Population-specific causal disease effect sizes in functionally important regions impacted by selection

Huwenbo Shi[1,2,*], Steven Gazal[1,2], Masahiro Kanai[2,3,4,5,6], Evan M. Koch[7,8], Armin P. Schoech[1,2,9], Katherine M. Siewert[1,2], Samuel S. Kim[1,2,10], Yang Luo[2,5,7,11,12], Tiffany Amariuta[2,5,11,12,13], Hailiang Huang[3,4,8], Yukinori Okada[6,14], Soumya Raychaudhuri[2,5,7,11,12,15], Shamil R. Sunyaev[7,8], and Alkes L. Price[1,2,9,†]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[4]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA
[5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[6]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
[7]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[8]Department of Medicine, Harvard Medical School, Boston, MA, USA
[9]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[10]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[11]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[12]Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[13]Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA
[14]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan
[15]Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK

Correspondence: *hshi@hsph.harvard.edu (HS), †aprice@hsph.harvard.edu (ALP)

Abstract

Many diseases and complex traits exhibit population-specific causal effect sizes with trans-ethnic genetic correlations significantly less than 1, limiting trans-ethnic polygenic risk prediction. We developed a new method, S-LDXR, for stratifying squared trans-ethnic genetic correlation across genomic annotations, and applied S-LDXR to genome-wide association summary statistics for 31 diseases and complex traits in East Asians (EAS) and Europeans (EUR) (average N_{EAS}=90K, N_{EUR}=267K) with an average trans-ethnic genetic correlation of 0.85 (s.e. 0.01). We determined that squared trans-ethnic genetic correlation was 0.82× (s.e. 0.01) smaller than the genome-wide average at SNPs in the top quintile of background selection statistic, implying more population-specific causal effect sizes. Accordingly, causal effect sizes were more population-specific in functionally important regions, including conserved and regulatory regions. In analyses of regions surrounding specifically expressed genes, causal effect sizes were most population-specific for skin and immune genes and least population-specific for brain genes. Our results could potentially be explained by stronger gene-environment interaction at loci impacted by selection, particularly positive selection.

Introduction

Trans-ethnic genetic correlations are significantly less than 1 for many diseases and complex traits,[1–6] implying that population-specific causal disease effect sizes contribute to the incomplete portability of genome-wide association study (GWAS) findings and polygenic risk scores to non-European populations.[6–12] However, current methods for estimating genome-wide trans-ethnic genetic correlations assume the same trans-ethnic genetic correlation for all categories of SNPs,[2,5,13] providing little insight into why causal disease effect sizes are population-specific. Understanding the biological processes contributing to population-specific causal disease effect sizes can help inform polygenic risk prediction in non-European populations and alleviate health disparities.[6,14,15]

Here, we introduce a new method, S-LDXR, for estimating enrichment of stratified squared trans-ethnic genetic correlation across functional categories of SNPs using GWAS summary statistics and population-matched linkage disequilibrium (LD) reference panels (e.g. the 1000 Genomes Project (1000G)[16]); we stratify the *squared* trans-ethnic genetic correlation across functional categories to robustly handle noisy heritability estimates. S-LDXR analyzes GWAS summary statistics of HapMap3[17] SNPs with minor allele frequency (MAF) greater than 5% in both East Asian (EAS) and European (EUR) populations (*regression SNPs*) to draw inferences about causal effects of all SNPs with MAF greater than 5% in both populations (*heritability SNPs*). We confirm that S-LDXR yields robust estimates in extensive simulations. We apply S-LDXR to 31 diseases and complex traits with GWAS summary statistics available in both East Asian (EAS) and European (EUR) populations, leveraging recent large studies in East Asian populations from the CONVERGE consortium and Biobank Japan;[18–20] we analyze a broad set of genomic annotations from the baseline-LD model,[21–23] as well as tissue-specific annotations based on specifically expressed gene sets.[24] Most results are meta-analyzed across the 31 traits to maximize power (analogous to ref.[21–23]), as we expect to see similar patterns of enrichment/depletion across traits (even though the underlying biological processes differ across traits). We also investigate trait-specific enrichments/depletions for the tissue-specific annotations (analogous to ref.[24]).

Results

Overview of methods

Our method (S-LDXR) for estimating stratified trans-ethnic genetic correlation is conceptually related to stratified LD score regression[21,22] (S-LDSC), a method for partitioning heritability from GWAS summary statistics while accounting for LD. S-LDXR determines that a category of SNPs is enriched for trans-ethnic genetic covariance if SNPs with high LD to that category have higher product of Z-scores than SNPs with low LD to that category. Unlike S-LDSC, S-LDXR models per-allele effect sizes (accounting for differences in MAF between populations).

In detail, the product of Z-scores of SNP j in two populations, $Z_{1j}Z_{2j}$, has the expectation

$$\mathrm{E}[Z_{1j}Z_{2j}] = \sqrt{N_1 N_2} \sum_C \ell_\times(j, C)\theta_C \ , \tag{1}$$

where N_p is the sample size for population p; $\ell_\times(j, C) = \sum_k r_{1jk} r_{2jk} \sigma_{1j} \sigma_{2j} a_C(k)$ is the trans-ethnic LD score of SNP j with respect to annotation C, whose value for SNP k, $a_C(k)$, can be either binary or continuous; r_{pjk} is the LD (Pearson correlation) between SNP j and k in population p; σ_{pj} is the standard deviation of SNP j genotypes in population p; and θ_C represents the per-SNP contribution to trans-ethnic genetic covariance of the *per-allele* causal disease effect size of annotation C. Here, r_{pjk} and σ_{pj} can be estimated from population-matched reference panels (e.g. 1000 Genomes Project[16]). We estimate θ_C for each annotation C using weighted least square regression. Subsequently, we estimate the trans-ethnic genetic covariance of each binary annotation C ($\rho_g(C)$) by summing trans-ethnic genetic covariance of each SNP in annotation C as $\sum_{j \in C} (\sum_{C'} a_{C'}(j)\theta_{C'})$, using coefficients ($\theta_{C'}$) for all binary and continuous-valued annotations C' included in the analysis; the heritabilities in each population ($h^2_{g1}(C)$ and $h^2_{g2}(C)$) are estimated analogously. We then estimate the stratified *squared* trans-ethnic genetic correlation, defined as

$$r_g^2(C) = \frac{\rho_g^2(C)}{h^2_{g1}(C) h^2_{g2}(C)} \ . \tag{2}$$

We define the enrichment/depletion of squared trans-ethnic genetic correlation as $\lambda^2(C) = \frac{r_g^2(C)}{r_g^2}$, where r_g^2 is the genome-wide squared trans-ethnic genetic correlation; $\lambda^2(C)$ can be meta-analyzed across traits with different r_g^2. S-LDXR analyzes GWAS summary statistics of common HapMap3[17] SNPs (*regression SNPs*) to estimate $\lambda^2(C)$ for (causal effects of) all common SNPs (*heritability SNPs*). Further details (quantities estimated, analytical bias

correction, shrinkage estimator to reduce standard errors, estimation of standard errors, significance testing, and factors impacting power) of the S-LDXR method are provided in the Methods section; we have publicly released open-source software implementing the method (see URLs).

