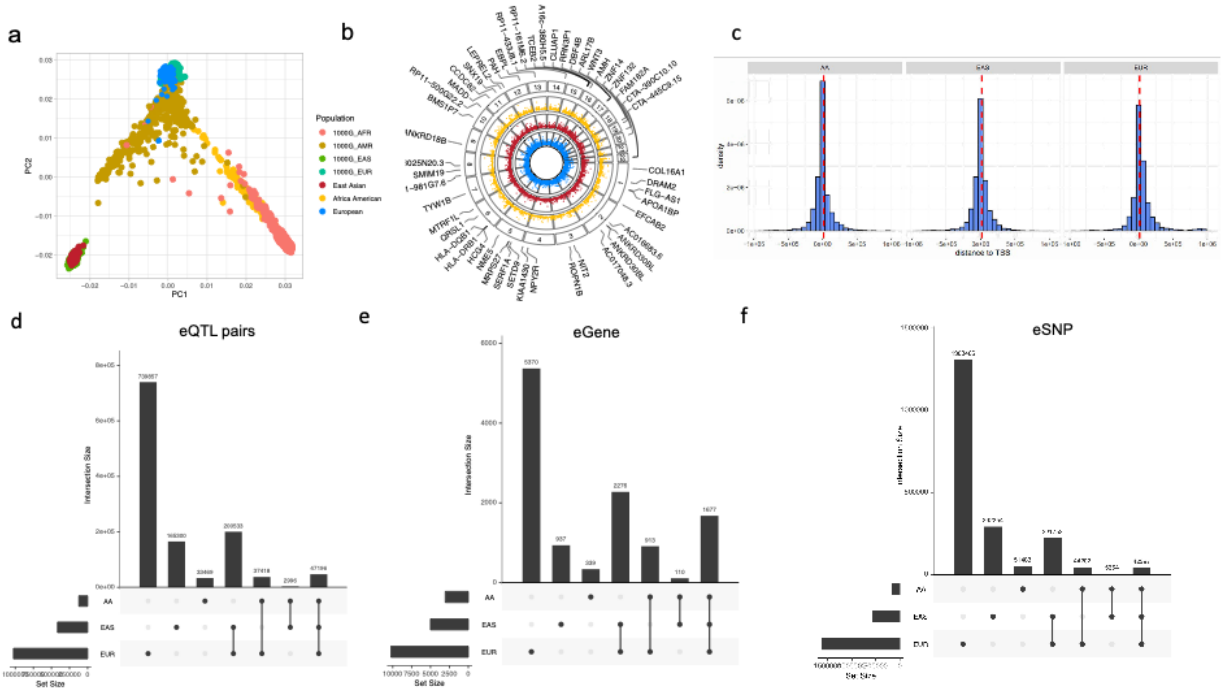


Fig. 2: Identification and characterization of eQTLs. a) PCA plot showing the population structure of individuals in our study as well as the 1000 Genomes Project. AFR: African; AMR: American; EAS: East Asian; EUR: European; b) Circos manhattan plot of significant eQTL genes among the three populations with highlighted top 50 fine-mapped eGenes. d-f) Upset plot showed overlap among the significantly associated eQTL pairs (d) eGenes (e) as well as eSNPs (f) between populations.

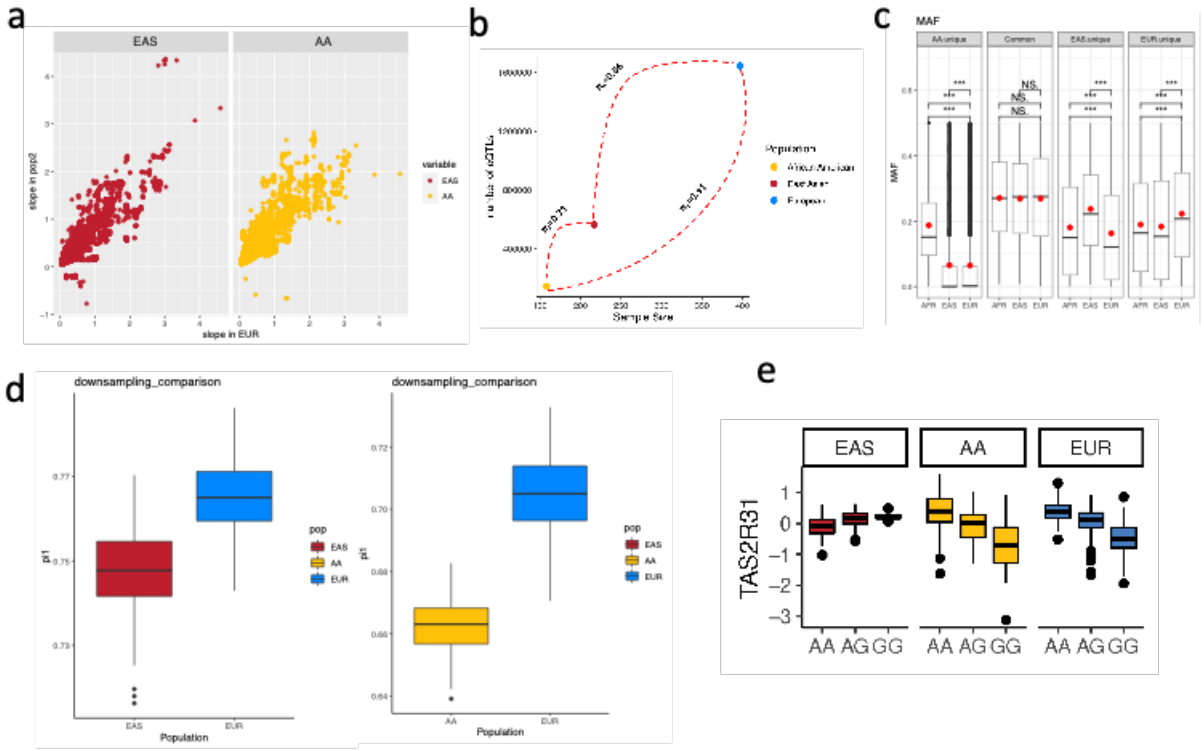
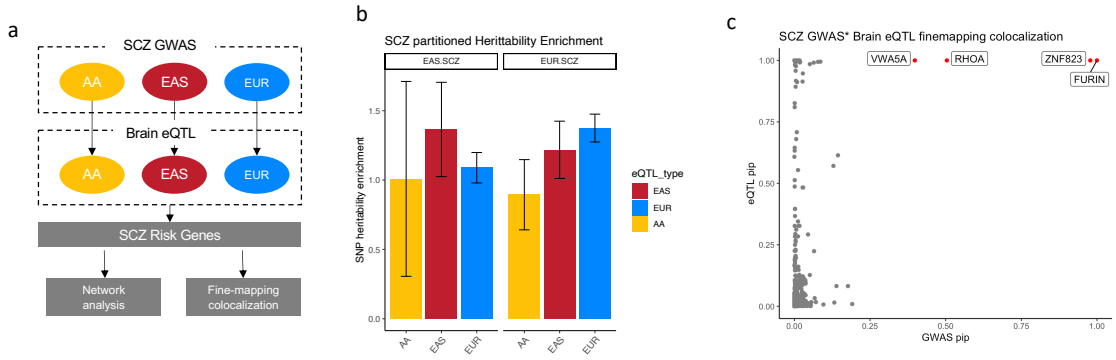


Fig. 3: Analysis of the regulatory patterns. a) relationship between sample size and the # of detected eQTLs. b) Effect sizes for common (MAF> 1%) sentinel cis-eQTLs across EA and AA populations; c) Comparison of MAF between population-shared and non-EUR unique eSNPs; d) downsampling to estimate π_1 between non-European and European eQTLs . e) example of opposite effect eQTL *TAS2R31*- rs2599400.

591



592

593 Fig. 4: Explanation of SCZ GWAS signals and prioritization of candidate genes. a) integration

594 strategies; b) GWAS enrichment results from LDSR. ***: Welch Modified Two-Sample t-Test P

595 < 0.001; c) Fine-mapped colocalization results. Each point represents an eQTL pais, the x-axis

596 represents the GWAS pip for that eSNP, and the y-axis represents the eQTL pip for that eSNP.

597 Red points represent pip_{GWAS}*pip_{eQTL}>0.1 and are labeled with the eGene

598

599

600

601

602

603 **Data availability**

604 The raw sequence data reported in this paper have been deposited in the Genome Sequence
605 Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences,
606 under accession numbers HRA000108, HRA000108 that can be accessed at
607 https://bigd.big.ac.cn/gsa-human. eQTLs summary results can be downloaded while requesting to
608 the corresponding author.

609 **Code availability**

610 Codes are available at https://github.com/liusihan/population-compare-pipeline.

611 **Acknowledgments**

612 We thank all donors and their families. Tissue was provided by the Human Brain Bank, Chinese
613 Academy of Medical Sciences & Peking Union Medical College, Beijing, China, the Chinese
614 Brain Bank Center, and the Xiangya School of Medicine Brain Bank. We acknowledge Stanley
615 Center for Psychiatric Research for supporting Hailiang Huang and Yu Chen working on this
616 project.

617 **Funding**

618 This study was supported by National Natural Science Foundation of China (Grants Nos.
619 82022024, 31571312, and 91632116), the National Key R&D Project of China (Grants No.
620 2016YFC1306000), The Science and Technology Innovation Program of Hunan Province
621 (2021RC4018, 2021RC5027), and NIH grant 1R01MH126459-01A1.

622 **Author contributions**

Y.C. and S.L. wrote the manuscript, analyzed the data and performed all computations. F.W. and Y. J. contributed to eQTLs analysis of EUR population. Z.R., F.D., C.H and M.L. extracted DNA and RNA, as well as collected sample information. R.D., Y.X., and R.K. substantively revised the manuscript. Z.N., S.X. and H.H. participated in the design of comparing the brain regulatory architecture. K.Y., W.Q., C.M., X.Y., A.B., J.D., J.H., B.T. and C.L. provided primary sequenced samples and data including their clinical information. C.M., C.L. and C.C. conceived, designed and supervised the study and modified the manuscript.

