Supplementary Information

Table of Contents

Supplementary Note 1: Phenotype mapping **3**

Supplementary Note 2. Stratified analyses **5**

Supplementary Note 3: Sensitivity analyses of using the EUR LD reference for the cross-ancestry analyses **6**

Supplementary Note 4: Polygenic prediction of EA in UK Biobank East Asian samples and in leave-one-cohort-out analyses **8**

Supplementary Note 5: Sensitivity analyses of EA-associated brain cell types based on ancestry-stratified GWAS **10**

Supplementary Note 6: Author contributions for each participating EAS cohort **11**

References **12**

Supplementary Figures ***14***

Supplementary Fig. S1 Quality control of East Asian GWAS of EA. **14**

Supplementary Fig. S2 Comparison of allele frequencies and effect sizes of genetic associations identified in EAS and EUR **15**

Supplementary Fig. S3 Polygenic prediction of EA in UKB East Asian samples. **16**

Supplementary Fig. S4 EAS-specific genetic associations in EUR-reported loci. **17**

Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods. **18**

Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued). **19**

Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued). **20**

Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued). **21**

Supplementary Fig. S6 Triangulated causal effects of human complex traits on EA in EAS across multiple MR methods. **22**

Supplementary Fig. S7 Stratified LD score regression (S-LDSC) analysis. **23**

Supplementary Fig. S8 Cell-type association across ten human brain regions based on cross-ancestry EA GWAS meta-analysis. **24**

Supplementary Fig. S9 Cell-type association in the human prefrontal cortex based on ancestry-stratified EA GWAS. **25**

Supplementary Fig. S10 Cell-type association across brain regions in the mouse brain based on ancestry-stratified EA GWAS. 26

Supplementary Fig. S11 Cell-type association across ten human brain regions based on ancestry-stratified EA GWAS. 28

Supplementary Fig. S12 Cohort-level characteristics of all EAS cohorts. 29

Supplementary Fig. S13 Consistency of the inverse-variance-weighted meta-analysis results between OSCA and METAL for genome-wide significant variants. 30

Supplementary Fig. S14 Locus plots of the *DCAKD* gene. 31

Supplementary Notes

Supplementary Note 1: Phenotype mapping

Following established protocol[1,2] for large-scale genome-wide association studies (GWAS) of educational attainment (EA), we requested that each cohort perform phenotype mapping based on the International Standard Classification of Education (ISCED) 2011 levels and official country-specific years-of-schooling mapping files at the following URL: https://www.uis.unesco.org/en/methods-and-tools/isced/mapping-and-diagrams. Each cohort converted educational attainment measures to ISCED levels and then mapped them to the corresponding years of schooling (*EduYears*) based on its cohort-specific educational context, within the range provided by the official mappings. The cohort-specific mappings and phenotype distributions are listed in Supplementary Table S35.

Compared with the previous study, our ISCED to *EduYears* mappings differ in two aspects. First, we used the 2011 version of the ISCED level definition, while the previous studies used the 1997 version. Compared to ISCED 1997, which had seven levels of education, ISCED 2011 has nine levels, with more detailed tertiary-education levels (the ISCED 5-7 levels in the 2011 version correspond to the ISCED 5 level in 1997). We show the difference and correspondence between these two ISCED versions in the following table (also see Supplementary Table S49):

Correspondence between ISCED 2011 and ISCED 1997 levels

ISCED 1997 level	1997 Definition	ISCED 2011 level	2011 Definition
0	Pre-primary education	0	Early childhood education
1	Primary education or first stage of basic education	1	Primary education
2	Lower secondary or second stage of basic education	2	Lower secondary education
3	(Upper) secondary education	3	Upper secondary education
4	Post-secondary non-tertiary education	4	Post-secondary non-tertiary education
5	First stage of tertiary education (not leading directly to an advanced research qualification)	5	Short-cycle tertiary education
		6	Bachelor's or equivalent level
		7	Master's or equivalent level

6	Second stage of tertiary education (leading to an advanced research qualification, e.g. a PhD)	8	Doctoral or equivalent level

By using the ISCED 2011 levels, our data capture the difference between samples who completed short-cycle tertiary education, undergraduate education, and a Master's degree. Second, we leveraged cohort-specific years-of-schooling mappings, while the previous EAS study from Chen et al. used US-based years-of-schooling mappings. By restricting only the range of each ISCED level, we gave each cohort the flexibility to determine which level-to-year mapping best suited their cohort-specific educational context. We summarized the common ISCED mappings across cohorts and classified them by country in the following table (also see Supplementary Table S50):

Summarized ISCED mappings for new cohorts from China and Japan.

ISCED 2011 level	**Definition**	**Years of schooling in China**	**Years of schooling in Japan**
0	Early childhood education	0	0
1	Primary education	6	6
2	Lower secondary education	9	9
3	Upper secondary education	12	12
4	Post-secondary non-tertiary education	-	14
5	Short-cycle tertiary education	15	14
6	Bachelor's or equivalent level	16	16
7	Master's or equivalent level	19	Actual recorded years of schooling
8	Doctoral or equivalent level	22 or 23	Actual recorded years of schooling

All Chinese cohorts used the same level-to-year mapping for ISCED 0 to ISCED 7. None of the samples were assigned to the ISCED 4 level in China. For samples who completed ISCED 8, the 23mofang cohort used 23 years of schooling, whereas the other cohorts used 22 years. The Japanese cohorts used the

same level-to-year mapping for ISCED 0 to ISCED 6, while using the actual recorded years of schooling for individuals who completed graduate school (years of schooling > 16). Individuals in the combined ISCED 1/2 category, in which primary and junior high school levels could not be distinguished, were uniformly assigned 9 years of schooling, reflecting the compulsory education period in Japan. Overall, our phenotype mapping approach captured fine-scale phenotypic variance and reflected cohort-specific environmental contexts, thereby reducing cross-cohort harmonization noise. We also provided detailed mappings for each cohort in Supplementary Table S35.

Supplementary Note 2. Stratified analyses

We grouped our newly collected cohorts from four different aspects for meta-analysis: 1. Age-stratified analysis. We requested each cohort to separate their samples by their enrollment age (birth year plus 6) into: 1) age group 1: before 1945, 2) age group 2: 1946-1965, 3) age group 3: 1966-1985, and 4) age group 4: after 1986. We then performed a meta-analysis across all cohorts using the age-stratified cohort-level GWAS. 2. Sex-stratified analysis. We requested that each cohort separate their samples into male and female samples. We then performed a meta-analysis across all cohorts using the sex-stratified cohort-level GWAS. 3. Country-stratified analysis: We performed meta-analysis of the cohort-level GWAS in Chinese (CKB, BGI-lcWGS, WGEG, WGEA, 23mofang, WeGut, and WGEC) and compared with the JPN-meta GWAS (the meta-analysis of six Japanese cohorts). 4. Sampling-characteristic-stratified analysis. We performed meta-analysis of the cohort-level GWAS across two groups: 1) Prospective studies, including CKB and WeGut. 2) Genetic testing platforms, including WeGene and 23mofang.

We estimated the SNP-based heritability across the GWAS meta-analysis stratified by these groups and the genetic correlations between group pairs (Supplementary Tables S44 and S45). We found specific SNP-based heritability and genetic correlations of less than one across age-, sex-, and country-stratified groups. We assume that potential uncaptured gene-by-environment effects may account for these heterogeneities. We also estimated genetic correlations between these groups and the EUR meta-analysis (Supplementary Table S46). Overall, these cross-ancestry genetic correlation estimates are not significantly different from one another given the large estimation uncertainty. All stratified groups showed genetic correlations > 0.8 with Chen et al.'s GWAS, which explains the 0.99 correlation between the meta-analysis of new EAS cohorts and Chen et al.'s GWAS. The meta-analysis of new EAS datasets from genetic testing platforms (WGEA, WGEC, WGEG, and 23mofang) showed the highest genetic correlation ($R_{\mathrm{g}} = 0.902$, $\mathrm{s.e.} = 0.137$) with the EUR GWAS from Lee et al. The genetic correlations

between the other large cohorts or stratified groups and the EUR meta-analysis ranged from 0.611 (meta-analysis across prospective cohorts including CKB and WeGut) to 0.806 (meta-analysis of male-only GWAS across all cohorts). Consistent trends were observed in the genetic correlations between these groups and IQ (R_g ranged from 0.459 for JPN-meta to 0.874 for meta-analysis across genetic testing platforms).

Supplementary Note 3: Sensitivity analyses of using the EUR LD reference for the cross-ancestry analyses

For cross-ancestry analyses that require an LD reference, such as LD-based clumping and stratified-LDSC, we leveraged the EUR samples from 1KGP3 as the LD reference. To ensure this did not introduce bias due to LD mismatch between the reference panel and the GWAS, we evaluated the LDSC intercept and attenuation ratio for the meta-analysis summary statistics using the EUR reference. The LDSC intercept and attenuation ratio not only estimate the genomic inflation caused by factors other than polygenic heritability (such as population stratification) but also reflect mismatches between the reference population from which LD scores were estimated and the discovery population in which the GWAS was performed[3]. Using a EUR LD reference for a cross-ancestry GWAS meta-analysis may bias the LDSC intercept due to LD misspecification. Such mismatch biases the intercept through either noise or heterogeneity in average LD scores across populations: noise always biases the intercept upward, whereas heterogeneity in LD scores can bias it upward or downward, depending on whether the average LD score in the reference population is lower or higher compared to the GWAS population. An LDSC intercept close to 1 and an attenuation ratio below 20%, as suggested by the LDSC authors, indicate that the estimation bias is within acceptable limits.

The LDSC intercept estimated from our cross-ancestry meta-analysis based on EUR LD was close to 1 (LDSC intercept = 1.0721, s.e. = 0.0147), and the attenuation ratio was 4.0%, below the 20% threshold recommended by the LDSC authors. Using the same EUR LD, this result is similar to the LDSC intercept and attenuation ratio (LDSC intercept = 1.0286, s.e. = 0.0138; ratio = 1.7%) estimated from the EUR-only GWAS, and smaller than the results (LDSC intercept = 1.1235, s.e. = 0.0154; ratio = 20.6%) estimated from the EAS-only GWAS. These results suggest that a mismatch dose exist when applying the EUR LD to our cross-ancestry meta-analysis, but that it is smaller than the mismatch between the EUR LD and the EAS-only GWAS. We also measured the LDSC intercept and attenuation ratio of our cross-ancestry meta-analysis data using a mixed EAS and EUR LD reference, calculated from 503 EUR samples and 294 EAS samples from 1KGP3, matching the sample ratio in our cross-ancestry meta-analysis (~37% EAS).