We apply S-LDXR to 62 annotations (defined in both EAS and EUR populations) from our baseline-LD-X model (Methods, Table S1, Figures S1, S2), primarily derived from the baseline-LD model[21–23] (v1.1; see URLs). We have publicly released all baseline-LD-X model annotations and LD scores for EAS and EUR populations (see URLs).

Simulations

We evaluated the accuracy of S-LDXR in simulations using genotypes that we simulated using HAPGEN2[25] from phased haplotypes of 481 EAS and 489 EUR individuals from the 1000 Genomes Project[16], preserving population-specific MAF and LD patterns (18,418 simulated EAS-like and 36,836 simulated EUR-like samples, after removing genetically related samples, ratio of sample sizes similar to empirical data; ∼2.5 million SNPs on chromosomes 1 – 3) (Methods); we did not have access to individual-level EAS data at sufficient sample size to perform simulations with real genotypes. For each population, we randomly selected a subset of 500 simulated samples to serve as the reference panel for estimating LD scores. We performed both null simulations (heritable trait with functional enrichment but no enrichment/depletion of squared trans-ethnic genetic correlation; $\lambda^2(C) = 1$) and causal simulations ($\lambda^2(C) \neq 1$). In our main simulations, we randomly selected 10% of the SNPs as causal SNPs in both populations, set genome-wide heritability to 0.5 in each population, and adjusted genome-wide genetic covariance to attain a genome-wide r_g of 0.60 (unless otherwise indicated). In the null simulations, we used heritability enrichments from analyses of real traits in EAS samples to specify per-SNP causal effect size variances and covariances. In the causal simulations, we directly specified per-SNP causal effect size variances and covariances to attain $\lambda^2(C) \neq 1$ values from analyses of real traits, as these were difficult to attain using the heritability and trans-ethnic genetic covariance enrichments from analyses of real traits.

First, we assessed the accuracy of S-LDXR in estimating genome-wide trans-ethnic genetic correlation (r_g); we note that S-LDXR does not use the shrinkage estimator for genome-wide estimates. Across a wide range of simulated r_g values (0.0 to 1.0), S-LDXR yielded approximately unbiased estimates and well-calibrated jackknife standard errors (Table S3, Figure S3).

Second, we assessed the accuracy of S-LDXR in estimating $\lambda^2(C)$ in quintiles of the 8 continuous-valued annotations of the baseline-LD-X model. We performed both null simu-

lations ($\lambda^2(C) = 1$) and causal simulations ($\lambda^2(C) \neq 1$). Results are reported in Figure 1a and Tables S4 – S9 . In both null and causal simulations, S-LDXR yielded approximately unbiased estimates of $\lambda^2(C)$ for most annotations, validating our analytical bias correction. As a secondary analysis, we tried varying the S-LDXR shrinkage parameter, α, which has a default value of 0.5. In null simulations, results remained approximately unbiased; in causal simulations, reducing α led to less precise (but less biased) estimates of $\lambda^2(C)$, whereas increasing α biased results towards the null ($\lambda^2(C) = 1$), demonstrating a bias-variance tradeoff in the choice of α (Figure S4, Tables S5, S8). Results were similar at other values of the proportion of causal SNPs (1% and 100%; Tables S4, S6, S7, S9). We also confirmed that S-LDXR produced well-calibrated jackknife standard errors (Tables S4-S9).

Third, we assessed the accuracy of S-LDXR in estimating $\lambda^2(C)$ for the 28 main binary annotations of the baseline-LD-X model (inherited from the baseline model of ref.[21]). We discarded $\lambda^2(C)$ estimates with the highest standard errors (top 5%), as estimates with large standard errors (which are particularly common for annotations of small size) are uninformative for evaluating unbiasedness of the estimator (in analyses of real traits, trait-specific estimates with large standard errors are retained, but contribute very little to meta-analysis results, and would be interpreted as inconclusive when assessing trait-specific results). Results are reported in Figure 1b and Tables S5, S8. In null simulations, S-LDXR yielded unbiased estimates of $\lambda^2(C)$, further validating our analytical bias correction. In causal simulations, estimates were biased towards the null ($\lambda^2(C) = 1$) – particularly for annotations of small size (proportion of SNPs $< 1\%$) – due to our shrinkage estimator; increasing the shrinkage parameter above its default value of 0.5 further biased the estimates towards the null ($\lambda^2(C) = 1$) in causal simulations (Tables S7, S8, S9). To ensure robust estimates, we focus on the 20 main binary annotations of large size ($> 1\%$ of SNPs) in analyses of real traits (see below); although results for these annotations may still be biased towards the null, we emphasize that S-LDXR is unbiased in null data. Results were similar at other values of the proportion of causal SNPs (1% and 100%; Tables S4, S6, S7, S9). We also confirmed that S-LDXR produced well-calibrated jackknife standard errors (Tables S4-S9) and conservative p-values (Figure S7, Table S4-S9).

Fourth, we performed additional null simulations in which causal variants differed across the two populations (Methods). S-LDXR yielded robust estimates of $\lambda^2(C)$, well-calibrated standard errors and conservative p-values in these simulations (Figure S11, Table S12).

Fifth, we performed additional null simulations with annotation-dependent MAF-dependent genetic architectures,[26–28] defined as architectures in which the level of MAF-dependence is annotation-dependent, to ensure that estimate of $\lambda^2(C)$ remain unbiased. We disproportionately sampled low-frequency causal variants from the top quintile of background selection

statistic, and set the variance of per-allele effect sizes of a causal SNP to be inversely proportional to its maximum MAF across both populations (Methods). Results are reported in Figure S8-S10, and Table S10, S11. S-LDXR yielded nearly unbiased estimates of $\lambda^2(C)$ for the 28 binary functional annotations (Figure S8) and nearly unbiased estimates of $\lambda^2(C)$ for most quintiles of continuously valued annotations (Figure S9); estimates were slightly biased estimates in the top and bottom quintile of the average level of LD annotation and the recombination rate annotation, likely due to less accurate reference LD scores at SNPs with extreme levels of LD. We repeated these simulations with 5 MAF bin annotations added to the baseline-LD-X model and obtained similar results (Figure S8a, S9b), supporting our decision not to include MAF bin annotations into the baseline-LD-X model.