Conflict of Interest

All the authors declare no competing financial interests.

Ethnics

Ethics Committee of Central South University gave ethical approval for this work (2015031007).

637 **References**

638 1. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
639 schizophrenia. *Nature* **604**, 502–508 (2022).
640 2. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations:
641 Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
642 3. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**,
643 175–185 (2018).
644 4. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–
645 585 (2013).
646 5. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide
647 range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
648 6. Wang, L. *et al.* Brain Banks Spur New Frontiers in Neuropsychiatric Research and Strategies
649 for Analysis and Validation. *Genomics Proteomics Bioinformatics* **17**, 402–414 (2019).
650 7. Mogil, L. S. *et al.* Genetic architecture of gene expression traits across diverse populations.
651 *PLoS Genet.* **14**, e1007586 (2018).
652 8. de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal
653 downstream effects and putative drivers for brain-related diseases. *Nat. Genet.* **55**, 377–388
654 (2023).
655 9. Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal
656 variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022).
657 10. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with
658 the 'Sum of Single Effects' model. *PLoS Genet.* **18**, e1010299 (2022).

659 11. Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in
660 epigenomic features using a systematic, data-driven approach. *Bioinforma. Oxf. Engl.* **31**,
661 2601–2606 (2015).
662 12. Pybus, M. *et al.* 1000 Genomes Selection Browser 1.0: a genome browser dedicated to
663 signatures of natural selection in modern humans. *Nucleic Acids Res.* **42**, D903-909 (2014).
664 13. Ning, Z. *et al.* Expression profiles of east-west highly differentiated genes in Uyghur
665 genomes. *Natl. Sci. Rev.* **10**, nwad077 (2023).
666 14. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and
667 European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
668 15. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics
669 Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu & Bipolar Disorder and
670 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic
671 Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**,
672 1705-1715.e16 (2018).
673 16. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia
674 among individuals of African and Latino ancestry. *Mol. Psychiatry* **25**, 2455–2467 (2020).
675 17. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide
676 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
677 18. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with
678 complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
679 19. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex
680 trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

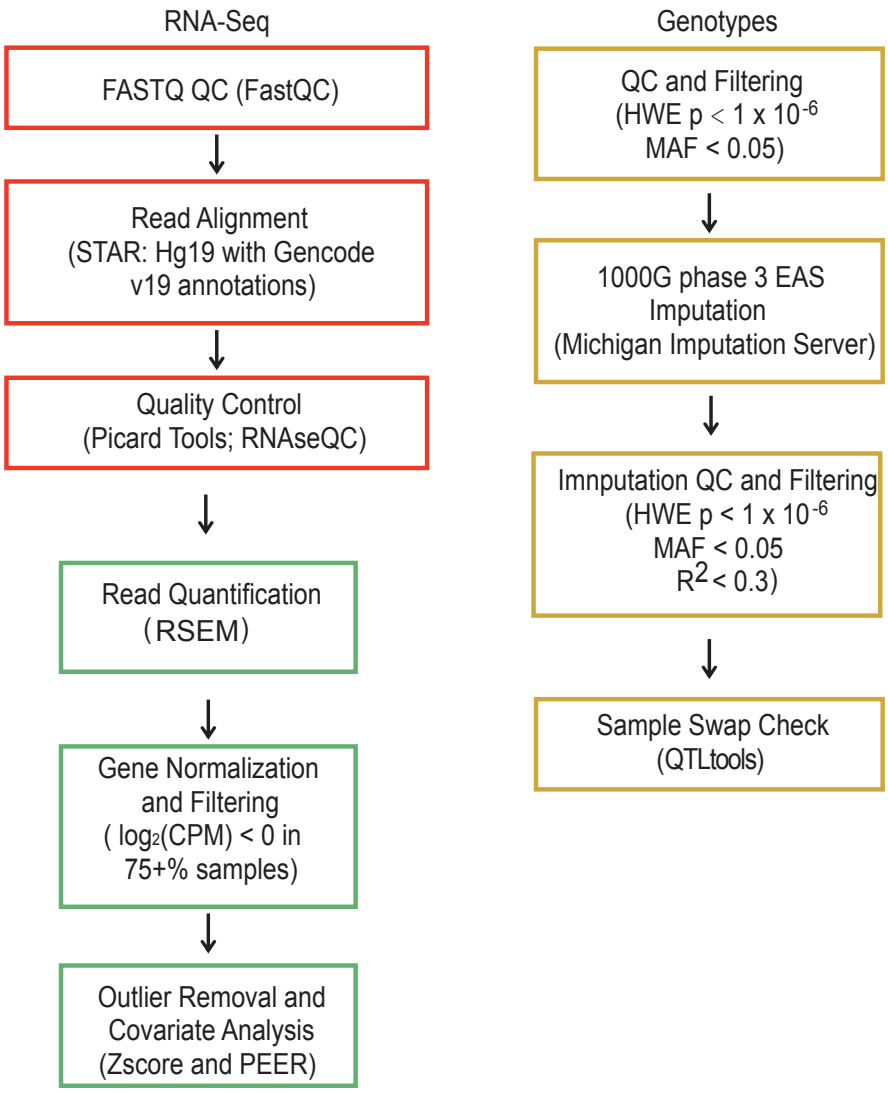
681 20. Feng, X., Liu, S., Li, K., Bu, F. & Yuan, H. NCAD v1.0: A database for non-coding variant
682 annotation and interpretation. *J. Genet. Genomics Yi Chuan Xue Bao* S1673-8527(23)00257–
683 6 (2023) doi:10.1016/j.jgg.2023.12.005.
684 21. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.*
685 **28**, 1243–1252 (2018).
686 22. Kachuri, L. *et al.* Gene expression in African Americans, Puerto Ricans and Mexican
687 Americans reveals ancestry-specific patterns of genetic architecture. *Nat. Genet.* **55**, 952–963
688 (2023).
689 23. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation
690 in humans. *Nature* **501**, 506–511 (2013).
691 24. Fraser, H. B., Lam, L. L., Neumann, S. M. & Kobor, M. S. Population-specificity of human
692 DNA methylation. *Genome Biol.* **13**, R8 (2012).
693 25. Tehranchi, A. *et al.* Fine-mapping cis-regulatory variants in diverse human populations. *eLife*
694 **8**, e39595 (2019).
695 26. Ning, Z. *et al.* Expression profiles of east-west highly differentiated genes in Uyghur
696 genomes. *Natl. Sci. Rev.* **10**, nwad077 (2023).
697 27. Wooding, S. P. & Ramirez, V. A. Global population genetics and diversity in the TAS2R
698 bitter taste receptor family. *Front. Genet.* **13**, 952299 (2022).
699 28. Andrade, A. *et al.* Genetic Associations between Voltage-Gated Calcium Channels and
700 Psychiatric Disorders. *Int. J. Mol. Sci.* **20**, 3537 (2019).
701 29. Heringa, S. M., Begemann, M. J. H., Goverde, A. J. & Sommer, I. E. C. Sex hormones and
702 oxytocin augmentation strategies in schizophrenia: A quantitative review. *Schizophr. Res.*
703 **168**, 603–613 (2015).

704 30. de Boer, J., Prikken, M., Lei, W. U., Begemann, M. & Sommer, I. The effect of raloxifene
705 augmentation in men and women with a schizophrenia spectrum disorder: a systematic
706 review and meta-analysis. *NPJ Schizophr.* **4**, 1 (2018).
707 31. Chiappelli, J. *et al.* Disrupted glucocorticoid--Immune interactions during stress response in
708 schizophrenia. *Psychoneuroendocrinology* **63**, 86–93 (2016).
709 32. Yan, X.-X., Ma, C., Bao, A.-M., Wang, X.-M. & Gai, W.-P. Brain banking as a cornerstone
710 of neuroscience in China. *Lancet Neurol.* **14**, 136 (2015).
711 33. Qiu, W. *et al.* Standardized Operational Protocol for Human Brain Banking in China.
712 *Neurosci. Bull.* **35**, 270–276 (2019).
713 34. Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated
714 quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**, 1–6
715 (2018).
716 35. Trivedi, U. H. *et al.* Quality control of next-generation sequencing data without a reference.
717 *Front. Genet.* **5**, 111 (2014).
718 36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
719 *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
720 37. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**,
721 2078–2079 (2009).
722 38. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
723 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
724 39. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat.*
725 *Commun.* **8**, 15452 (2017).

726 40. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–
727 21 (2013).
728 41. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or
729 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
730 42. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model
731 analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
732 43. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of
733 expression residuals (PEER) to obtain increased power and interpretability of gene
734 expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
735 44. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
736 genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).
737 45. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
738 epigenomes. *Nature* **518**, 317–330 (2015).
739 46. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor
740 binding sites. *Nat. Genet.* **45**, 723–729 (2013).
741 47. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl.*
742 *Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
743 48. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST:
744 the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
745 49. Pawitan, Y., Seng, K. C. & Magnusson, P. K. E. How many genetic variants remain to be
746 discovered? *PloS One* **4**, e7969 (2009).
747 50. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies
748 disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