Specifically, we retained all available EUR samples and randomly selected 294 EAS samples from the same 1KGP3 reference panel. The genotypes of these 797 individuals were merged using PLINK2, retaining only variants that passed standard quality control thresholds (variant-level missingness < 5%, individual-level missingness < 5%, MAF > 1%, and Hardy-Weinberg equilibrium $P < 1 \times 10^{-6}$). The LDSC intercept and ratio were biased upward (LDSC intercept = 1.1317, s.e. = 0.0164; attenuation ratio = 7.2%), indicating a larger mismatch than using the EUR LD reference (Supplementary Table S48). To distinguish the difference in LDSC estimates due to LD structure versus variant inclusion, we performed further comparison between the EUR LD reference and the mixed-ancestry LD reference using the same set of regression variants. We restricted both LD references to the exact same set of overlapping variants used in the previous LDSC regressions with the original EUR and mixed EAS-EUR LD references, and generated new reference LD score files. The corresponding LDSC analyses were then performed using this identical set of regression variants. The differences between the estimates were evaluated using a paired block-jackknife approach across 200 matched genomic blocks, with statistical significance assessed by two-sided Wald tests. Using the same summary statistics and regression variants, the mixed-ancestry reference yielded a significantly higher LDSC intercept than the EUR-only reference (1.1393 vs. 1.1000, difference s.e. = 0.0069; $P = 1.32 \times 10^{-8}$) and a correspondingly higher attenuation ratio (0.0764 vs. 0.0548; the test is identical to that of the intercept because the ratio is a fixed transformation of the intercept given the same summary statistics) (Supplementary Table S48). Notably, restricting the EUR LD reference to this smaller subset of variants also increased the LDSC intercept and attenuation ratio compared to the original, more complete EUR LD reference (intercept = 1.0721, s.e. = 0.0147; ratio = 0.0397). These findings indicate that LDSC estimates are inherently sensitive to the LD structure of the reference panel, while other confounders, such as variant inclusion, can also contribute to the overall estimation bias.

Moreover, we performed a sensitivity analysis using the mixed EAS and EUR reference for LD-based clumping. We found that 741 independent lead variants were identified with the mixed LD reference, of which 612 overlapped the 809 independent variants identified based on the EUR LD. The mixed LD reference identified 129 additional unique variants, but all of these variants had already been captured within the clumps of the 809 independent variants based on the EUR LD, suggesting that the EUR LD captures longer-range LD patterns than the mixed LD. We then studied the 197 unique variants identified with the EUR LD. We found that 46 of them were included in the 741 clumps of the mixed LD, while the remaining 151 variants were missing because they were present only in the EUR LD reference

and absent from the mixed EAS-EUR panel. This is because, during the quality control process for generating the mixed LD reference, only SNPs common across ancestries were retained. We further checked the clumping log files and found that among the 33,245 genome-wide significant variants in our cross-ancestry meta-analysis data, 15.3% were missing from the mixed LD reference, whereas only 0.5% were missing from the EUR LD reference.

While it is difficult to prove which reference panel has the smaller underlying population mismatch relative to the GWAS summary statistics, the practical advantages of the EUR LD reference are clear. First, as demonstrated by our sensitivity analyses above, although estimation bias can be introduced by multiple factors, using a more complete set of variants minimizes overall estimation bias. Second, while both reference panels perform acceptably in LDSC regression (attenuation ratios < 20%), missing variants in the mixed reference severely compromise LD-based clumping by reducing the number of identified lead variants. Third, the EUR LD reference represents a true biological population rather than an artificially admixed cohort. A statistically mixed reference panel may create spurious LD correlations that do not exist in nature and introduce unknown noise into the analyses. Therefore, we utilized the EUR LD reference for our primary cross-ancestry analyses to maximize variant retention while keeping estimation bias within acceptable limits. To ensure the robustness of our findings, we employed the mixed EAS and EUR LD reference as a sensitivity check.

Supplementary Note 4: Polygenic prediction of EA in UK Biobank East Asian samples and in leave-one-cohort-out analyses

We evaluated the predictive performance of our East Asian (EAS) EA GWAS in a holdout cohort comprising UK Biobank (UKB[4]) EAS samples and compared it to the European (EUR) EA GWAS from Lee et al. (EA3[1]). To optimize cross-ancestry prediction, we leveraged a weighted PRS model (Methods). Consistent with the higher relevance of ancestry-matched discovery data, the EAS-derived PRS exhibited higher norminal predictive power than the EUR-derived PRS in the EAS holdout cohort. Furthermore, the performance of the EAS-derived PRS was comparable to that of the cross-ancestry PRS. This suggests that the predictive power of our cross-ancestry PRS framework in EAS individuals is primarily driven by the EAS GWAS data (Supplementary Table S8).

We also benchmarked our summary statistics with those of a previous study (Chen et al.[5]) using the full UKB EAS sample (N=1,615) and the established framework. We utilized PRS-CS[6] to construct the ancestry-

stratified EA PRS and PRS-CSx[7] to build cross-ancestry PRS from the meta-analysis of EAS and EUR GWAS. This approach bypasses the need for an independent tuning set by performing an inverse-variance-weighted meta-analysis of the posterior effect sizes across ancestries. Under this framework, our EAS-derived PRS achieved a nominal incremental R^2 of 1.72%, outperforming the EUR-derived PRS (0.62%), the cross-ancestry PRS (1.16%) derived from the meta-analysis, and the EAS-derived PRS from Chen et al. (1.18% in 1,747 UKB EAS samples). The attenuation of cross-ancestry PRS predictive power in this framework is likely due to the large heterogeneity between the prediction accuracies of the EAS-only and EUR-only PRS in this dataset. The EUR-only PRS, despite its low prediction accuracy in the test cohort, was assigned a large weight due to the larger sample size of the EUR GWAS discovery dataset. Consequently, the posterior effect sizes of EAS-specific tagging variants may be underestimated. Although the limited sample size of the UKB EAS holdout cohort warrants cautious interpretation, the consistent nominal trends across both PRS frameworks underscore the EAS-specific polygenic architecture captured by our GWAS.

Moreover, we performed a sensitivity analysis using a leave-one-cohort-out (LOCO) strategy based on the WeGut ($n = 4{,}254$) and 23mofang ($n = 9{,}155$) cohorts. We used the same parameters to construct ancestry-specific EA PRS. For cross-ancestry PRS, we also applied the weighted PRS model and the meta model built in PRS-CSx. For the weighted PRS model, we performed only one iteration with a single random sampling of the validation and test samples, and used the 1,000 bootstrap replicates within the dataset to estimate the confidence intervals. The EAS-only PRS derived from the LOCO meta-analyses demonstrated incremental R^2 values of 0.96% in WeGut and 1.12% in 23mofang. The prediction accuracies of the EUR-only PRS in these two cohorts were comparable to the EAS-only PRS. The attenuated prediction accuracy of the EAS-only EA PRS in these two cohorts could be due to reduced statistical power in the meta-analysis, cohort-specific SNP-based heritability of EA, and estimation uncertainty. The cross-ancestry PRS from both the weighted and meta models in the LOCO PRS analyses increased relative to the ancestry-specific PRS (Supplementary Table S37).

As a sensitivity analysis, we also repeated the PRS analysis using the UKB LD reference provided by the PRS-CSx authors, which have much larger sample sizes than the 1KGP3 reference ($n_{\mathrm{EAS}} = 2{,}181$; $n_{\mathrm{EUR}} = 375{,}120$), but which may cause over-fitting since our test cohort is from the same population. For the weighted PRS model, we performed only one iteration with a single random sampling of the validation and test samples, and used the 1,000 bootstrap replicates within the dataset to estimate the confidence

intervals. The results were nearly identical to those obtained when using 1KGP3 as the LD reference (Supplementary Table S37).

Supplementary Note 5: Sensitivity analyses of EA-associated brain cell types based on ancestry-stratified GWAS

To evaluate the concordance of EA-associated cell types between EAS and EUR, we conducted a series of sensitivity analyses using ancestry-stratified GWAS summary statistics and LD references (1KGP3). Our validation framework integrated three transcriptomics datasets using three cell mapping methods: (1) single-cell RNA sequencing (scRNA-seq) data from the human prefrontal cortex (PFC)[8] using scDRS[9] and stratified-LDSC (S-LDSC[10]); (2) spatial transcriptomics (ST) data from the mouse brain[11] using gsMap[12]; and (3) scRNA-seq data across 10 human brain regions[13] using scDRS.

Cell-type association analyses in the human PFC using both our EAS GWAS and EA3 identified excitatory and inhibitory neurons as significant EA-associated cell types ($\mathrm{FDR} < 0.05$; Supplementary Fig. S9, Supplementary Table S28). Although oligodendrocyte precursor cells (OPCs) were identified as an associated cell type exclusively in EUR and the overall number of associated cells in EUR is much larger than in EAS, we observed high concordance in cell-type-level association z-scores ($r = 0.908$). These results suggest that although discrepancies exist between ancestries, they likely reflect the superior statistical power in the larger EUR discovery GWAS, while the overall cell-type association patterns remained highly consistent. At the single-cell level, we observed moderate but stable correlations in association Z-scores ($r = 0.503$) and normalized trait scores (weighted sum of MAGMA genes accounting for control gene sets, referred to as 'disease score' in scDRS; $r = 0.500$). These results are consistent with expectations, given that a substantial proportion of cells were not significantly associated. Furthermore, S-LDSC analysis confirmed that EA heritability is significantly enriched in genes preferentially expressed in excitatory and inhibitory neurons in both ancestries (Supplementary Table S29), with a correlation in cell-type-level enrichment Z-scores of 0.788. These results collectively demonstrate a conserved single-cell association pattern with EA in the human PFC across ancestries.

Integrating ancestry-stratified GWAS data with mouse brain ST data via gsMap, we identified excitatory neurons as the most significant EA-associated cell type in both ancestries (Supplementary Table S30). Spatial association patterns were highly conserved as well, with a strong correlation in cell-type association significance (Cauchy combination *P*-values on a $-log_{10}$ scale) across brain regions ($r = 0.865$;

Supplementary Fig. S10; Supplementary Table S31) and single-cell-level association significance ($r = 0.756$). These results validate that the spatial distribution of EA-associated cells in mouse ST data is conserved across ancestries.