Sixth, we performed additional null simulations, in which we increased or decreased the reference panel size from 500 to 250 or 1,000, to assess the impact of reference panel size on the accuracy of S-LDXR (Methods). We simulated GWAS summary statistics based on the baseline-LD-X model as well as the model with annotation-dependent MAF-dependent genetic architectures. We determined that the small systematic biases in null simulations of continuous-valued annotations were on the same order of magnitude as for 500 reference samples (Figures S12, S13 and Table S13 for 250 reference samples; Figures S14, S15 and Table S14 for 1,000 reference samples). We also performed simulations in which we reduced the simulated GWAS sample size by half, from N_{EAS}=18K, N_{EUR}=37K to N_{EAS}=9K, N_{EUR}=18K (while fixing the reference panel size at 500). We again determined that the small systematic biases were generally on the same order of magnitude as for N_{EAS}=18K, N_{EUR}=37K (although estimates were less stable and sometimes subject to larger biases, likely because our analytical bias correction starts to break down when the GWAS has low power) (Figures S16, S17 and Table S15). Although it was not computationally feasible to perform simulations at larger GWAS sample sizes, these analyses do not provide a reason to believe that the small systematic biases that we observed in some of our null simulations of continuously valued annotations would substantially increase at larger GWAS sample sizes.

In summary, S-LDXR produced approximately unbiased estimates of enrichment/depletion of squared trans-ethnic genetic correlation in null simulations, and conservative estimates in causal simulations of both quintiles of continuous-valued annotations and binary annotations.

Analysis of baseline-LD-X model annotations across 31 diseases and complex traits

We applied S-LDXR to 31 diseases and complex traits with summary statistics in East Asians (average N=90K) and Europeans (average N=267K) available from Biobank Japan,

UK Biobank, and other sources (Table S16 and Methods). First, we estimated the trans-ethnic genetic correlation (r_g) (as well as population-specific heritabilies) for each trait. Results are reported in Figure S18 and Table S16. The average r_g across 31 traits was 0.85 (s.e. 0.01) (average $r_g^2 = 0.72$ (s.e. 0.02)). 28 traits had $r_g < 1$, and 11 traits had r_g significantly less than 1 after correcting for 31 traits tested ($P < 0.05/31$); the lowest r_g was 0.34 (s.e. 0.07) for Major Depressive Disorder (MDD), although this may be confounded by different diagnostic criteria in the two populations.[29] Several other complex traits, including Age at Menopause ($r_g = 0.57$ (s.e. 0.09)) and LDL ($r_g = 0.66$ (s.e. 0.11)) also had low trans-ethnic r_g, likely due to pervasive gene-environment interaction across the genome. These estimates were consistent with estimates obtained using Popcorn[2] (Figure S19) and those reported in previous studies.[2,5,6] We note that our estimates of trans-ethnic genetic correlation for 31 complex traits are higher than those reported for gene expression traits[2] (average estimate of 0.32, increasing to 0.77 when restricting the analysis to gene expression traits with (cis) heritability greater than 0.2 in both populations), which are expected to have different genetic architectures.

Second, we estimated the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$) in quintiles of the 8 continuous-valued annotations of the baseline-LD-X model, meta-analyzing results across traits; these annotations are moderately correlated (Figure 2a and Table S1). We used the default shrinkage parameter ($\alpha = 0.5$) in all analyses. Results are reported in Figure 2b and Table S17. We consistently observed a depletion of $r_g^2(C)$ ($\lambda^2(C) < 1$, implying more population-specific causal effect sizes) in functionally important regions. For example, we estimated $\lambda^2(C) = 0.82$ (s.e. 0.01) for SNPs in the top quintile of background selection statistic (defined as $1 -$ McVicker B statistic / 1000;[30] see ref.[22]); $\lambda^2(C)$ estimates were less than 1 for 29/31 traits, including 2 traits (Height and EGFR) with two-tailed $p < 0.05/31$. The background selection statistic quantifies the genetic distance of a site to its nearest exon; regions with high background selection statistic have higher per-SNP heritability, consistent with the action of selection, and are enriched for functionally important regions.[22] We observed the same pattern for CpG content and SNP-specific F_{st} (which are positively correlated with background selection statistic; Figure 2a) and the opposite pattern for nucleotide diversity (which is negatively correlated with background selection statistic). We also estimated $\lambda^2(C) = 0.87$ (s.e. 0.03) for SNPs in the top quintile of average LLD (which is positively correlated with background selection statistic), although these SNPs have *lower* per-SNP heritability due to a competing positive correlation with predicted allele age.[22] We caution that average LLD was the annotation most susceptible to bias in our simulations; see Simulations. Likewise, we estimated $\lambda^2(C) = 0.84$ (s.e. 0.02) for SNPs in the *bottom* quintile of recombination rate (which is negatively correlated with

background selection statistic), although these SNPs have average per-SNP heritability due to a competing negative correlation with average LLD.[22] However, $\lambda^2(C) < 1$ estimates for the bottom quintile of GERP (NS) (which is positively correlated with both background selection statistic and recombination rate) and the middle quintile of predicted allele age are more difficult to interpret. For all annotations analyzed, heritability enrichments did not differ significantly between EAS and EUR, consistent with previous studies.[20,31] Results were similar at a more stringent shrinkage parameter value ($\alpha = 1.0$; Figure S20), and for a meta-analysis across a subset of 20 approximately independent traits (Methods; Figure S21).