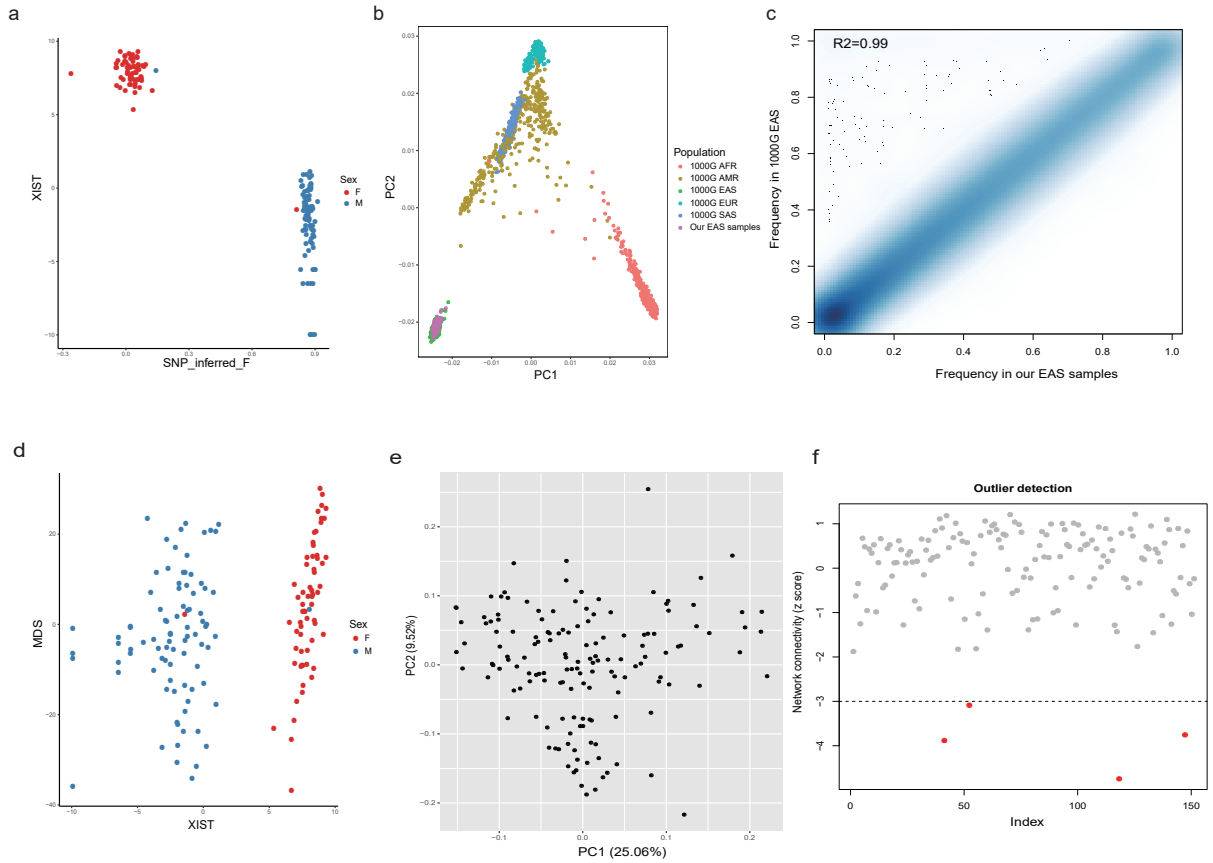
749 51. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene
750 expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
751

752

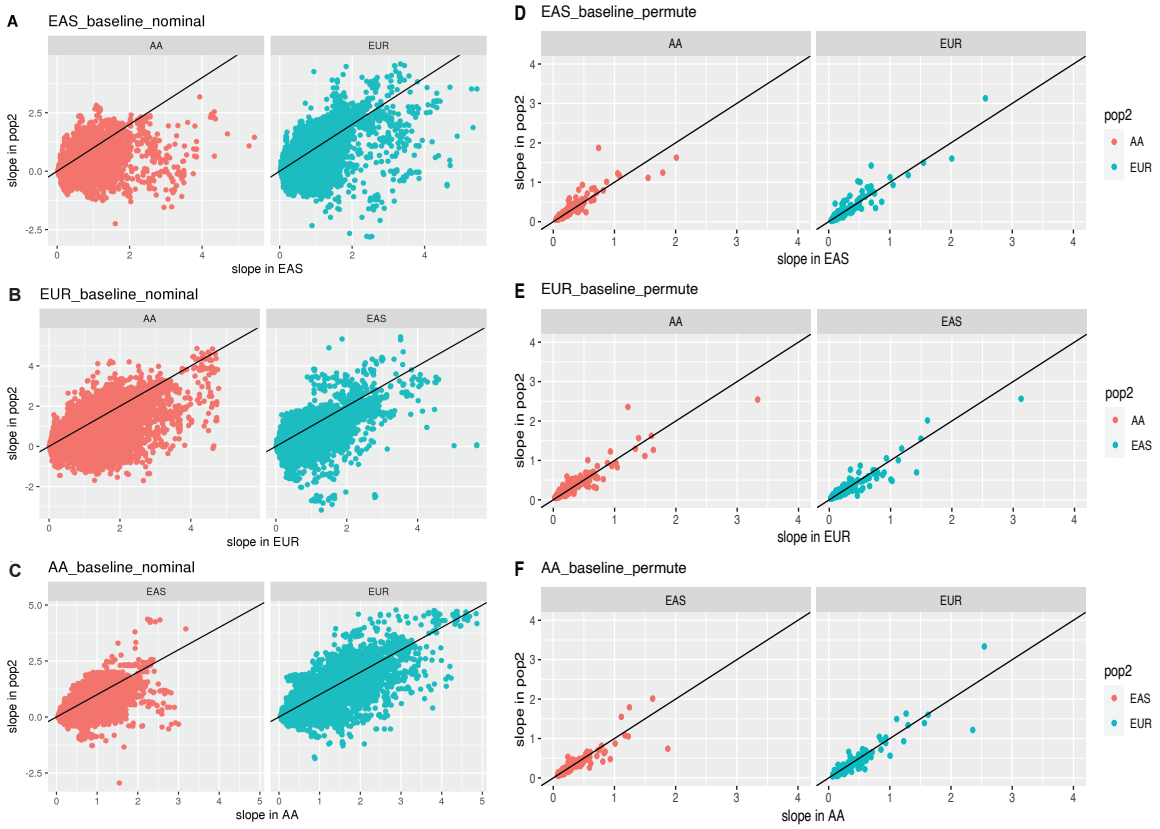


753

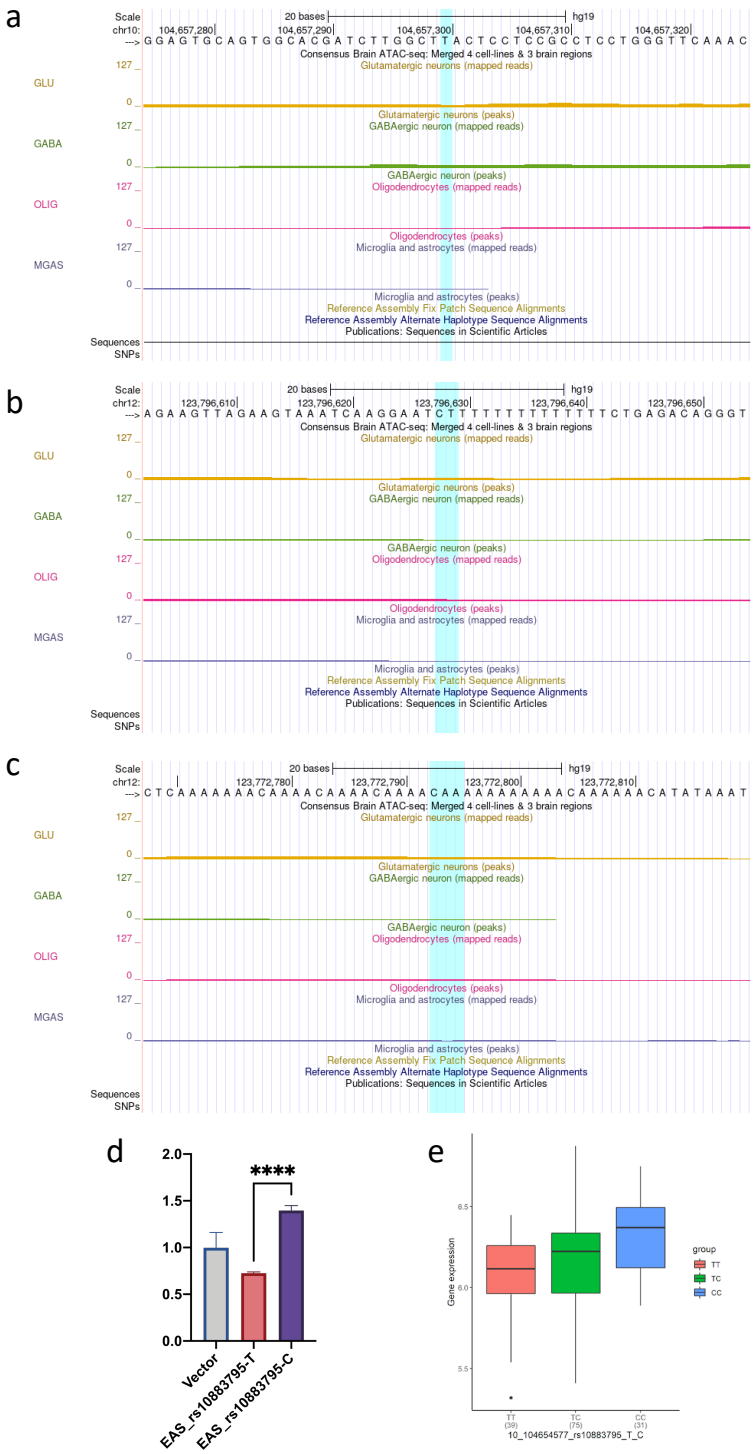
754 Extended Data Figure 1**:** Overview of methods and QC pipeline for EAS samples.