Finally, we applied scDRS to the scRNA-seq data of 10 human brain regions. Due to the high multiple-testing burden (183 cell-type-region pairs) and lower statistical power, we did not identify significant cell-type associations across brain regions based on EAS GWAS (Supplementary Fig. S11, Supplementary Table S32). However, we observed consistent cross-ancestry correlations in cell-type association Z-scores across brain regions ($r = 0.778$), single-cell-level association Z-scores ($r = 0.482$), and cellular level normalized trait scores ($r = 0.480$). These results reinforce a conserved, region-dependent cellular mechanism for EA between EAS and EUR.

Supplementary Note 6: Author contributions for each participating EAS cohort

Contributors to JPN-meta

Statistical analysis: Keiko Hikino, Takahisa Kawaguchi, Xiaoxi Liu, Taisei Mushiroda

Study supervision: Kouichi Ozaki, Yasuharu Tabara, Chikashi Terao

Contributors to CKB

Statistical analysis: Yuxuan Zhao, Qi Liu

Genotyping QC: Robin G. Walters

Study supervision: Zhengming Chen, Jun Lv, Liming Li, Dianjianyi Sun

Contributors to BGI-lcWGS

Statistical analysis: Linxuan Li, Hebin Zhang, Han Xiao

Study supervision: Aifen Zhou, Huanhuan Zhu, Xin Jin

Contributors to WGEA, WGEG, and WGEC

Statistical analysis: Lizhong Wang, Senwei Tang, Teng Li

Study supervision: Gang Chen

Contributors to 23mofang

Statistical analysis: Hao Mi

Study supervision: Kun Zhou

Contributors to WeGut

Statistical analysis: Fengzhe Xu

Study supervision: Xiong-Fei Pan, An Pan, Yu-ming Chen, Yimin Zhu, Dan Zhou, Ju-Sheng Zheng

References

1 Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018). https://doi.org:10.1038/s41588-018-0147-3

2 Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54**, 437-449 (2022). https://doi.org:10.1038/s41588-022-01016-z

3 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015). https://doi.org:10.1038/ng.3211

4 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018). https://doi.org:10.1038/s41586-018-0579-z

5 Chen, T. T. *et al.* Shared genetic architectures of educational attainment in East Asian and European populations. *Nat Hum Behav* **8**, 562-575 (2024). https://doi.org:10.1038/s41562-023-01781-9

6 Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* **10**, 1776 (2019). https://doi.org:10.1038/s41467-019-09718-5

7 Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics* **54**, 573-580 (2022). https://doi.org:10.1038/s41588-022-01054-7

8 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019). https://doi.org:10.1038/s41586-019-1195-2

9 Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nature Genetics* **54**, 1572-1580 (2022). https://doi.org:10.1038/s41588-022-01167-z

10 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235 (2015). https://doi.org:10.1038/ng.3404

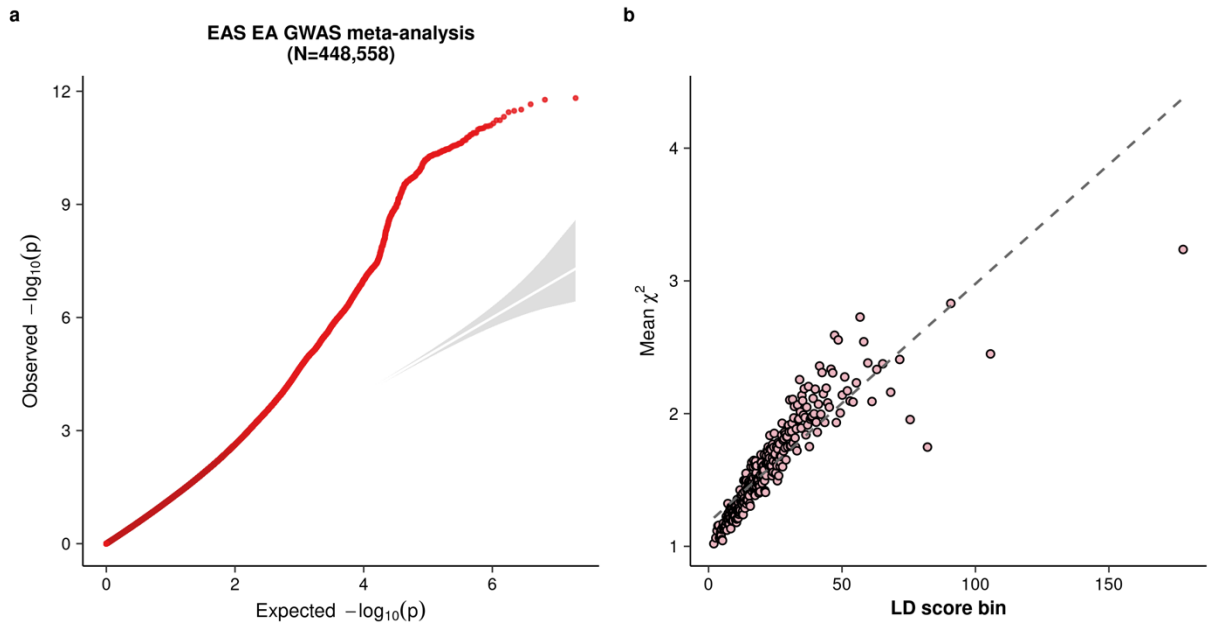
11 Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792.e1721 (2022). https://doi.org:10.1016/j.cell.2022.04.003

12 Song, L., Chen, W., Hou, J., Guo, M. & Yang, J. Spatially resolved mapping of cells associated with human complex traits. *Nature* (2025). https://doi.org:10.1038/s41586-025-08757-x

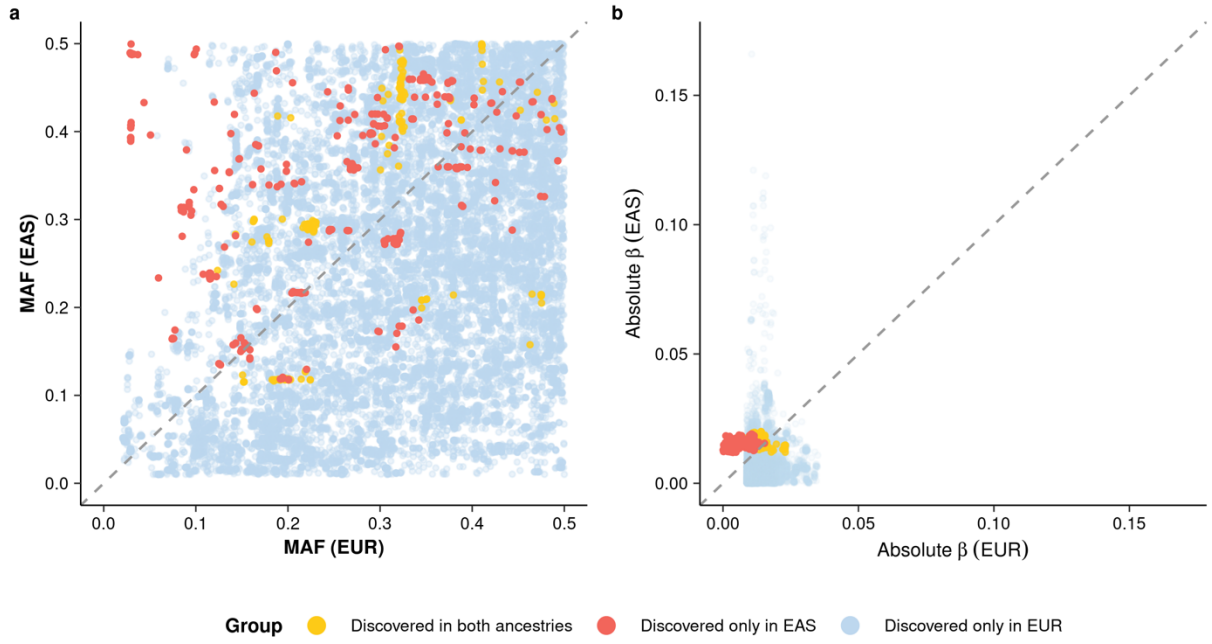
13 Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023). https://doi.org:10.1126/science.add7046

Supplementary Figures

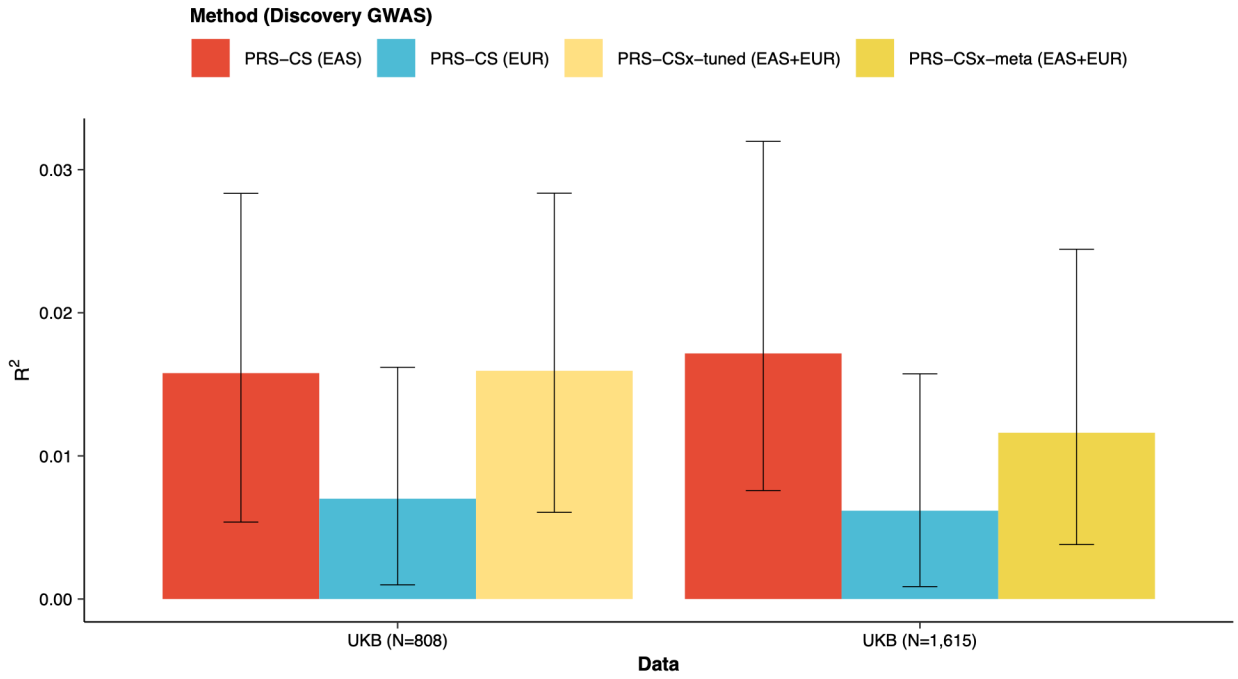
Supplementary Fig. S1 Quality control of East Asian GWAS of EA. a Quantile-Quantile (Q-Q) plot of the East Asian EA GWAS meta-analysis. The gray shaded areas represent the 95% confidence intervals under the null hypothesis. **b** LD score regression (LDSC) plot. Each point represents an LD score quantile containing 1000 SNPs.