Finally, we estimated $\lambda^2(C)$ for the 28 main binary annotations of the baseline-LD-X model (Table S1), meta-analyzing results across traits (as we did not observe significant trait-specific enrichment/depletion of squared trans-ethnic genetic correlation for these annotations due to limited power). Results are reported in Figure 3a and Table S18. Our primary focus is on the 20 annotations of large size ($> 1\%$ of SNPs), for which our simulations yielded robust estimates; results for remaining annotations are reported in Table S18. We consistently observed a depletion of $\lambda^2(C)$ (implying more population-specific causal effect sizes) within these annotations: 17 annotations had $\lambda^2(C) < 1$, and 5 annotations had $\lambda^2(C)$ significantly less than 1 after correcting for 20 annotations tested ($P < 0.05/20$). These annotations included Conserved ($\lambda^2(C) = 0.93$ (s.e. 0.02)), Promoter ($\lambda^2(C) = 0.85$ (s.e. 0.04)) and Super Enhancer ($\lambda^2(C) = 0.93$ (s.e. 0.02)), each of which was significantly enriched for per-SNP heritability, consistent with ref.[21]. For all annotations analyzed, heritability enrichments did not differ significantly between EAS and EUR (Figure 3a), consistent with previous studies.[20,31] Results were similar at a more stringent shrinkage parameter value ($\alpha = 1.0$; Figure S20), and for a meta-analysis across a subset of 20 approximately independent traits (Methods; Figure S22). As a secondary analysis, we also estimated $\lambda^2(C)$ across 10 MAF bin annotations; we did not observe variation in $\lambda^2(C)$ estimates across MAF bins (Table S19), further supporting our decision to not include MAF bin annotations in the baseline-LD-X model.

Since the functional annotations are moderately correlated with the 8 continuous-valued annotations (Table S1c, Figure S1), we investigated whether the depletions of squared trans-ethnic genetic correlation ($\lambda^2(C) < 1$) within the 20 binary annotations could be explained by the 8 continuous-valued annotations. For each binary annotation, we estimated its expected $\lambda^2(C)$ based on values of the 8 continuous-valued annotations for SNPs in the binary annotation (Methods), meta-analyzed this quantity across traits, and compared observed vs. expected $\lambda^2(C)$ (Figure 3b and Table S20). We observed strong concordance, with a slope of 0.57 (correlation of 0.61) across the 20 binary annotations. This implies that the deple-

tions of $r_g^2(C)$ ($\lambda^2(C) < 1$) within binary annotations are largely explained by corresponding values of continuous-valued annotations.

In summary, our results show that causal disease effect sizes are more population-specific in functionally important regions impacted by selection. Further interpretation of these findings, including the role of positive and/or negative selection, is provided in the Discussion section.

Analysis of specifically expressed gene annotations

We analyzed 53 specifically expressed gene (SEG) annotations, defined in ref.[24] as $\pm$100kb regions surrounding the top 10% of genes specifically expressed in each of 53 GTEx[32] tissues (Table S2), by applying S-LDXR with the baseline-LD-X model to the 31 diseases and complex traits (Table S16). We note that although SEG annotations were previously used to prioritize disease-relevant tissues based on disease-specific heritability enrichments,[20,24] enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$) is standardized with respect to heritability (i.e. increase in heritability in the denominator would lead to increase in trans-ethnic genetic covariance in the numerator (Equation (2))), hence not expected to produce extremely disease-specific signals. Thus, we first assess meta-analyzed $\lambda^2(C)$ estimates across the 31 diseases and complex traits (trait-specific estimates are assessed below).

Results are reported in Figure 4a and Table S21. $\lambda^2(C)$ estimates were less than 1 for all 53 tissues and significantly less than 1 ($p < 0.05/53$) for 37 tissues, with statistically significant heterogeneity across tissues ($p < 10^{-20}$; Methods). The strongest depletions of squared trans-ethnic genetic correlation were observed in skin tissues (e.g. $\lambda^2(C) = 0.83$ (s.e. 0.02) for Skin Sun Exposed (Lower Leg)), Prostate and Ovary ($\lambda^2(C) = 0.84$ (s.e. 0.02) for Prostate, $\lambda^2(C) = 0.86$ (s.e. 0.02) for Ovary) and immune-related tissues (e.g. $\lambda^2(C) = 0.85$ (s.e. 0.02) for Spleen), and the weakest depletions were observed in Testis ($\lambda^2(C) = 0.98$ (s.e. 0.02); no significant depletion) and brain tissues (e.g. $\lambda^2(C) = 0.98$ (s.e. 0.02) for Brain Nucleus Accumbens (Basal Ganglia); no significant depletion). Results were similar at less stringent and more stringent shrinkage parameter values ($\alpha = 0.0$ and $\alpha = 1.0$; Figures S23, S24 and Table S21). A comparison of 14 blood-related traits and 16 other traits yielded highly consistent $\lambda^2(C)$ estimates ($R = 0.82$; Figure S25, Table S22), confirming that these findings were not extremely disease-specific.

These $\lambda^2(C)$ results were consistent with the higher background selection statistic[30] in Skin Sun Exposed (Lower Leg) ($R = 0.17$), Prostate ($R = 0.16$) and Spleen ($R = 0.14$) as compared to Testis ($R = 0.02$) and Brain Nucleus Accumbens (Basal Ganglia) ($R = 0.08$) (Figure S26, Table S2), and similarly for CpG content (Figure S27, Table S2). Although

these results could in principle be confounded by gene size,[33] the low correlation between gene size and background selection statistic ($R = 0.06$) or CpG content ($R = -0.20$) (in $\pm$100kb regions) implies limited confounding. We note the well-documented action of recent positive selection on genes impacting skin pigmentation,[34–38] the immune system,[34–37,39] and Ovary;[40] we are not currently aware of any evidence of positive selection impacting Prostate. We further note the well-documented action of negative selection on fecundity- and brain-related traits,[26,28,41] but it is possible that recent positive selection may more closely track differences in causal disease effect sizes across human populations, which have split relatively recently[42] (see Discussion).

More generally, since SEG annotations are moderately correlated with the 8 continuous-valued annotations (Figure S28, Table S2), we investigated whether these $\lambda^2(C)$ results could be explained by the 8 continuous-valued annotations (analogous to Figure 3b). Results are reported in Figure 4b and Table S23. We observed strong concordance, with a slope of 0.96 (correlation of 0.76) across the 53 SEG annotations. This implies that the depletions of $\lambda^2(C)$ within SEG annotations are explained by corresponding values of continuous-valued annotations.

The strong depletion of squared trans-ethnic genetic correlation in tissues impacted by positive selection (as opposed to negative selection) suggests a possible connection between positive selection and population-specific causal effect sizes. To further assess this, we estimated the enrichment/depletion of squared trans-ethnic genetic correlation in SNPs with high integrated haplotype score (iHS),[43,44] which quantifies the action of positive selection (Methods). We observed a significant depletion ($\lambda^2(C) = 0.88$ (s.e. 0.03)), further implicating positive selection (however, it is difficult to assess whether the iHS annotation contains unique information about $\lambda^2(C)$ conditional on other annotations; see Discussion). In addition, we observed a high genome-wide trans-ethnic genetic correlation for schizophrenia ($r_g = 0.95$ (s.e. 0.04) vs. average of 0.85 (s.e. 0.01) across traits), a psychiatric disorder hypothesized to be strongly impacted by negative selection,[45,46] suggesting that negative selection may play a limited role in population-specific causal effect sizes. As noted above, these estimates pertain to parameters that were defined based on common variants (see Overview of methods); we note that although negative selection has the strongest impact on low-frequency variants,[26] common variants are also impacted by negative selection and can inform inferences about negative selection.[22] The role of positive selection (as opposed to negative selection) in population-specific causal effect sizes is discussed further in the Discussion section.