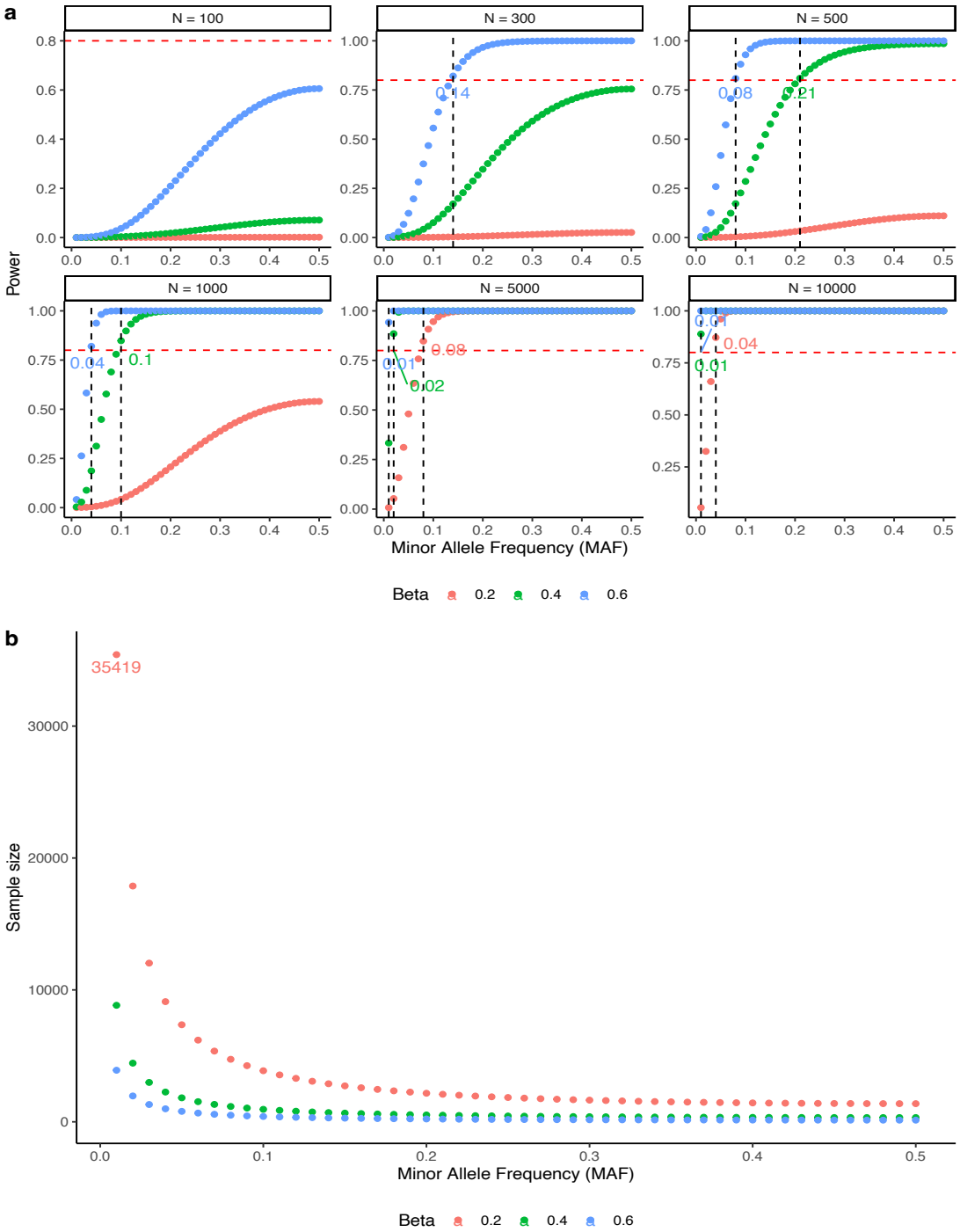
Extended Data Figure 2: Preprocessing of RNA-sequencing and whole-genome sequencing (WGS) data of EAS samples. a, Sex-mismatch checked by WGS data. b, Population PCA plot with 1000G genotype data. c, Imputation accuracy. d, Sex-mismatch checked by Xist expression. e, PCA plot for EAS samples. f, Distribution of Z-score.



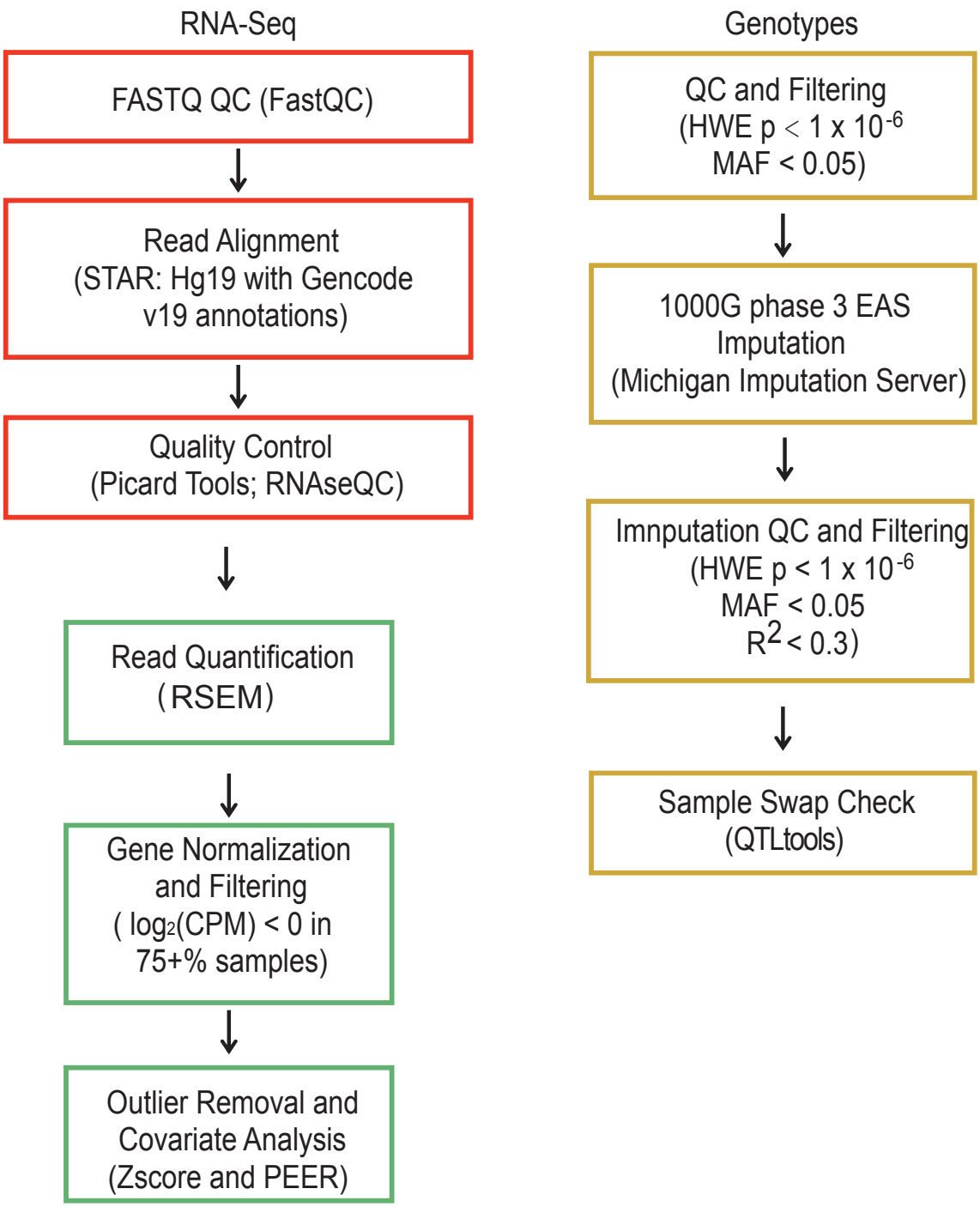
Extended Data Figure 3: Effect size correlation of population-shared eQTLs between EUR and non-EUR population. (a-c) nominal pass; (d-f) permutation pass