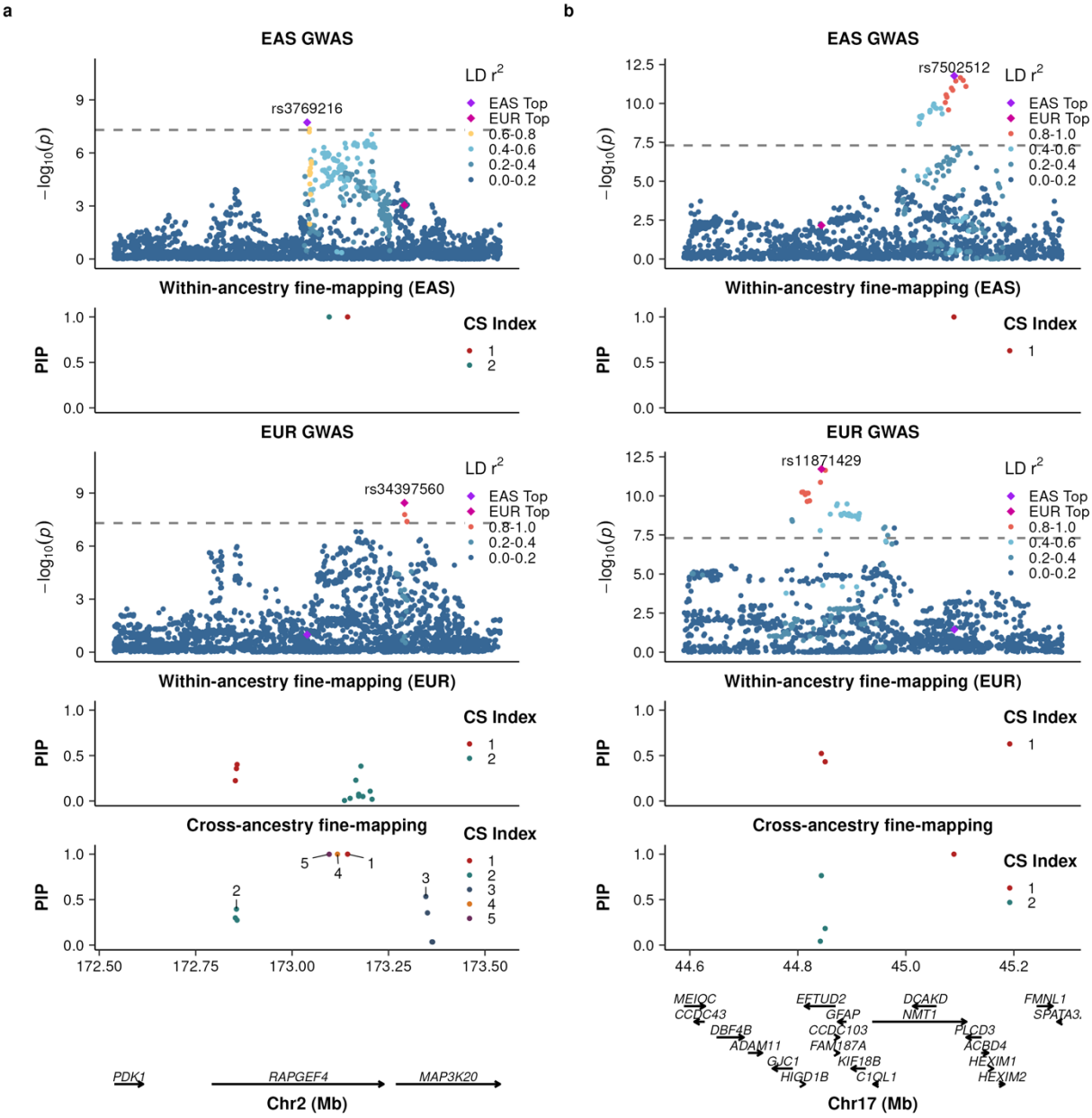
Supplementary Fig. S2 Comparison of allele frequencies and effect sizes of genetic associations identified in EAS and EUR. a Scatter plot of minor allele frequencies for genome-wide significant variants identified in ancestry-stratified analyses. **b** Absolute effect sizes of genome-wide significant variants identified in ancestry-stratified analyses. Colors denote the discovery ancestry.



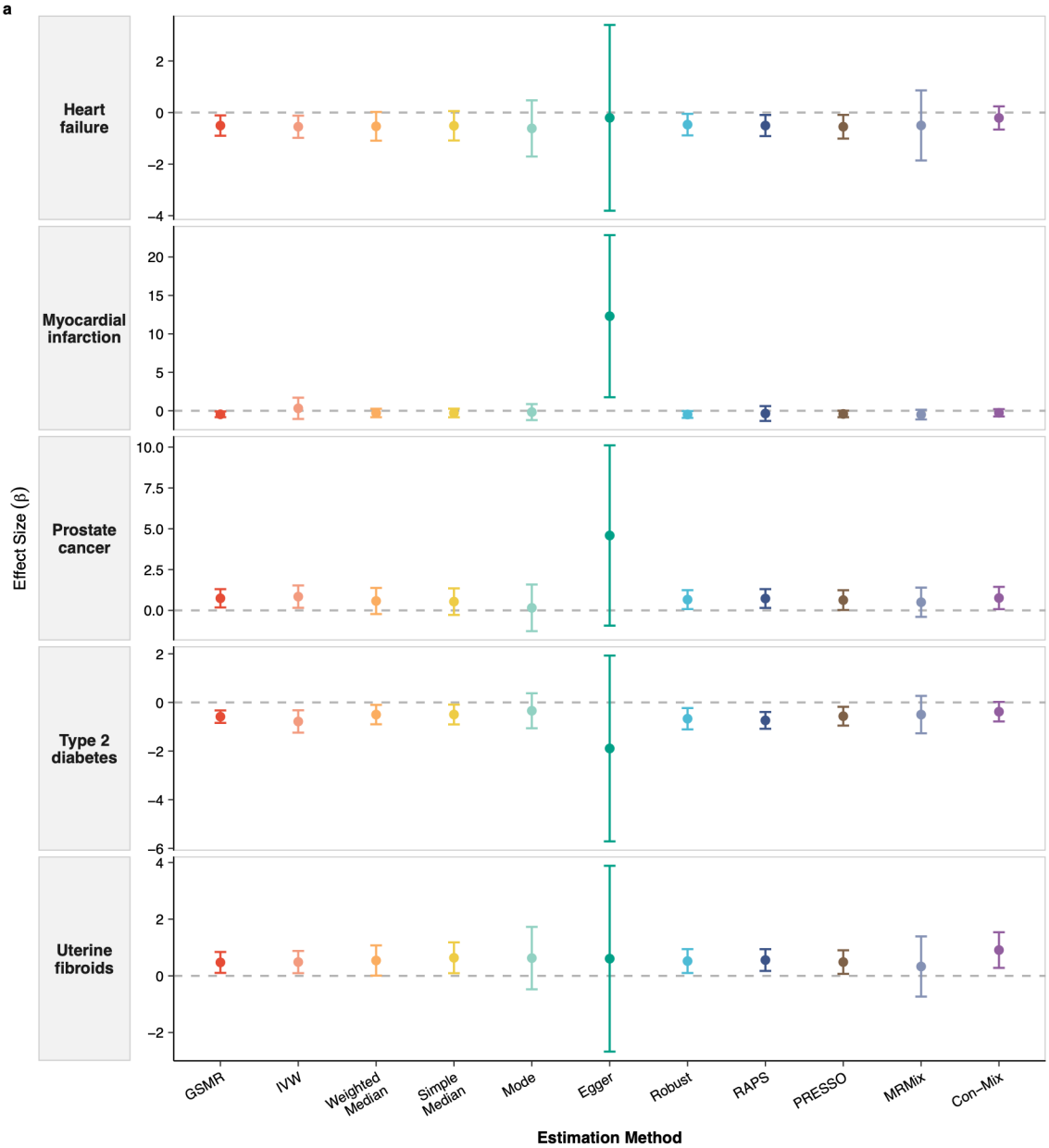
Supplementary Fig. S3 Polygenic prediction of EA in UKB East Asian samples. The EA PRS was constructed using a weight-tuning model with half of the samples ($\mathrm{N} = 808$) and the established model (Supplementary Note 4) with all EAS samples in UKB ($\mathrm{N} = 1{,}615$). The x-axis denotes the testing cohort, and the y-axis represents the incremental R^2 for the EA PRS. Colors represent different discovery GWAS datasets and methods. Error bars indicate 95% confidence intervals. PRS-CSx-tuned (EAS+EUR): the weighted sum of ancestry-specific PRS-CSx results used as the cross-ancestry EA PRS, with weights derived from validation sets (N=807). PRS-CSx-meta (EAS+EUR): cross-ancestry EA PRS generated using the --meta=TRUE command in PRS-CSx.