We investigated the enrichment/depletion of $\lambda^2(C)$ in the 53 specifically expressed gene annotations for each individual trait (Table S24). We identified 6 significantly depleted (vs.

0 significantly enriched) trait-tissue pairs at per-trait $p < 0.05/53$. The limited number of statistically significant results was expected, due to the reduced power of trait-specific analyses; however, $\lambda^2(C)$ estimates were generally consistent across traits. Results for BMI and height, two widely studied anthropometric traits, are reported in Figure 5. For BMI, we observed significant depletion of squared trans-ethnic genetic correlation ($\lambda^2(C) = 0.84$ (s.e. 0.05)) in Pituitary. Previous studies have highlighted the role of Pituitary in obesity;[47–49] our results suggest that this tissue-specific mechanism is population-specific. For height, we observed significant depletion of squared trans-ethnic genetic correlation for Transformed fibroblasts ($\lambda^2(C) = 0.87$ (s.e. 0.03)), a connective tissue linked to human developmental disorders;[50] again, our results suggest that this tissue-specific mechanism is population-specific. Although Pituitary was significantly depleted for BMI but not height, and Transformed fibroblasts was significantly depleted for height but not BMI, we caution that for both tissues our $\lambda^2(C)$ estimates did not differ significantly between BMI and height.

In summary, our results show that causal disease effect sizes are more population-specific in regions surrounding specifically expressed genes.

Discussion

We developed a new method (S-LDXR) for stratifying squared trans-ethnic genetic correlation across functional categories of SNPs that yields approximately unbiased estimates in extensive simulations. By applying S-LDXR to East Asian and European summary statistics across 31 diseases and complex traits, we determined that SNPs with high background selection statistic[30] have substantially depleted squared trans-ethnic genetic correlation (vs. the genome-wide average), implying that causal effect sizes are more population-specific. Accordingly, squared trans-ethnic genetic correlations were substantially depleted for SNPs in many functional categories and enriched in less functionally important regions (although the power of S-LDXR to detect enrichment of squared trans-ethnic genetic correlation is limited due to depletion of heritability in less functionally important regions). In analyses of specifically expressed gene annotations, we observed substantial depletion of squared trans-ethnic genetic correlation for SNPs near skin and immune-related genes, which are strongly impacted by recent positive selection, but not for SNPs near brain genes. We also observed trait-specific depletions of squared-trans-ethnic genetic correlation for specifically expressed gene annotations, which indicate population-specific disease mechanisms.

Reductions in trans-ethnic genetic correlation have several possible underlying explanations, including gene-environment (G×E) interaction, gene-gene (G×G) interaction, and dominance variation (but not differences in heritability across populations, which would not affect trans-ethnic genetic correlation and were not observed in our study). Given the increasing evidence of the role of G×E interaction in complex trait architectures,[51] and evidence that G×G interaction and dominance variation explain limited heritability,[52–54] we hypothesize that depletion of squared trans-ethnic genetic correlation in the top quintile of background selection statistic and in functionally important regions may be primarily attributable to stronger G×E interaction in these regions. Interestingly, a recent study on plasticity in Arabidopsis observed a similar phenomenon: lines with more extreme phenotypes exhibited stronger G×E interaction.[55] Although depletion of squared trans-ethnic genetic correlation is often observed in regions with higher per-SNP heritability, which may often be subject to stronger G×E, depletion may also occur in regions with lower per-SNP heritability that are subject to stronger G×E; we hypothesize that this is the case for SNPs in the top quintile of average LLD and the bottom quintile of GERP (NS) (Figure 2).

Distinguishing between stronger G×E interaction in regions impacted by selection and stronger G×E interaction in functionally important regions as possible explanations for our findings is a challenge, because functionally important regions are more strongly impacted by selection. To this end, we constructed an annotation that is similar to the background

selection statistic but does not make use of recombination rate, instead relying solely on a SNP's physical distance to the nearest exon (Methods). Applying S-LDXR to the 31 diseases and complex traits using a joint model incorporating baseline-LD-X model annotations and the nearest exon annotation, the background selection statistic remained highly conditionally informative for trans-ethnic genetic correlation, whereas the nearest exon annotation was not conditionally informative (Table S25). This result implicates stronger G×E interaction in regions with reduced effective population size that are impacted by selection, and not just proximity to functional regions, in explaining depletions of squared trans-ethnic genetic correlation; however, we emphasize that selection acts on allele frequencies rather than causal effect sizes, and could help explain our findings only in conjunction with other explanations such as G×E interaction. Our results on specifically expressed genes implicate stronger G×E interaction near skin, immune, and ovary genes and weaker G×E interaction near brain genes, potentially implicating positive selection (as opposed to negative selection). This conclusion is further supported by the significant depletion of squared trans-ethnic genetic correlation in the integrated haplotype score (iHS) annotation that specifically reflects positive selection, high genome-wide trans-ethnic genetic correlation for schizophrenia (Table S16), and lack of variation in squared trans-ethnic genetic correlation across genes in different deciles of probability of loss-of-function intolerance[56] (Methods, Figure S29, S30, Table S26). We conclude that depletions of squared trans-ethnic genetic correlation could potentially be explained by stronger G×E interaction at loci impacted by positive selection. We caution that other explanations are also possible; in particular, evolutionary modeling using an extension of the Eyre-Walker model[57] to two populations suggests that our results for the background selection statistic could also be consistent with negative selection (Supplementary Note, Figure S31, S32, Table S27). Additional information, such as genomic annotations that better distinguish different types of selection or data from additional diverse populations, may help elucidate the relationship between selection and population-specific causal effect sizes.