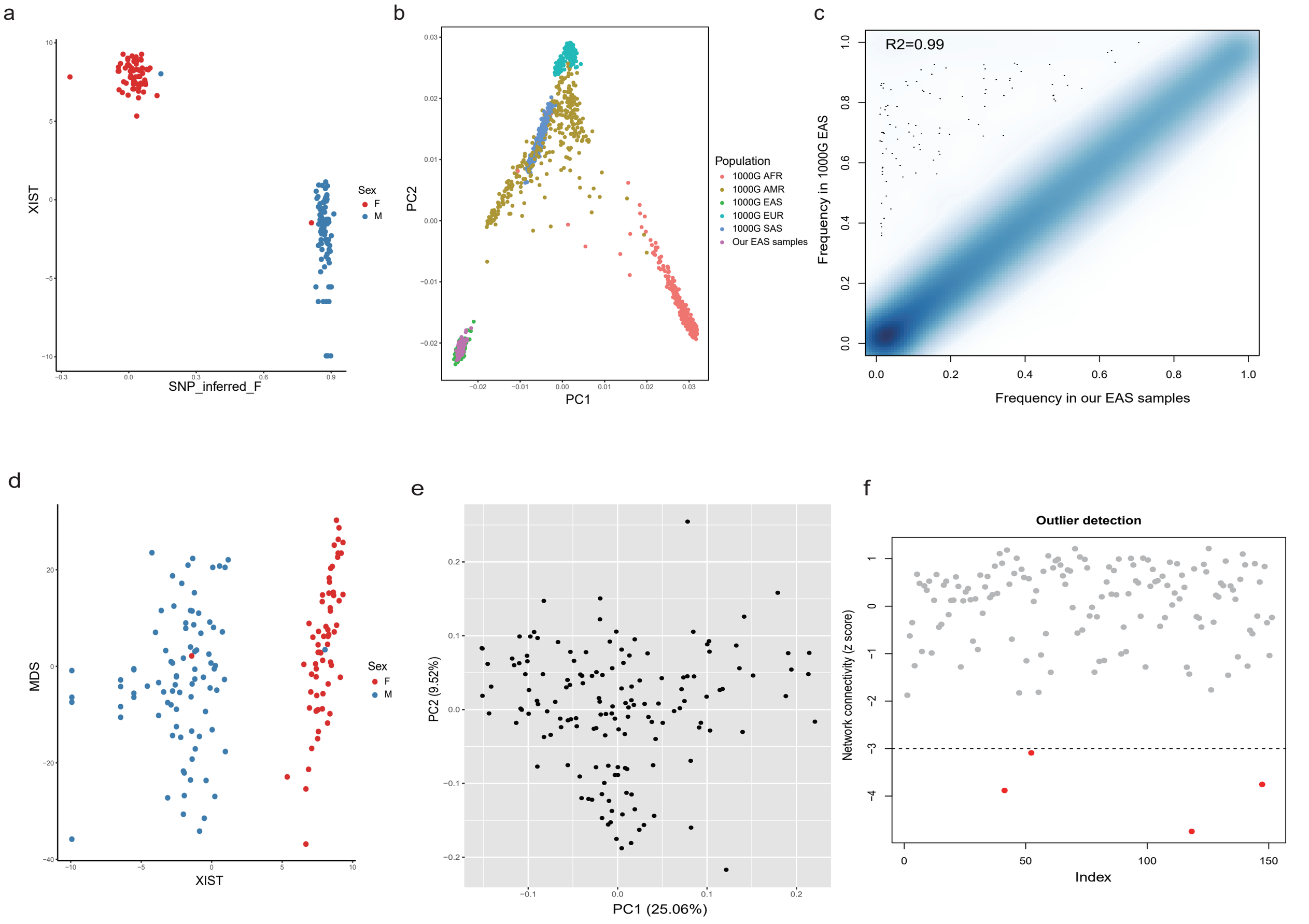


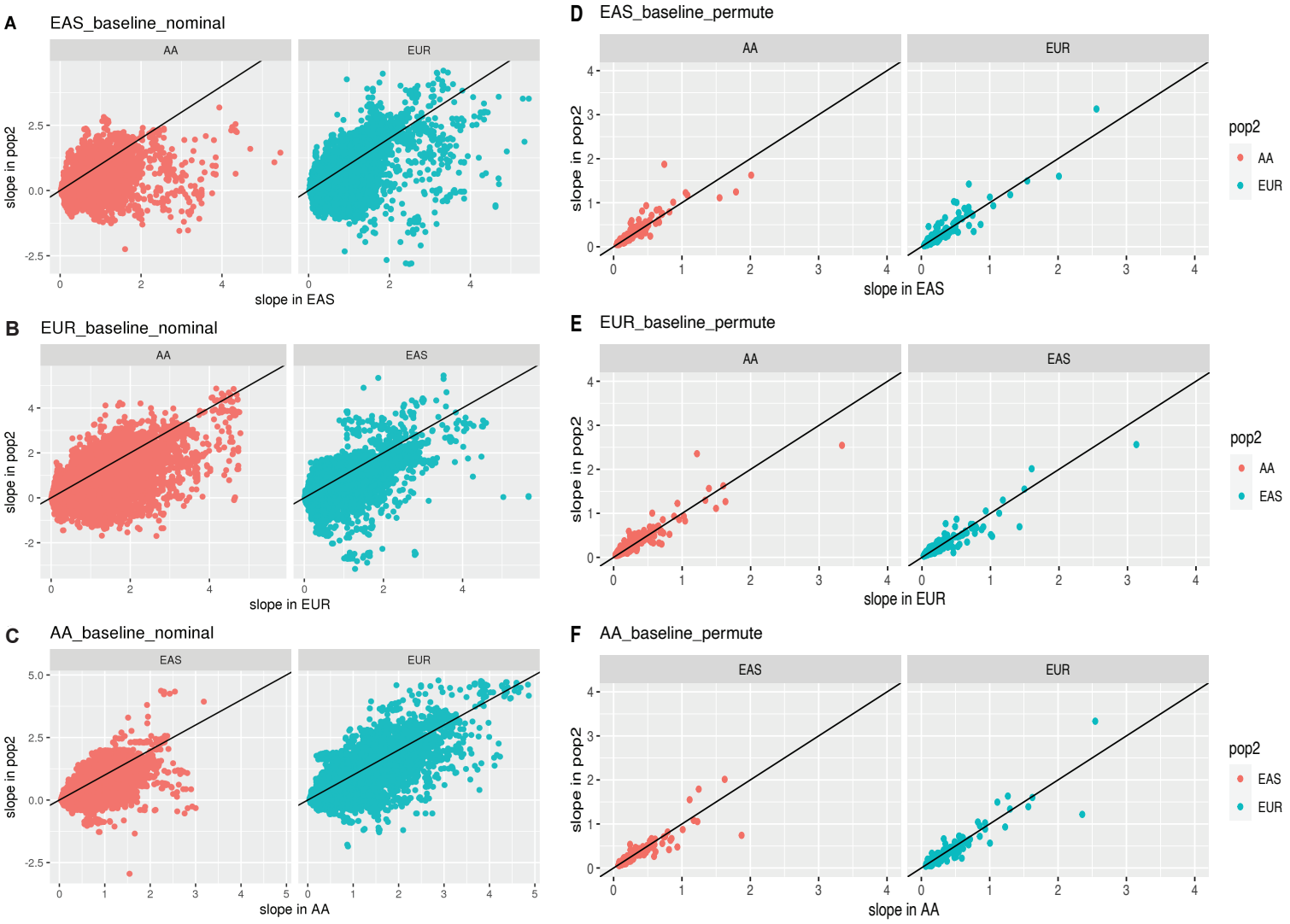
Extended Data Figure 4: Regulatory effect across glutamate neuron, GABA neuron, oligodendrocytes, and microglia for new regulatory SNPs within population-shared risk genes for (a) CNNM2, (b)C12orf65, (c) MPHOSPH9. (d) Dual luciferase reporter assay for EAS eSNP at risk gene CNNM2. (f) eQTL result for the eSNP and expression risk gene CNNM2 in EAS cohort

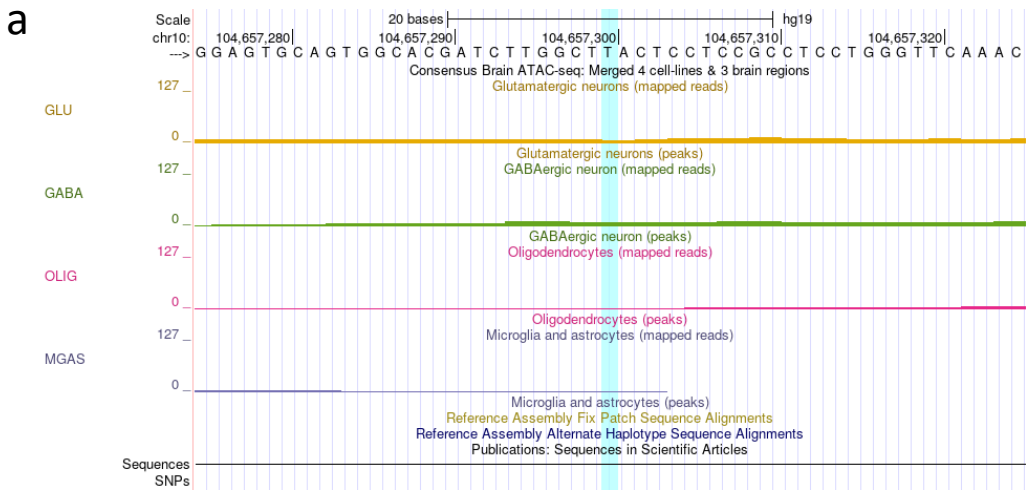


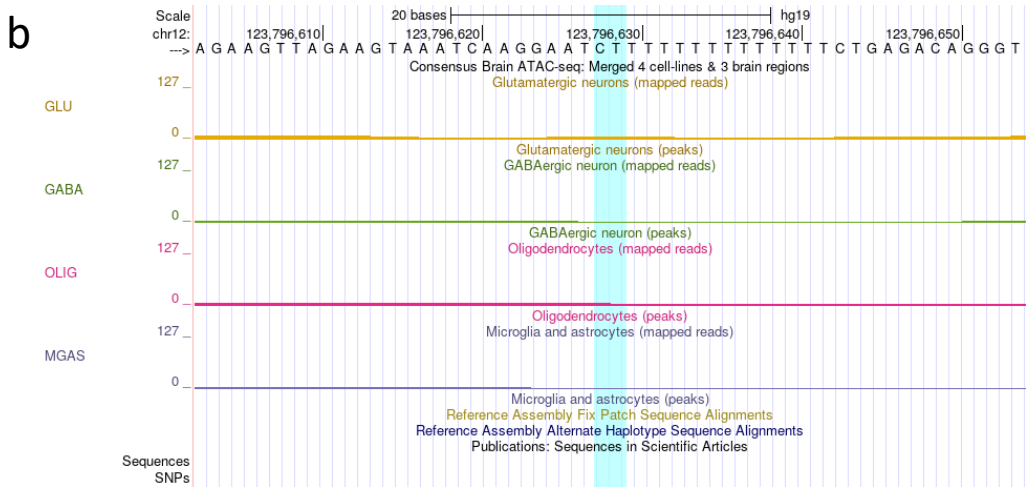
Extended Data Figure 5: The sample size required for well-powered brain eQTL detection in diverse populations. (a) The percentage of brain eQTLs detected power under different sample sizes and effect sizes is shown as a function of log-scaled sample size. (b) The required sample size achieving 80% power based on the effect size estimated form non-EUR specific eQTLs.