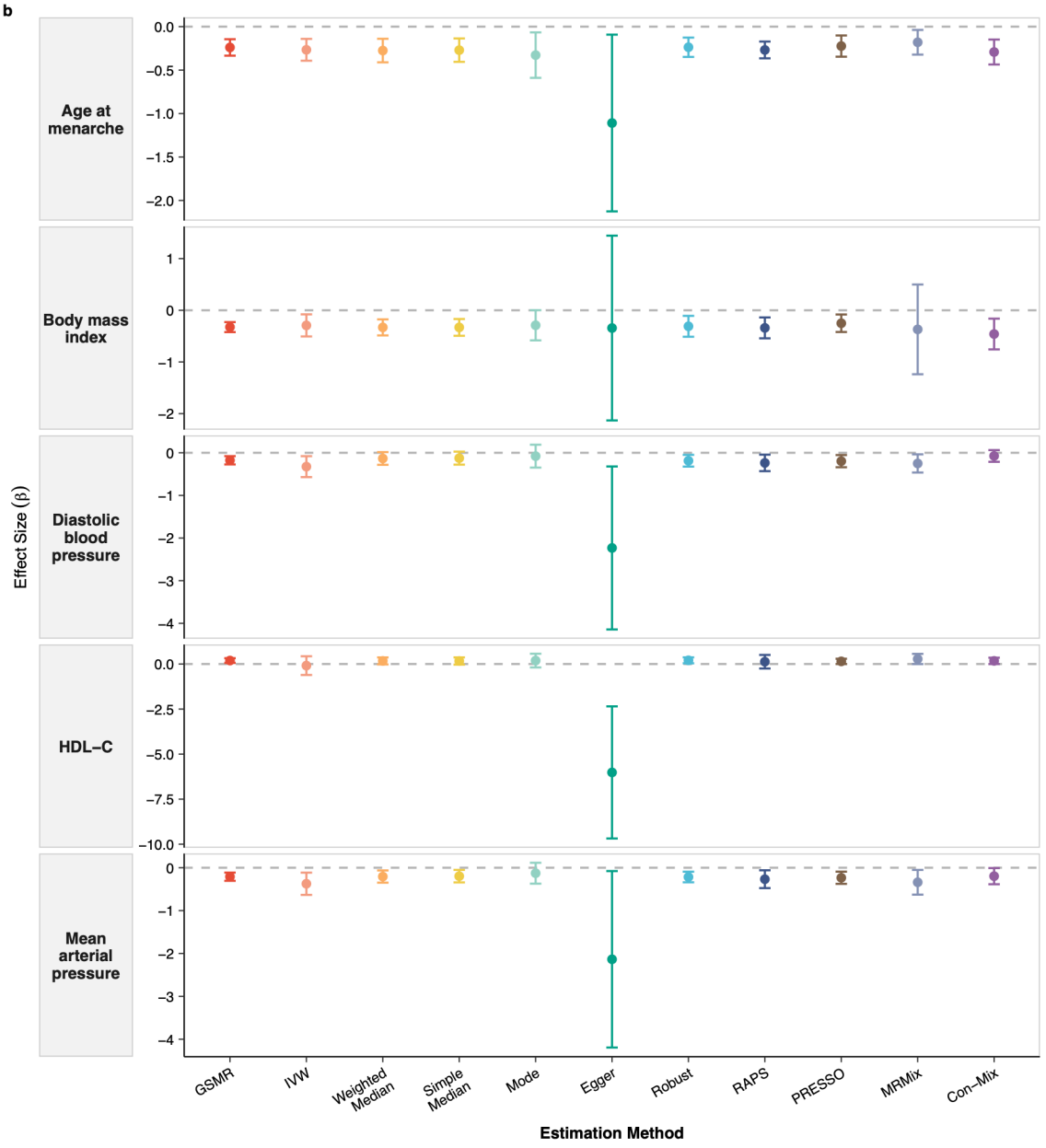
Supplementary Fig. S4 EAS-specific genetic associations in EUR-reported loci. a, b Locus plots of EAS-specific genetic associations at EUR-reported loci, together with cross-ancestry and within-ancestry fine-mapping results. The x-axis represents chromosomal position. The y-axis represents the association significance ($-log_{10}P$) from the EA GWAS, or the variant posterior inclusion probability (PIP) in each 95% credible set (CS) from the cross-ancestry or within-ancestry fine-mapping results (by SuSiEx). In the GWAS layers, the purple and pink diamonds represent the lead variant in EAS and EUR, respectively, while other dots are colored according to their LD r^2 with the lead variant. In the fine-mapping layers, different dot colors represent different CSs, and only variants included in a CS are plotted.



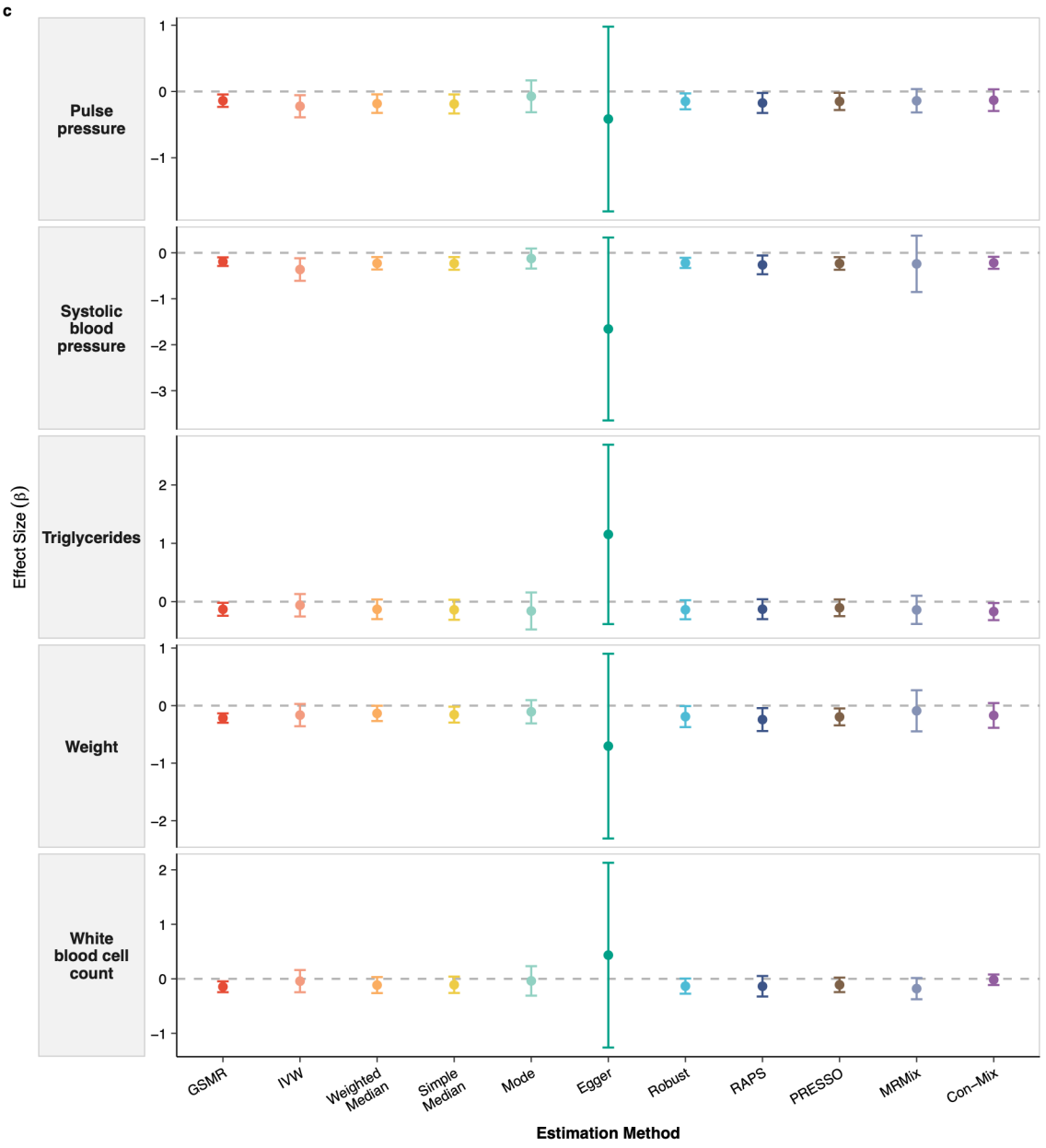
Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods. a-d, Causal estimates of EA on human complex traits. Traits with independent GWAS and significant GSMR causal effects after FDR correction ($\mathrm{FDR} < 0.05$) were shown and categorized into diseases (**a**), biomarkers (**b** and **c**), and drug intake (**d**). The row-wise panels show different traits. Circles are colored to denote different MR methods. The y-axis represents b_{xy} estimates from each method. The error bars denote 95% confidence intervals. The horizontal dashed line indicates $b_{xy} = 0$. Results with $s.e. > 10 \times b_{xy}$ were omitted from visualization due to uninformative confidence intervals.



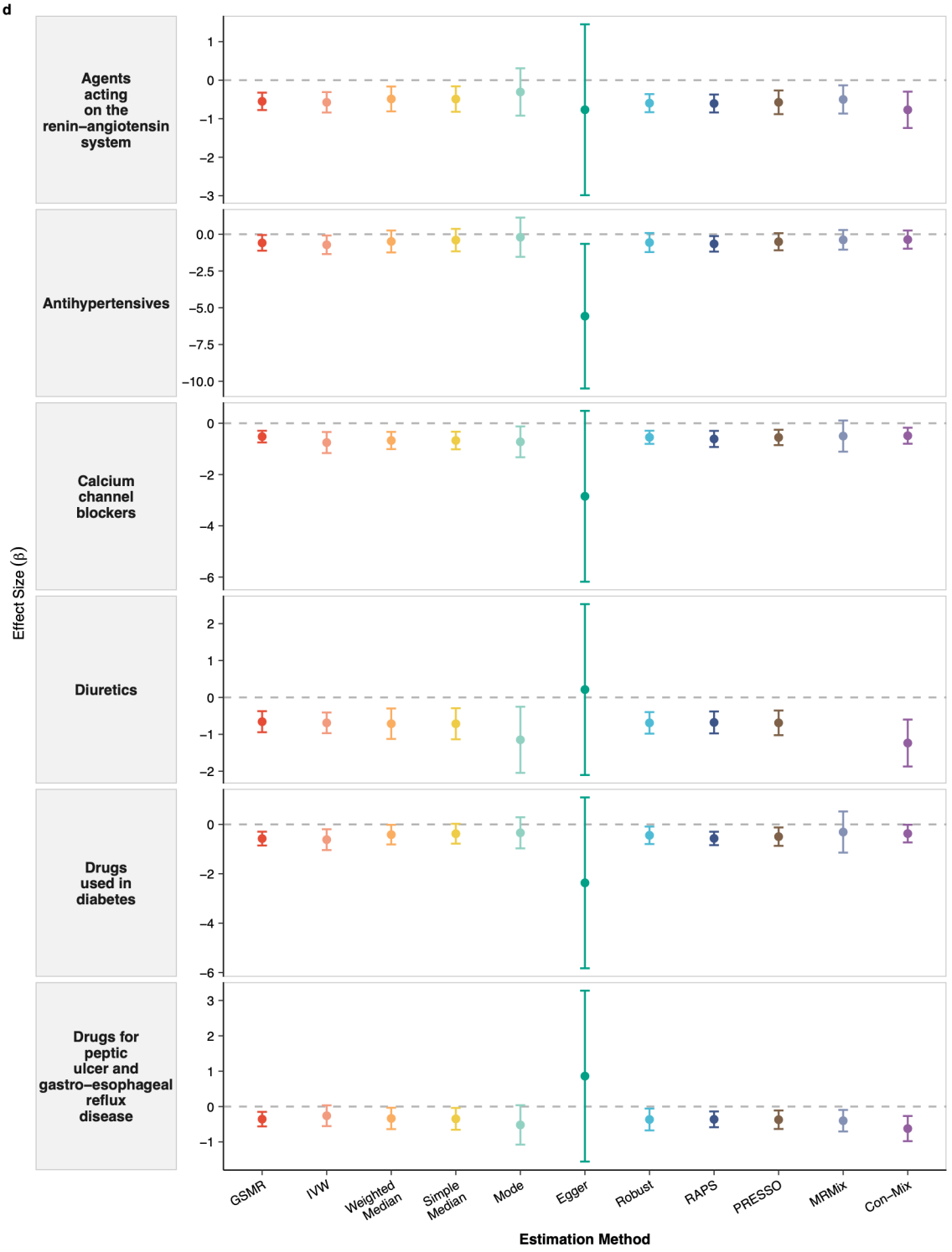
Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued).