Our study has several implications. First, polygenic risk scores (PRS) in non-European populations that make use of European training data[6,9,11] may be improved by reweighting SNPs based on the expected enrichment/depletion of squared trans-ethnic genetic correlation, helping to alleviate health disparities.[6,14,15] For example, when applying LD-pruning + p-value thresholding methods,[58,59] both the strength of association and trans-ethnic genetic correlation should be accounted for when prioritizing SNPs for trans-ethnic PRS, as our results suggest that trans-ethnic genetic correlation is likely depleted near functional SNPs with significant p-values (due to stronger G×E). In particular, when multiple SNPs have similar level of significance, the SNPs enriched for trans-ethnic genetic correlation should be priori-

tized. Analogously, when applying more recent methods that estimate posterior mean causal effect sizes[60–66] (including functionally informed methods[62,66]), these estimates should subsequently be weighted according to the expected enrichment/depletion for squared trans-ethnic genetic correlation based on their functional annotations. Second, modeling population-specific genetic architectures may improve trans-ethnic fine-mapping. Our results suggest that causal effect sizes and/or causal variants are likely to differ across different populations, contrary to standard assumptions.[31,67] Thus, incorporating information about trans-ethnic genetic correlations in trans-ethnic fine-mapping may lead to more accurate identification of both population-specific and shared causal variants.[68] Third, modeling population-specific genetic architectures may also increase power in trans-ethnic meta-analysis,[69] e.g. by adapting MTAG[70] to two populations (instead of two traits) , leveraging trans-ethnic (instead of cross-trait) genetic correlation between pairs of populations to improve estimation of SNP effect sizes in both populations. Fourth, it may be of interest to stratify G×E interaction effects[51] across genomic annotations. Fifth, modeling and incorporating environmental variables, where available, may provide additional insights into population-specific causal effect sizes. In our simulations, we did not explicitly simulate G×E. However, G×E would induce population-specific causal effect sizes, which we did explicitly simulate. Sixth, the S-LDXR method could potentially be extended to stratify squared *cross-trait* genetic correlations[71] across genomic annotations.[72]

We note several limitations of this study that pertain to the S-LDXR method. First, S-LDXR is designed for populations of homogeneous continental ancestry (e.g. East Asians and Europeans) and is not currently suitable for analysis of admixed populations[73] (e.g. African Americans or admixed Africans from UK Biobank[74]), analogous to LDSC and its published extensions.[21,71,75] However, a recently proposed extension of LDSC to admixed populations[76] could be incorporated into S-LDXR, enabling its application to the growing set of large studies in admixed populations.[10] Second, S-LDXR estimates of enrichment of stratified squared trans-ethnic genetic correlation ($\lambda^2(C)$) are slightly downward biased in null simulations of the top quintile of the background selection statistic and average LLD annotations, especially in simulations involving annotation-dependent MAF-dependent genetic architectures. However, these biases are small compared to the depletions of $\lambda^2(C)$ observed in analysis of real traits. We further note that our estimates are unbiased in null simulations of binary annotations, implying that our results on real traits for binary annotations are extremely robust. Third, since S-LDXR applies shrinkage to reduce standard error in estimating stratified squared trans-ethnic genetic correlation and its enrichment, estimates are conservative – true depletions of squared trans-ethnic genetic correlation in functionally important regions may be stronger than the estimated depletions. However, we

emphasize that S-LDXR is approximately unbiased in null data. Fourth, the optimal value of the shrinkage parameter α may be specific to the pair of populations analyzed. In our simulations, we determined that $\alpha = 0.5$ provides a satisfactory bias-variance tradeoff across a wide range of values of polygenicity and power. Thus, $\alpha = 0.5$ may also be satisfactory for other pairs of populations. However, we recommend that one should ideally perform simulations on the pair of populations being analyzed to selection the optimal value of α. Fifth, it is difficult to assess whether a focal annotation contains unique information about $\lambda^2(C)$ conditional on other annotations, as squared trans-ethnic genetic correlation is a non-linear quantity defined by the quotient of squared trans-ethnic genetic covariance and the product of heritabilities in each population.

We also note several limitations of this study that pertain to our analysis of real traits. First, we focused on comparisons of East Asians and Europeans, due to limited availability of very large GWAS in other populations. For other pairs of continental populations, if differences in environment are similar, then we would expect similar genome-wide trans-ethnic genetic correlation and similar enrichment/depletion of squared trans-ethnic genetic correlation, based on our hypothesis that imperfect trans-ethnic genetic correlation is primarily attributable to G×E. We also note that different set of SNPs, with different MAF and LD patterns, would be analyzed for different pairs of populations. However, we expect that these differences would not contribute to differences in trans-ethnic genetic correlation, if G×E is the fundamental factor impacting trans-ethnic genetic correlation. Second, the specifically expressed gene (SEG) annotations analyzed in this study are defined predominantly (but not exclusively) based on gene expression measurements of Europeans.[24] We hypothesize that results based on SEG annotations defined in East Asian populations would likely be similar, as heritability enrichment of functional annotations (predominantly defined in Europeans) are consistent across continental populations,[20,31] despite the fact that gene expression patterns and genetic architectures of gene expression differ across diverse populations.[12,77,78] Thus, SEG annotations derived from gene expression data from diverse populations may provide additional insights into population-specific causal effect sizes. Third, we restricted our analyses to SNPs that were relatively common (MAF>5%) in both populations (estimating parameters that were defined based on common SNPs), due to the lack of a large LD reference panel for East Asians. Extending our analyses to lower-frequency SNPs may provide further insights into the role of negative selection in shaping population-specific genetic architectures, as negative selection has the strongest impact on variants with low frequency.[26,27] Fourth, we did not consider population-specific variants in our analyses, due to the difficulty in defining trans-ethnic genetic correlation for population-specific variants[2,5], a more fundamental challenge than analyzing low-frequency SNPs; a

recent study[79] has reported that population-specific variants substantially limit trans-ethnic genetic risk prediction accuracy. Fifth, estimates of genome-wide trans-ethnic genetic correlation may be confounded by different trait definitions or diagnostic criteria in the two populations, particularly for major depressive disorder. However, this would not impact estimates of enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$), which is defined relative to genome-wide values. Sixth, we have not pinpointed the exact underlying phenomena (e.g. environmental heterogeneity coupled with gene-environment interaction) that lead to population-specific causal disease effect sizes at functionally important regions. Despite these limitations, our study provides an improved understanding of the underlying biology that contribute to population-specific causal effect sizes, and highlights the need for increasing diversity in genetic studies.