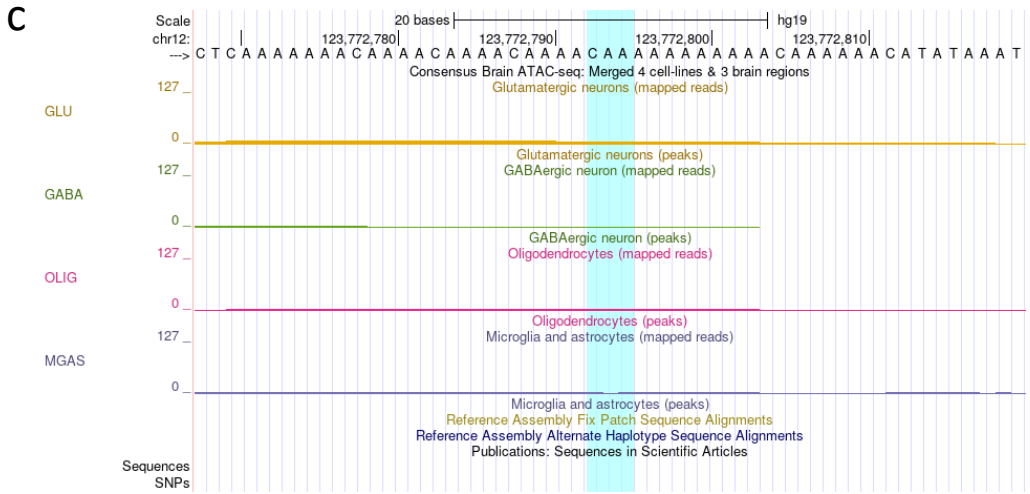

RNA-Seq
FASTQ QC (FastQC)
Read Alignment
(STAR: Hg19 with Gencode
v19 annotations)
Quality Control
(Picard Tools; RNAseQC)
Read Quantification
(RSEM)
Gene Normalization
and Filtering
(log2(CPM) < 0 in
75+% samples)
Outlier Removal and
Covariate Analysis
(Zscore and PEER)
Genotypes
QC and Filtering
(HWE p < 1 x 10-6
MAF < 0.05)
1000G phase 3 EAS
Imputation
(Michigan Imputation Server)
Imnputation QC and Filtering
(HWE p < 1 x 10-6
MAF < 0.05
R2 < 0.3)
Sample Swap Check
(QTLtools)

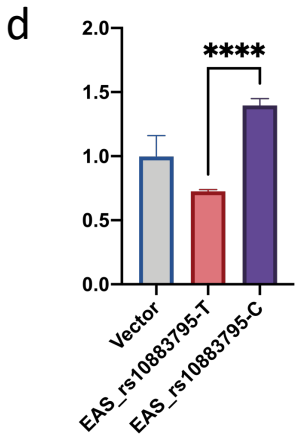

a
XIST
SNP_inferred_F
Sex
F
M
b
PC2
PC1
Population
1000G AFR
1000G AMR
1000G EAS
1000G EUR
1000G SAS
Our EAS samples
c
R2=0.99
Frequency in 1000G EAS
Frequency in our EAS samples
d
MDS
XIST
Sex
F
M
e
PC2 (9.52%)
PC1 (25.06%)
f
Outlier detection
Network connectivity (z score)
Index

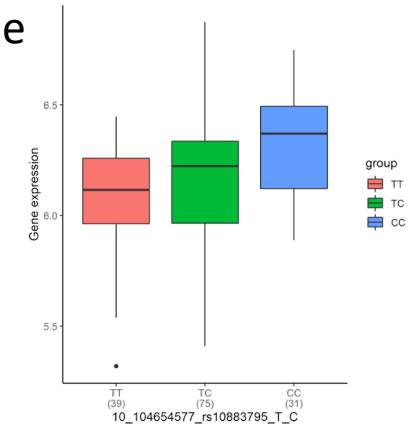

A
EAS_baseline_nominal
AA
EUR
slope in pop2
slope in EAS
B
EUR_baseline_nominal
AA
EAS
slope in pop2
slope in EUR
C
AA_baseline_nominal
EAS
EUR
slope in pop2
slope in AA
D
EAS_baseline_permute
AA
EUR
slope in pop2
slope in EAS
pop2
AA
EUR
E
EUR_baseline_permute
AA
EAS
slope in pop2
slope in EUR
pop2
AA
EAS
F
AA_baseline_permute
EAS
EUR
slope in pop2
slope in AA
pop2
EAS
EUR

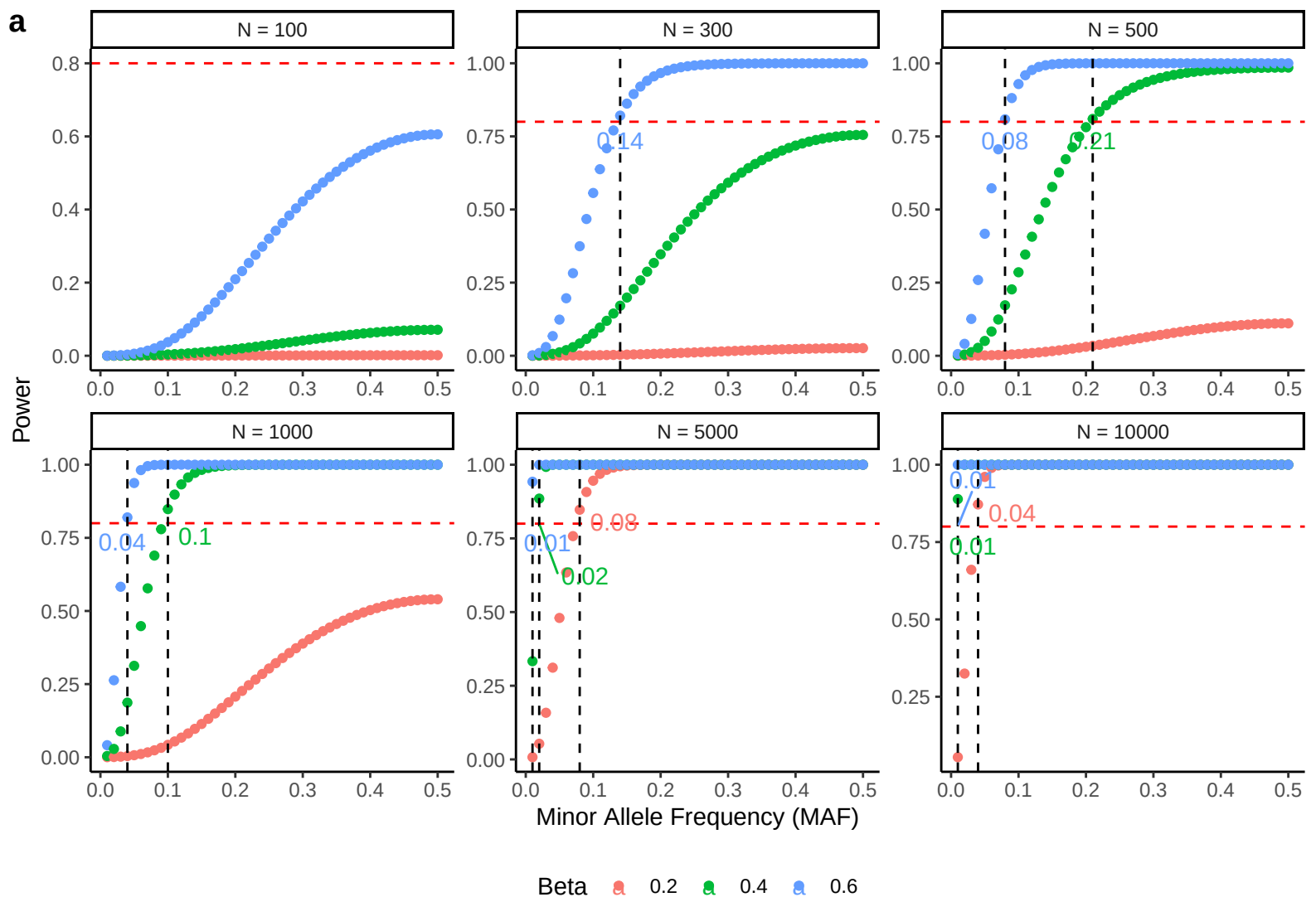

a
Scale
chr10:
20 bases
hg19
104,657,280
104,657,290
104,657,300
104,657,310
104,657,320
---> G G A G T G C A G T G G C A C G A T C T T G G C T T A C T C C T C C G C C T C C T G G G T T C A A A C
Consensus Brain ATAC-seq: Merged 4 cell-lines & 3 brain regions
Glutamatergic neurons (mapped reads)
GLU
Glutamatergic neurons (peaks)
GABAergic neuron (mapped reads)
GABA
GABAergic neuron (peaks)
Oligodendrocytes (mapped reads)
OLIG
Oligodendrocytes (peaks)
Microglia and astrocytes (mapped reads)
MGAS
Microglia and astrocytes (peaks)
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
Publications: Sequences in Scientific Articles
Sequences
SNPs