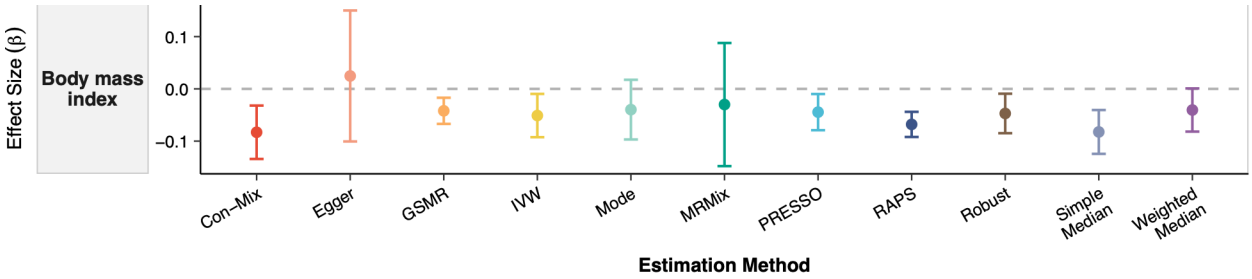
Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued).



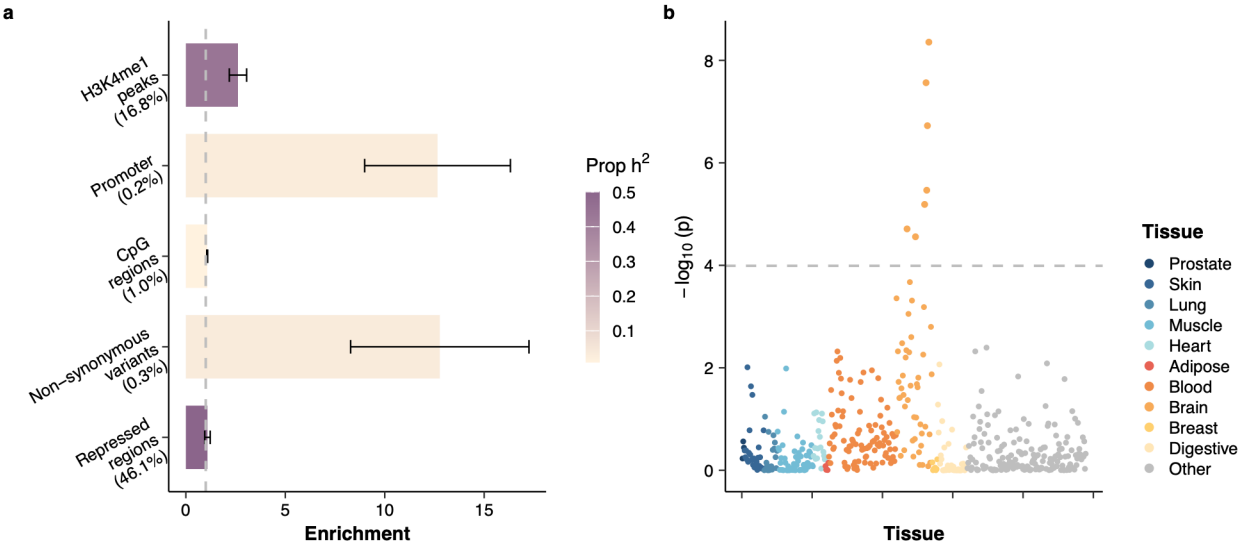
Supplementary Fig. S5 Triangulated causal effects of EA on human complex traits in EAS across multiple MR methods (Continued).



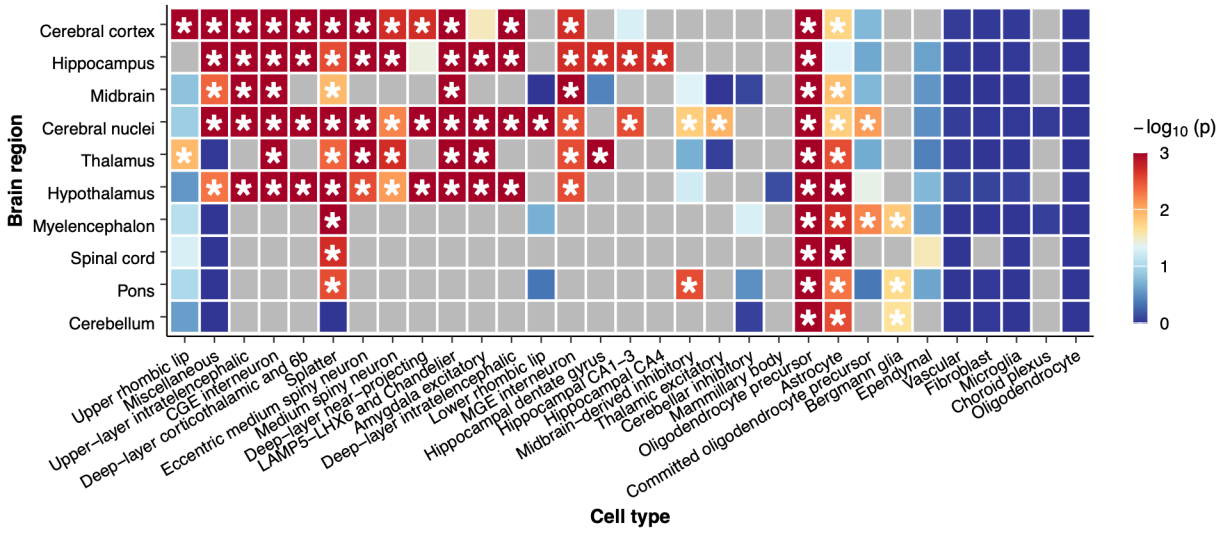
Supplementary Fig. S6 Triangulated causal effects of human complex traits on EA in EAS across multiple MR methods. Traits with significant GSMR causal effects on EA in EAS after FDR correction ($\mathrm{FDR} < 0.05$) were shown. The row-wise panels show different traits. Circles are colored to denote different MR methods. The y-axis represents b_{xy} estimates from each method. The error bars denote 95% confidence intervals. The horizontal dashed line indicates $b_{xy} = 0$.



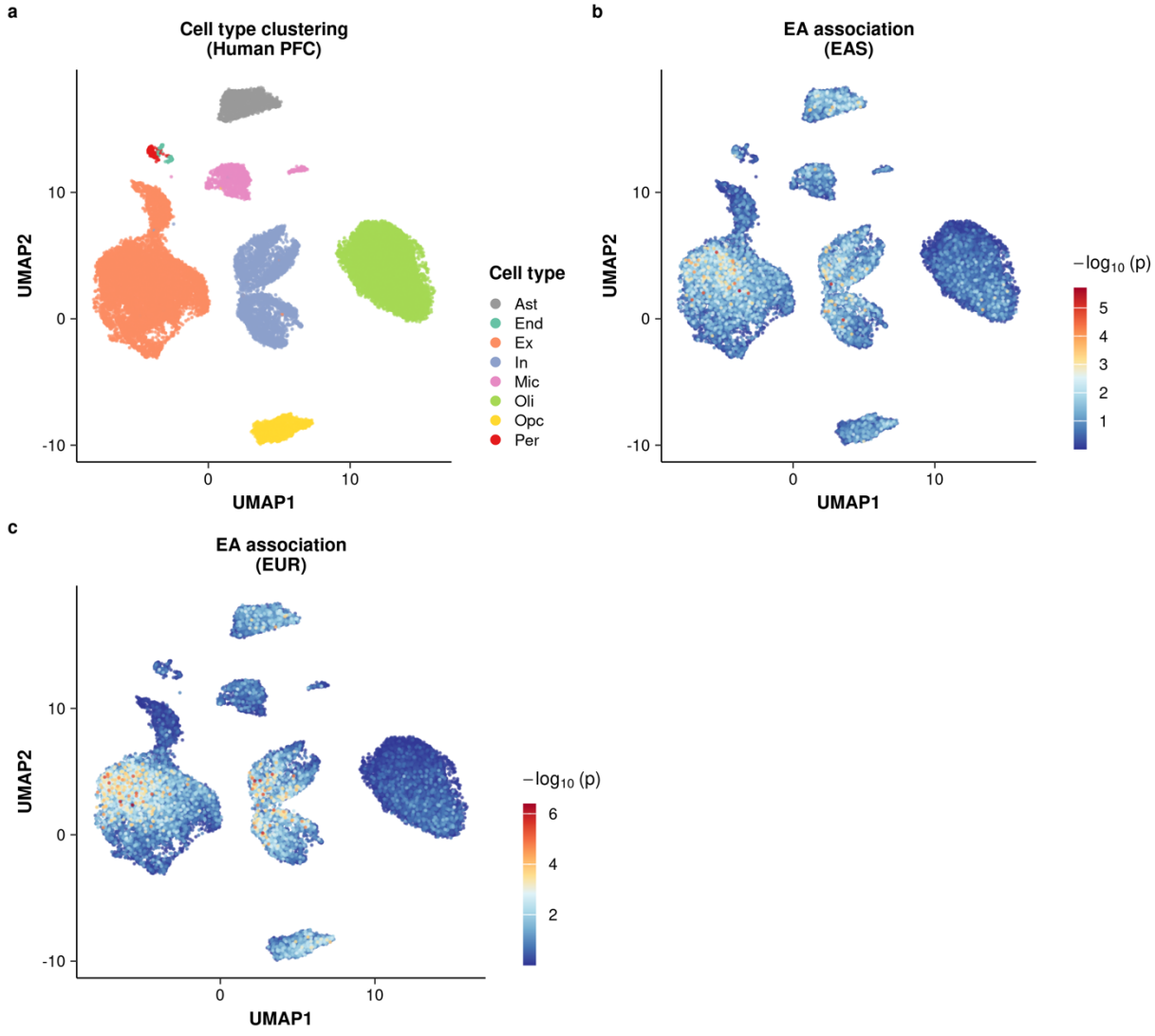
Supplementary Fig. S7 Stratified LD score regression (S-LDSC) analysis. a Heritability enrichment of EA across functional genomic annotations. The x-axis represents the per-SNP heritability. The y-axis represents different genomic functional regions, with the proportion in parentheses indicating the fraction of SNPs belonging to each category. Bar colors denote the proportion of heritability explained. The gray dashed line indicates an enrichment ratio of 1 (null). Significantly enriched functional annotations ($\mathrm{FDR} < 0.05$) are displayed in descending order of significance ($-log_{10}(P_{\mathrm{enrichment}})$), with repressed regions included as a negative control ($P_{\mathrm{enrichment}} = 0.531$). **b** Heritability enrichment of EA in tissue-specific open chromatin regions. Each point represents a tissue, colored by major tissue category. The y-axis denotes enrichment significance. The gray dashed line represents the Bonferroni-corrected significance threshold ($P = 0.05/490$) on a $-log_{10}$ scale.