URLs

- S-LDXR software: https://github.com/huwenboshi/s-ldxr/
- Python code for simulating GWAS summary statistics: https://github.com/huwenboshi/s-ldxr-sim/
- baseline-LD-X model annotations and LD scores: https://data.broadinstitute.org/alkesgroup/S-LDXR/
- Distance to nearest exon annotation and LD scores: https://data.broadinstitute.org/alkesgroup/S-LDXR/
- baseline-LD model annotations: https://alkesgroup.broadinstitute.org/LDSCORE/
- 1000 Genomes Project: https://www.internationalgenome.org/
- PLINK2: https://www.cog-genomics.org/plink/2.0/
- HAPGEN2: https://mathgen.stats.ox.ac.uk/genetics_software/hapgen/hapgen2.html
- UCSC Genome Browser: https://genome.ucsc.edu/
- Exome Aggregation Consortium (ExAC): https://exac.broadinstitute.org/
- Integrated haplotype scores (iHS): http://coruscant.itmat.upenn.edu/data/JohnsonEA_iHSscores.tar.gz

Methods

Definition of stratified squared trans-ethnic genetic correlation

We model a complex phenotype in two populations using linear models, $\boldsymbol{Y}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1$ and $\boldsymbol{Y}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2$, where $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are vectors of phenotype measurements of population 1 and population 2 with sample size N_1 and N_2 , respectively; $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are mean-centered *but not normalized* genotype matrices at M SNPs in the two populations; $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are *per-allele causal* effect sizes of the M SNPs; and $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are environmental effects in the two populations. We assume that in each population, genotypes, causal effect sizes, and environmental effects are independent from each other. We assume that the per-allele effect size of SNP j in the two populations has variance and covariance,

$$\begin{aligned} \mathrm{Var}[\beta_{1j}] = \sum_C a_C(j)\tau_{1C},\ \mathrm{Var}[\beta_{2j}] = \sum_C a_C(j)\tau_{2C}, \\ \mathrm{Cov}[\beta_{1j}, \beta_{2j}] = \sum_C a_C(j)\theta_C, \end{aligned} \tag{3}$$

where $a_C(j)$ is the value of SNP j for annotation C, which can be binary or continuous-valued; τ_{1C} and τ_{2C} are the net contribution of annotation C to the variance of β_{1j} and β_{2j}, respectively; and θ_C is the net contribution of annotation C to the covariance of β_{1j} and β_{2j}.

We define stratified trans-ethnic genetic correlation of a binary annotation C (e.g. functional annotations[21] or quintiles of continuous-valued annotations[22]) as,

$$r_g(C) = \frac{\rho_g(C)}{\sqrt{h^2_{g1}(C)}\sqrt{h^2_{g2}(C)}}, \tag{4}$$

where $\rho_g(C) = \sum_{j\in C} \mathrm{Cov}[\beta_{1j}, \beta_{2j}] = \sum_{j\in C}\sum_{C'} a_{C'}(j)\theta_{C'}$ is the trans-ethnic genetic covariance of annotation C; and $h^2_{gp}(C) = \sum_{j\in C} \mathrm{Var}[\beta_{pj}] = \sum_{j\in C}\sum_{C'} a_{C'}(j)\tau_{pC'}$ is the "allelic-scale heritability" (sum of per-SNP variance of per-allele causal effect sizes; different from heritability on the standardized scale) of annotation C in population p. Here, C' includes all binary and continuous-valued annotations included in the analysis. Since estimates of $h^2_{gp}(C)$ can be noisy (possibly negative), we estimate *squared* stratified trans-ethnic genetic correlation,

$$r^2_g(C) = \frac{\rho^2_g(C)}{h^2_{g1}(C)h^2_{g2}(C)}, \tag{5}$$

to avoid bias or undefined values in the square root. In this work, we only estimate $r^2_g(C)$ for SNPs with minor allele frequency (MAF) greater than 5% in both populations. To assess

whether causal effect sizes are more or less correlated for SNPs in annotation C compared with the genome-wide average, r_g^2, we define the enrichment/depletion of stratified squared trans-ethnic genetic correlation as

$$\lambda^2(C) = \frac{r_g^2(C)}{r_g^2}. \tag{6}$$

We meta-analyze $\lambda^2(C)$ instead of $r_g^2(C)$ across diseases and complex traits. For continuous-valued annotations, defining $r_g^2(C)$ and $\lambda^2(C)$ is challenging, as squared correlation is a non-linear term involving a quotient of squared covariance and a product of variances; we elected to instead estimate $\lambda^2(C)$ for quintiles of continuous-valued annotations (analogous to ref.[22]). We note that the average value of $\lambda^2(C)$ across quintiles of continuous-valued annotations is not necessarily equal to 1, as squared trans-ethnic genetic correlation is a non-linear quantity.

S-LDXR method

S-LDXR is conceptually related to stratified LD score regression[21,22] (S-LDSC), a method for stratifying heritability from GWAS summary statistics, to two populations. The S-LDSC method determines that a category of SNPs is enriched for heritability if SNPs with high LD to that category have higher expected χ^2 statistic than SNPs with low LD to that category. Analogously, the S-LDXR method determines that a category of SNPs is enriched for trans-ethnic genetic covariance if SNPs with high LD to that category have higher expected product of Z-scores than SNPs with low LD to that category.

S-LDXR relies on the regression equation

$$\mathrm{E}[Z_{1j}Z_{2j}] = \sqrt{N_1 N_2} \sum_C \ell_\times(j, C)\theta_C \tag{7}$$

to estimate θ_C, where Z_{pj} is the Z-score of SNP j in population p; $\ell_\times(j, C) = \sum_k r_{1jk} r_{2jk} \sigma_{1j} \sigma_{2j} a_C(k)$ is the trans-ethnic LD score of SNP j with respect to annotation C, whose value for SNP k, $a_C(k)$, can be either binary or continuous; r_{pjk} is the LD between SNP j and k in population p; and σ_{pj} is the standard deviation of SNP j in population p. We obtain unbiased estimates of $\ell_\times(j, C)$ using genotype data of 481 East Asian and 489 European samples in the 1000 Genomes Project.[16] To account for heteroscedasticity and increase statistical efficiency, we use weighted least square regression to estimate θ_C. We use regression equations analogous to those described in ref.[21] to estimate τ_{1C} and τ_{2C}. We include only well-imputed (imputation INFO>0.9) and common (MAF>5% in both populations) SNPs that are present in

HapMap 3[17] (irrespective of GWAS significance level) in the regressions (*regression SNPs*), analogous to our previous work.[21,71,75] We use all SNPs present in either population in 1000 Genomes[16] to estimate trans-ethnic LD scores $\ell_{\times}(j, C)$ (*reference SNPs*; analogous to S-LDSC[21]), so that the resulting coefficients θ_C also pertain to these SNPs. However, we estimate $r_g^2(C)$ and $\lambda^2(C)$ (see below; defined as a function of causal effect sizes) for all SNPs with MAF>5% in both populations (*heritability SNPs*), accounting for tagging effects (analogous to S-LDSC[21]).