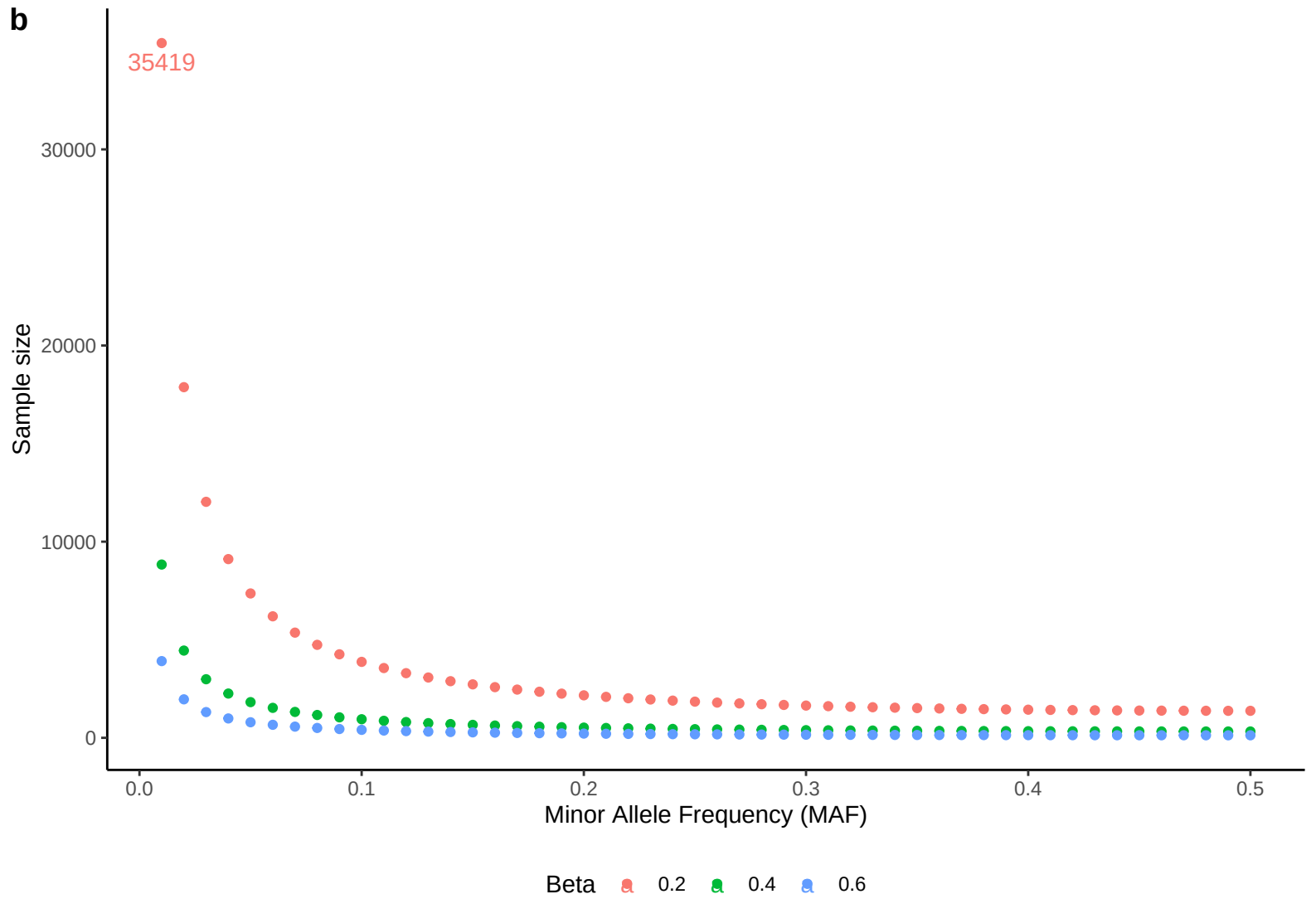

b
Scale
chr12:
20 bases
hg19
123,796,610
123,796,620
123,796,630
123,796,640
123,796,650
---> A G A A G T T A G A A G T A A A T C A A G G A A T C T T T T T T T T T T T T T T T C T G A G A C A G G G T
Consensus Brain ATAC-seq: Merged 4 cell-lines & 3 brain regions
Glutamatergic neurons (mapped reads)
GLU
Glutamatergic neurons (peaks)
GABAergic neuron (mapped reads)
GABA
GABAergic neuron (peaks)
Oligodendrocytes (mapped reads)
OLIG
Oligodendrocytes (peaks)
Microglia and astrocytes (mapped reads)
MGAS
Microglia and astrocytes (peaks)
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
Publications: Sequences in Scientific Articles
Sequences
SNPs

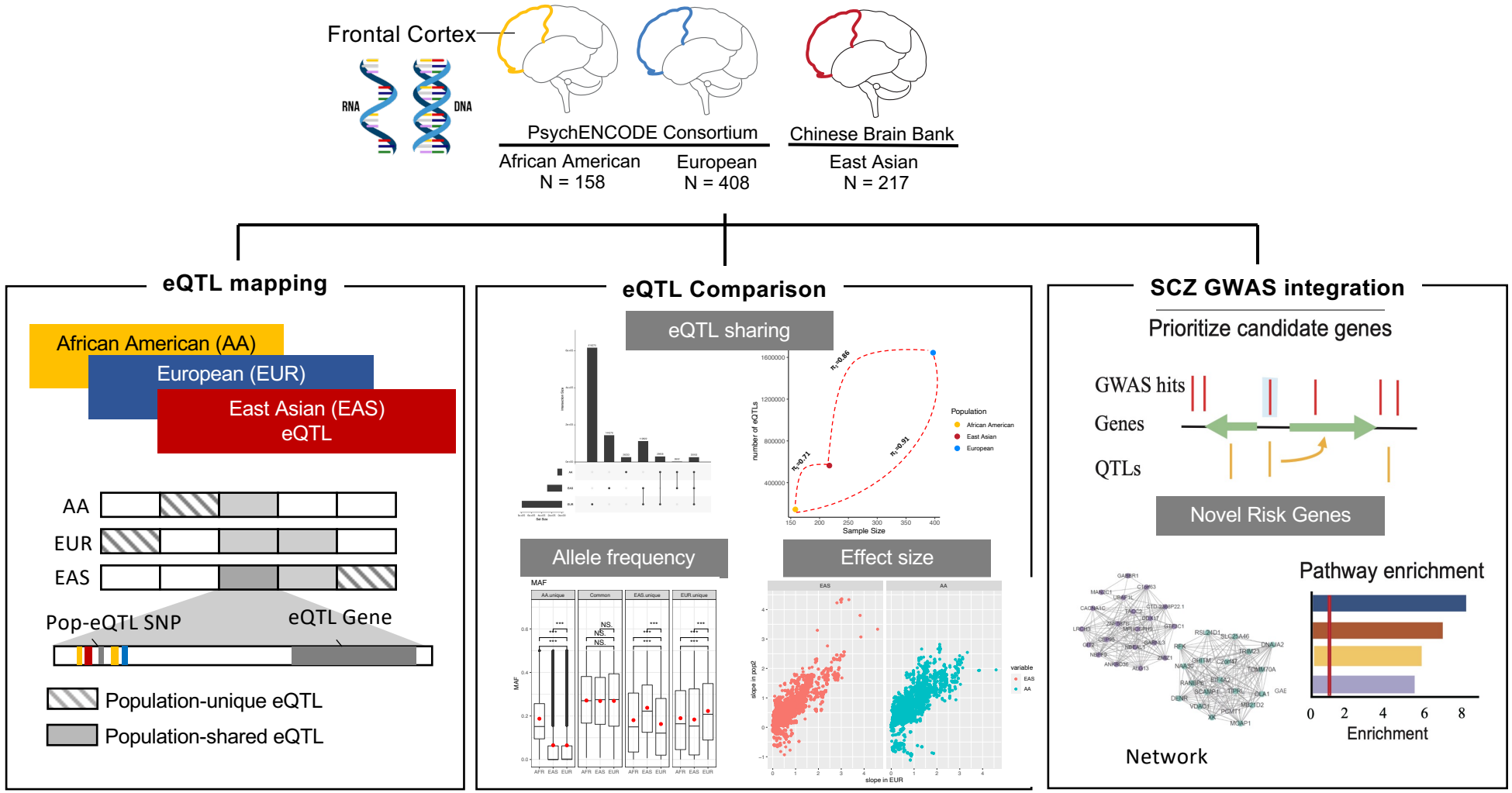

c
Scale
chr12:
20 bases
hg19
123,772,780
123,772,790
123,772,800
123,772,810
---> C T C A A A A A A A C A A A A C A A A A C A A A A C A A A A A A A A A A A C A A A A A A A C A T A T A A A T
Consensus Brain ATAC-seq: Merged 4 cell-lines & 3 brain regions
Glutamatergic neurons (mapped reads)
GLU
Glutamatergic neurons (peaks)
GABAergic neuron (mapped reads)
GABA
GABAergic neuron (peaks)
Oligodendrocytes (mapped reads)
OLIG
Oligodendrocytes (peaks)
Microglia and astrocytes (mapped reads)
MGAS
Microglia and astrocytes (peaks)
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
Publications: Sequences in Scientific Articles
Sequences
SNPs