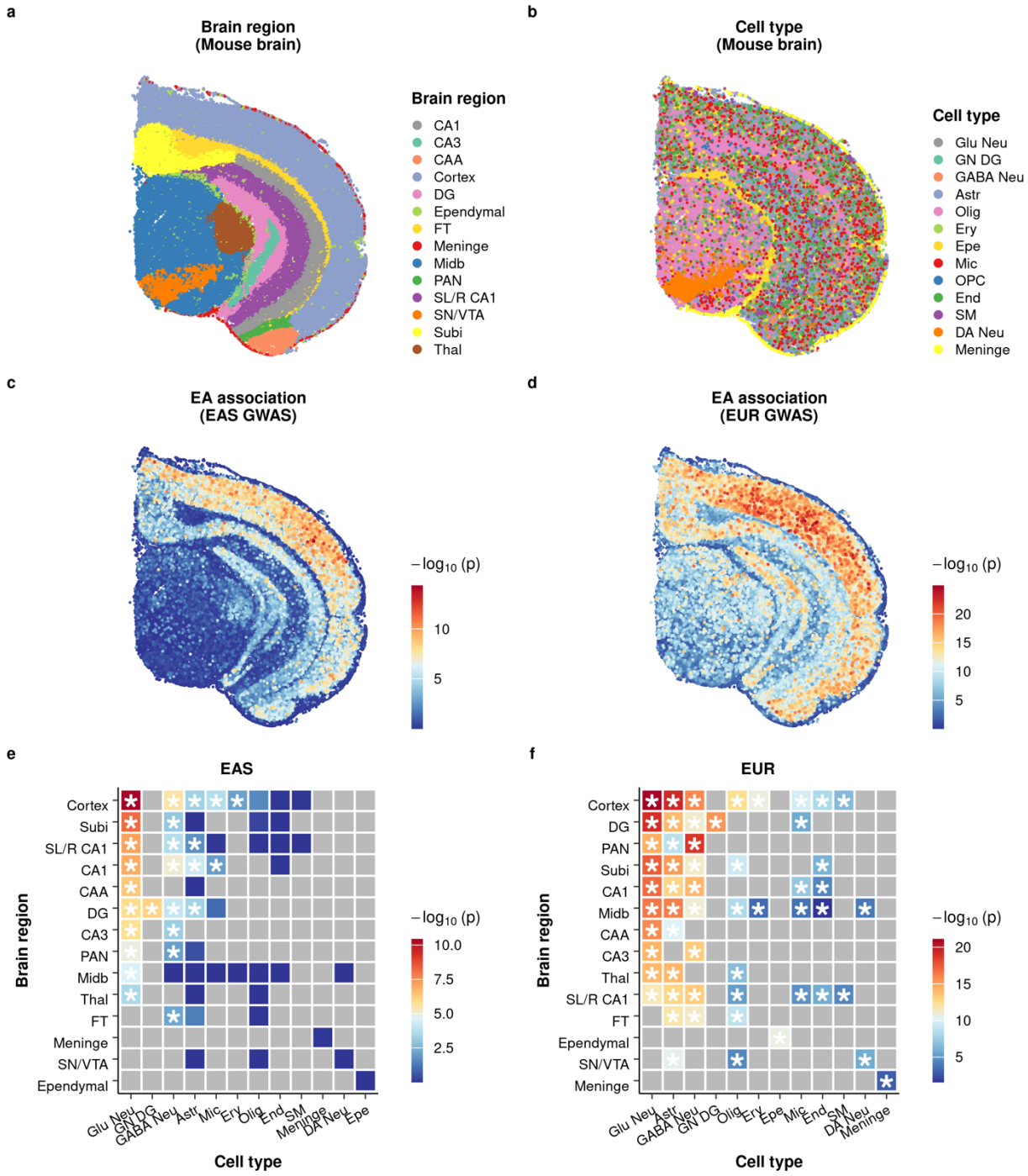
Supplementary Fig. S8 Cell-type association across ten human brain regions based on cross-ancestry EA GWAS meta-analysis. Cell-type association was calculated using scDRS. Colors indicate association significance, with significantly associated subpopulations of cells after FDR correction marked by asterisks. Gray tiles indicate subpopulations with fewer than 100 cells, which were excluded from the analysis. Rows represent anatomical brain regions, and columns represent established cell superclusters.



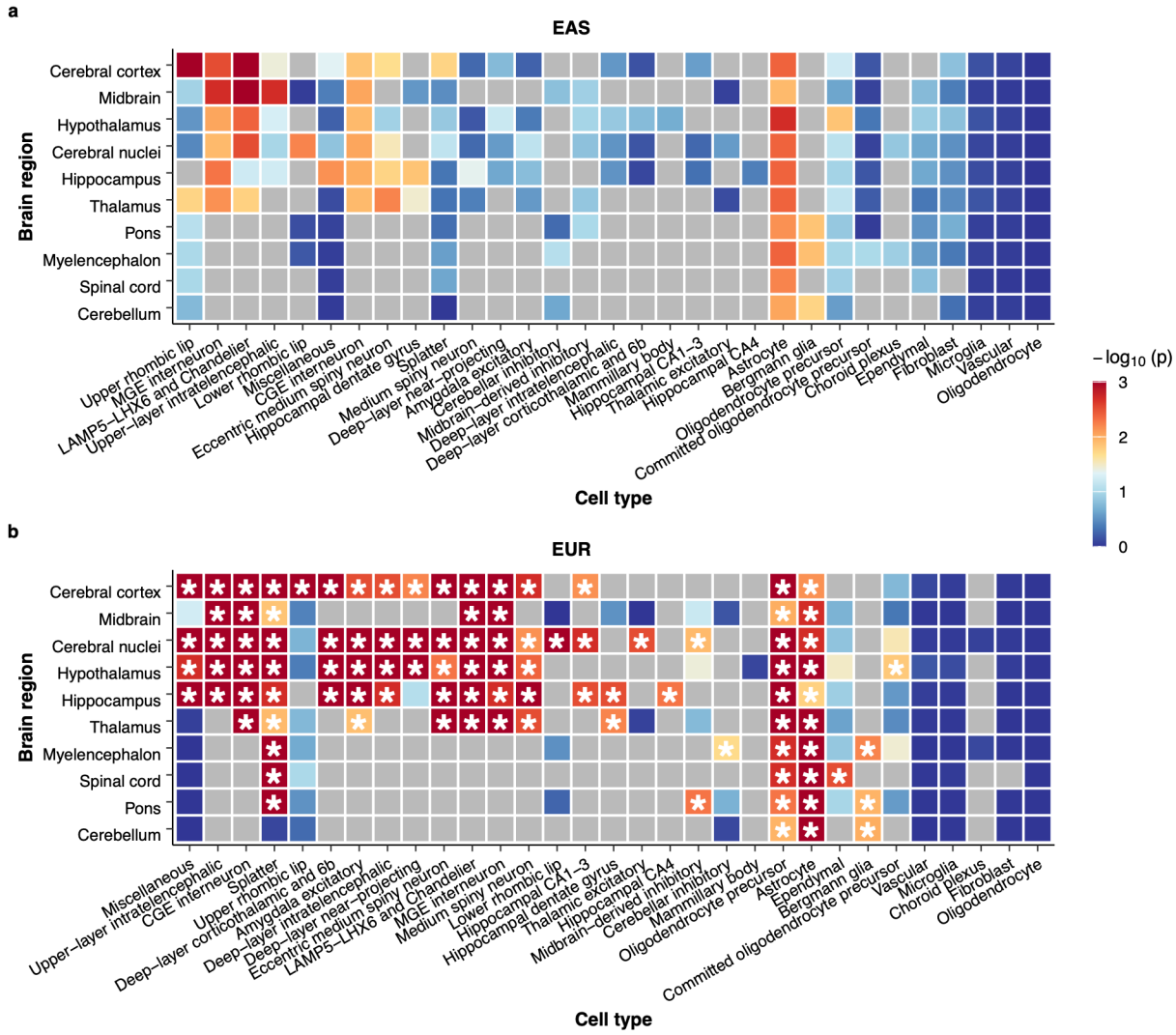
Supplementary Fig. S9 Cell-type association in the human prefrontal cortex based on ancestry-stratified EA GWAS. a UMAP plot of cell type clustering in the human prefrontal cortex. Abbreviations: Ast, astrocyte; End, endothelial cell; Ex, excitatory neuron; In, inhibitory neuron; Mic, microglia; Oli, oligodendrocyte; Opc, oligodendrocyte precursor cell; Per, pericyte. **b** Cellular-level associations calculated by scDRS based on the EAS GWAS. **c** Cellular-level associations calculated by scDRS based on the EUR GWAS. Cluster boundaries remain consistent with cell type clusters from panel **a**. The association significance of each cell is indicated by color intensity.