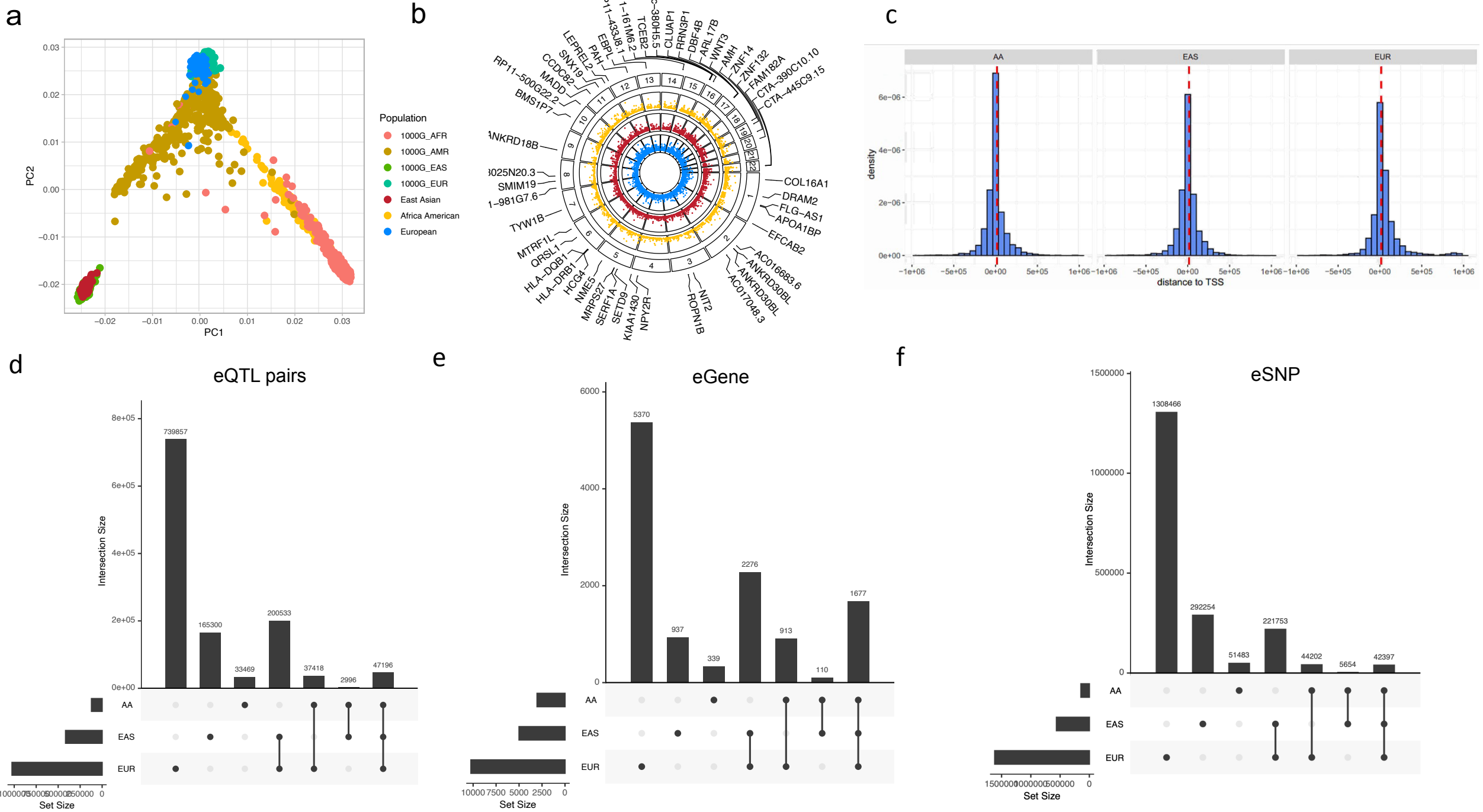

d
2.0
1.5
1.0
0.5
0.0
****
Vector
EAS_rs10883795-T
EAS_rs10883795-C

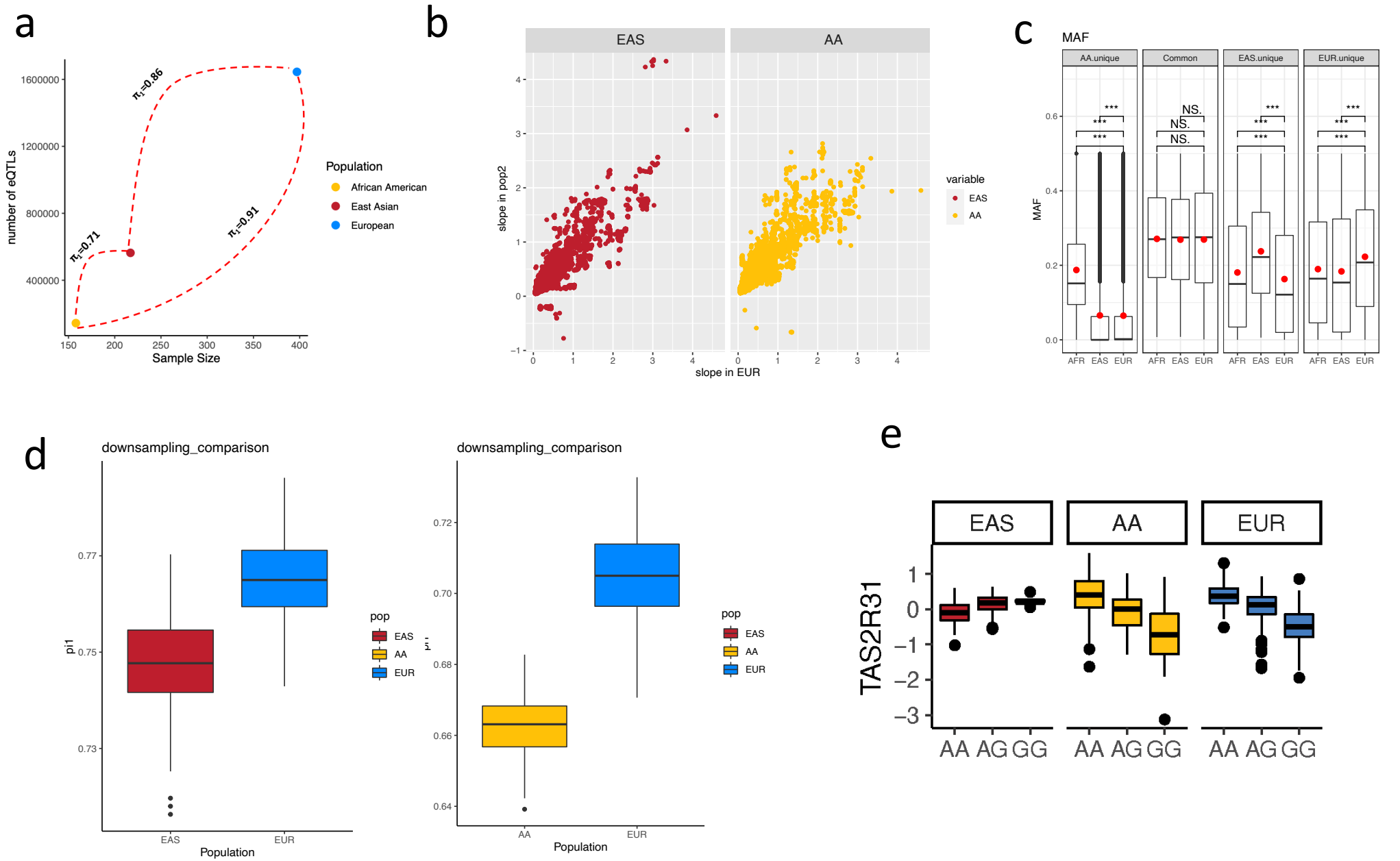

e
Gene expression
6.5
6.0
5.5
group
TT
TC
CC
TT (39)
TC (75)
CC (31)
10_104654577_rs10883795_T_C

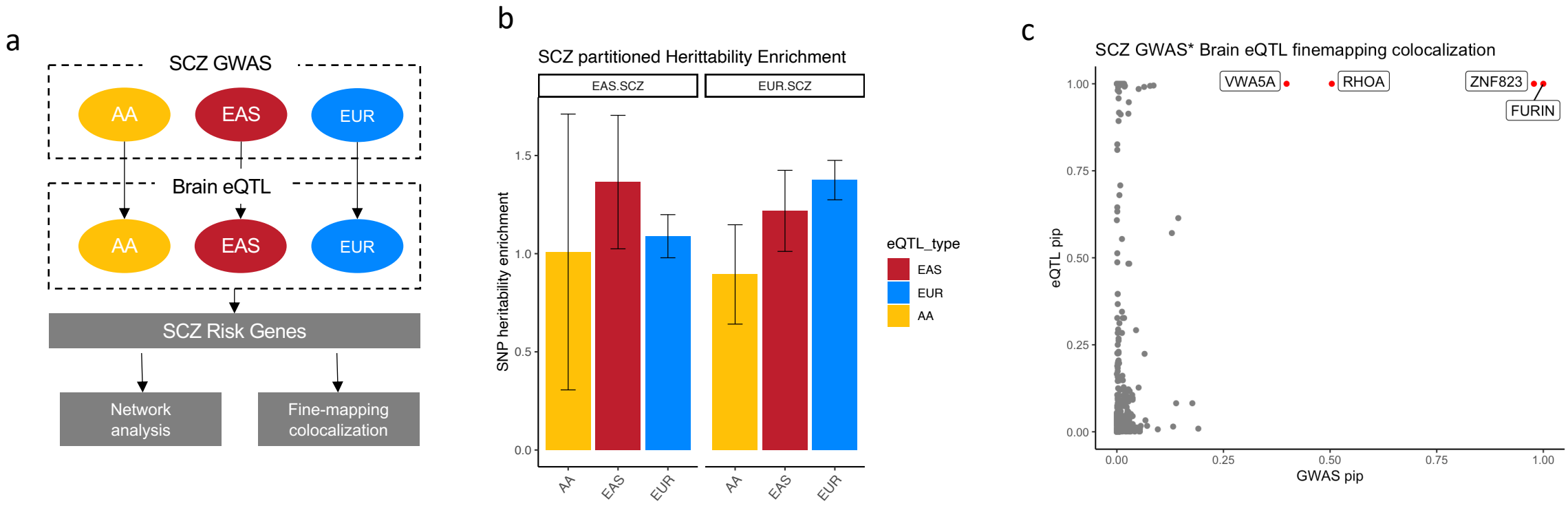

a
N = 100
N = 300
N = 500
N = 1000
N = 5000
N = 10000
Power
0.14
0.08
0.21
0.04
0.1
0.08
0.01
0.02
0.01
0.04
0.01
Minor Allele Frequency (MAF)
Beta
0.2
0.4
0.6


b
35419
Sample size
Minor Allele Frequency (MAF)
Beta
0.2
0.4
0.6


Frontal Cortex
RNA
DNA
PsychENCODE Consortium
African American
N = 158
European
N = 408
Chinese Brain Bank
East Asian
N = 217
eQTL mapping
African American (AA)
European (EUR)
East Asian (EAS)
eQTL
AA
EUR
EAS
Pop-eQTL SNP
eQTL Gene
Population-unique eQTL
Population-shared eQTL
eQTL Comparison
eQTL sharing
Allele frequency
Effect size
Sample Size
number of eQTLs
Population
African American
East Asian
European
MAF
slope in pop2
slope in EUR
variable
EAS
AA
SCZ GWAS integration
Prioritize candidate genes
GWAS hits
Genes
QTLs
Novel Risk Genes
Pathway enrichment
Enrichment
Network


a
PC1
PC2
Population
1000G_AFR
1000G_AMR
1000G_EAS
1000G_EUR
East Asian
Africa American
European
b
c
AA
EAS
EUR
density
distance to TSS
d
eQTL pairs
Intersection Size
Set Size
739857
165300
33469
200533
37418
2996
47196
e
eGene
5370
937
339
2276
913
110
1677
f
eSNP
1308466
292254
51483
221753
44202
5654
42397


a
number of eQTLs
Sample Size
π1=0.86
π1=0.71
π1=0.91
Population
African American
East Asian
European
b
EAS
AA
slope in pop2
slope in EUR
variable
EAS
AA
c
MAF
AA.unique
Common
EAS.unique
EUR.unique
NS.
AFR EAS EUR
d
downsampling_comparison
pi1
Population
pop
EAS
AA
EUR
e
EAS
AA
EUR
TAS2R31
AA AG GG


a
SCZ GWAS
AA
EAS
EUR
Brain eQTL
AA
EAS
EUR
SCZ Risk Genes
Network analysis
Fine-mapping colocalization
b
SCZ partitioned Herittability Enrichment
EAS.SCZ
EUR.SCZ
SNP heritability enrichment
0.0
0.5
1.0
1.5
AA
EAS
EUR
AA
EAS
EUR
eQTL_type
EAS
EUR
AA
c
SCZ GWAS* Brain eQTL finemapping colocalization
VWA5A
RHOA
ZNF823
FURIN
eQTL pip
0.00
0.25
0.50
0.75
1.00
GWAS pip
0.00
0.25
0.50
0.75
1.00