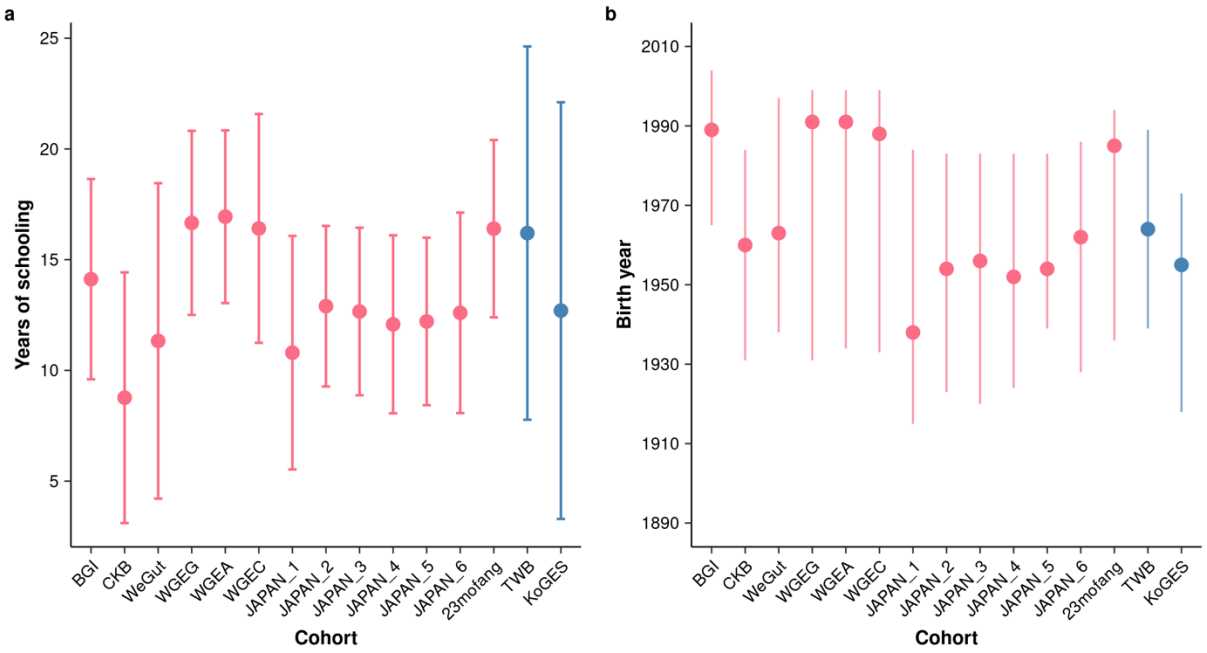
Supplementary Fig. S10 Cell-type association across brain regions in the mouse brain based on ancestry-stratified EA GWAS. a Mouse brain ST data, with spots colored by anatomical brain region. **b** Mouse brain ST data, with spots colored by established cell types. **c** Spatial mapping of EA-associated cellular signals using gsMap based on the EAS GWAS. **d** Spatial mapping of EA-associated cellular signals using gsMap based on the EUR GWAS. Spatial distributions of cells remain consistent with those in panel **a,b**. The association significance of each cell is indicated by colors. **e** Cell-type association across brain regions in mouse ST data via gsMap based on the EAS GWAS. **f** Cell-type association across brain regions in mouse ST data by gsMap based on the EUR GWAS. Colors indicate association significance, with significantly associated subpopulations of cells after FDR correction marked by asterisks. Gray tiles indicate subpopulations with fewer than 100 cells, which were excluded from the analysis. Rows represent anatomical brain regions, and columns represent established cell superclusters. Abbreviations: Subi, subiculum; CA1, cornu ammonis 1 region; SL/R CA1, stratum lacunosum-moleculare/radiatum of the hippocampal cornu ammonis area 1; CAA, cortical amygdalar area; DG, dentate gyrus; PAN, posterior amygdalar nucleus; Midb, midbrain; CA3, cornu ammonis area 3; Thal, thalamus; FT, fiber tract; SN/VTA, substantia nigra/ventral tegmental area; Glu Neu, Excitatory glutamatergic neuron; GN DG, granule cell of dentate gyrus; Astr, astrocyte; GABA Neu, GABAergic neuron; Ery, erythrocytes; Olig, oligodendrocytes; Epe, ependymal cell; Mic, microglia; Opc, oligodendrocyte precursor cell; SM, smooth muscle cell; End, endothelial cell; DA Neu, dopaminergic neuron.


a
Brain region
(Mouse brain)
Brain region
CA1
CA3
CAA
Cortex
DG
Ependymal
FT
Meninge
Midb
PAN
SL/R CA1
SN/VTA
Subi
Thal
b
Cell type
(Mouse brain)
Cell type
Glu Neu
GN DG
GABA Neu
Astr
Olig
Ery
Epe
Mic
OPC
End
SM
DA Neu
Meninge
c
EA association
(EAS GWAS)
−log10 (p)
10
5
d
EA association
(EUR GWAS)
−log10 (p)
20
15
10
5
e
EAS
Brain region
Cortex
Subi
SL/R CA1
CA1
CAA
DG
CA3
PAN
Midb
Thal
FT
Meninge
SN/VTA
Ependymal
Glu Neu
GN DG
GABA Neu
Astr
Mic
Ery
Olig
End
SM
Meninge
DA Neu
Epe
Cell type
−log10 (p)
10.0
7.5
5.0
2.5
f
EUR
Brain region
Cortex
DG
PAN
Subi
CA1
Midb
CAA
CA3
Thal
SL/R CA1
FT
Ependymal
SN/VTA
Meninge
Glu Neu
Astr
GABA Neu
GN DG
Olig
Ery
Epe
Mic
End
SM
DA Neu
Meninge
Cell type
−log10 (p)
20
15
10
5

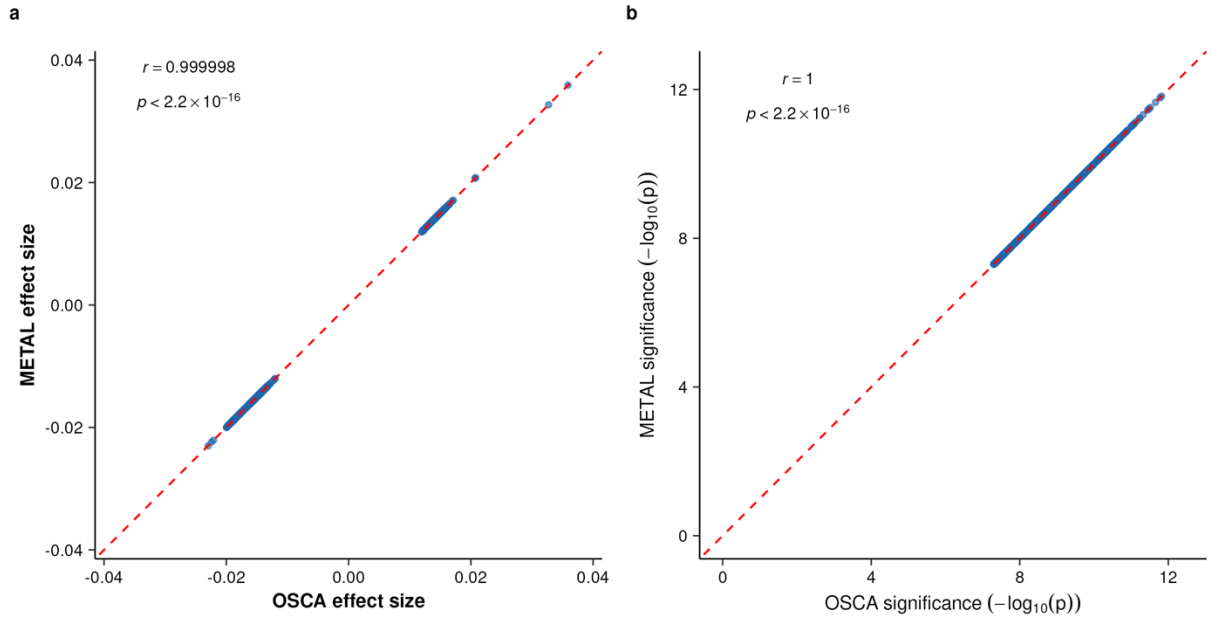
Supplementary Fig. S11 Cell-type association across ten human brain regions based on ancestry-stratified EA GWAS. a Cell-type association across ten human brain regions by scDRS based on the EAS GWAS. **b** Cell-type association across ten human brain regions by scDRS based on the EUR GWAS. Colors indicate association significance, with significantly associated subpopulations of cells after FDR correction marked by asterisks. Gray tiles indicate subpopulations with fewer than 100 cells, which were excluded from the analysis. Rows represent anatomical brain regions, and columns represent established cell superclusters.



Supplementary Fig. S12 Cohort-level characteristics of all EAS cohorts. a Distribution of years of schooling across all EAS cohorts. Dots represent mean values, and error bars represent 95% confidence intervals. New datasets are colored pink; datasets from the previous EAS study (Chen et al.) are colored blue. **b** Mean birth years and birth-year ranges across all EAS cohorts. Dots represent mean values for new datasets (pink) and median values for datasets from the previous EAS study (blue), and extended lines represent the range of birth years for each cohort.



Supplementary Fig. S13 Consistency of the inverse-variance-weighted meta-analysis results between OSCA and METAL for genome-wide significant variants. a Comparison of estimated effect sizes between the two tools for genome-wide significant variants. **b** Comparison of association *P*-values between the two tools for genome-wide significant variants.



Supplementary Fig. S14 Locus plots of the *DCAKD* gene. Each panel displays genetic associations of EA, fine-mapping posterior inclusion probabilities (PIP) by CARMA, and relevant molecular QTL signals in EAS, alongside genetic associations in EUR (EA3) at the locus. The GWAS lead SNP in EAS is denoted by a purple diamond, and the colors of other dots represent LD r^2 with the EAS lead SNP. Below the locus plots, chromatin interaction tracks illustrate epigenomic signatures, with red bars denoting promoter regions and purple arcs indicating interactions between the target gene promoter and the epigenomic marker harboring the lead SNP. The genomic position of the focal gene is shown at the bottom.