Let $\hat{\tau}_{1C}$, $\hat{\tau}_{2C}$, and $\hat{\theta}_C$ be the estimates of τ_{1C}, τ_{2C}, and θ_C, respectively. For each binary annotation C, we estimate the stratified heritability of annotation C in each population, $h_{g1}^2(C)$ and $h_{g2}^2(C)$, and trans-ethnic genetic covariance, $\rho_g(C)$, as

$$\hat{h}_{g1}^2(C) = \sum_{j \in C} \sum_{C'} a_{jC'} \hat{\tau}_{1C'},\ \hat{h}_{g2}^2(C) = \sum_{j \in C} \sum_{C'} a_{jC'} \hat{\tau}_{2C'},\ \hat{\rho}_g(C) = \sum_{j \in C} \sum_{C'} a_{jC'} \hat{\theta}_{C'}, \tag{8}$$

respectively, restricting to causal effects of SNPs with MAF>5% in both populations (*heritability SNPs*), using coefficients ($\tau_{1C'}$, $\tau_{2C'}$, and $\theta_{C'}$) of both binary and continuous-valued annotations. We estimate genome-wide trans-ethnic genetic correlation as $\hat{r}_g = \frac{\hat{\rho}_g(\mathcal{C})}{\sqrt{\hat{h}_{g1}^2(\mathcal{C})\hat{h}_{g1}^2(\mathcal{C})}}$, where $\mathcal{C}$ represents the set of all SNPs with MAF>5% in both populations. We then estimate $r_g^2(C)$ as

$$\hat{r}_g^2(C) = \left\{ \tilde{r}_g^2(C) + \frac{\mathrm{Cov}\left[\hat{\rho}_g^2(C), \hat{h}_{g1}^2(C)\hat{h}_{g2}^2(C)\right]}{\hat{h}_{g1}^2(C)\hat{h}_{g2}^2(C)} \right\} \Big/ \left\{ 1 + \frac{\mathrm{Var}\left[\hat{h}_{g1}^2(C)\hat{h}_{g2}^2(C)\right]}{\hat{h}_{g1}^2(C)\hat{h}_{g2}^2(C)} \right\}, \tag{9}$$

where $\tilde{r}_g^2(C) = \frac{\hat{\rho}_g^2(C) - \mathrm{Var}[\hat{\rho}_g(C)]}{\hat{h}_{g1}^2(C)\hat{h}_{g2}^2(C) - \mathrm{Cov}[\hat{h}_{g1}^2(C), \hat{h}_{g2}^2(C)]}$. The correction to $\tilde{r}_g^2(C)$ in Equation (9) is necessary for obtaining an unbiased estimate of $r_g^2(C)$, as computing quotients of two random variables introduces bias (Supplementary Note). (We do not constrain the estimate of $r_g^2(C)$ to its plausible range of $[-1, 1]$, as this would introduce bias.) Subsequently, we estimate enrichment of stratified squared trans-ethnic genetic correlation as

$$\hat{\lambda}^2(C) = \left\{ \tilde{\lambda}^2(C) + \frac{\mathrm{Cov}\left[\hat{r}_g^2(C), \hat{r}_g^2\right]}{\hat{r}_g^2(C)} \right\} \Big/ \left\{ 1 + \frac{\mathrm{Var}\left[\hat{r}_g^2(C)\right]}{\hat{r}_g^2(C)} \right\} \tag{10}$$

where $\tilde{\lambda}^2(C) = \frac{\hat{r_g^2}(C)}{\hat{r_g^2}}$, the ratio between estimated stratified ($\hat{r}_g^2(C)$) and genome-wide ($\hat{r}_g^2$) squared trans-ethnic genetic correlation. We use block jackknife over 200 non-overlapping and equally sized blocks to obtain standard error of all estimates. The standard error of $\lambda^2(C)$ primarily depends on the total allelic-scale heritability of SNPs in the annotation

(sum of per-SNP variances of causal per-allele effect sizes), which appears as the denominator ($h^2_{1g}(C)h^2_{2g}(C)$) in the estimation of stratified squared trans-ethnic genetic correlation ($r^2_g(C)$); if this denominator is small, estimation of $r^2_g(C)$ becomes noisy. The standard error of $\lambda^2(C)$ indirectly depends on the size of the annotation, because larger annotations tend to have larger total heritability. However, estimates of $\lambda^2(C)$ for a large annotation may have large standard error if the annotation is depleted for heritability.

To assess the informativeness of each annotation in explaining disease heritability and trans-ethnic genetic covariance, we define standardized annotation effect size on heritability and trans-ethnic genetic covariance for each annotation C analogous to ref.[22],

$$\tau^*_{1C} = \frac{Mh^2_{g1}}{h^2_{g1}(C)} \times \sigma_C \times \tau_{1C},\ \tau^*_{2C} = \frac{Mh^2_{g2}}{h^2_{g2}(C)} \times \sigma_C \times \tau_{2C},$$
$$\theta^*_C = \frac{M\rho_g}{\rho_g(C)} \times \sigma_C \times \theta_C, \tag{11}$$

where τ^*_{1C}, τ^*_{2C}, and θ^*_C represent proportionate change in per-SNP heritability in population 1 and 2 and trans-ethnic genetic covariance, respectively, per standard deviation increase in annotation C; τ_{1C}, τ_{2C}, and θ_C are the corresponding unstandardized effect sizes, defined in Equation (3); and σ_C is the standard deviation of annotation C.

We provide a more detailed description of the method, including derivations of the regression equation and unbiased estimators of the LD scores, in the **Supplementary Note**.

S-LDXR shrinkage estimator

Estimates of $r^2_g(C)$ can be imprecise with large standard errors if the denominator, $h^2_{g1}(C)h^2_{g2}(C)$, is close to zero and noisily estimated. This is especially the case for annotations of small size ($< 1\%$ SNPs). We introduce a shrinkage estimator to reduce the standard error in estimating $r^2_g(C)$.

Briefly, we shrink the estimated per-SNP heritability and trans-ethnic genetic covariance of annotation C towards the genome-wide averages, which are usually estimated with smaller standard errors, prior to estimating $r^2_g(C)$. In detail, let M_C be the number of SNPs in annotation C, we shrink $\frac{\hat{h}^2_{1g}(C)}{M_C}$, $\frac{\hat{h}^2_{2g}(C)}{M_C}$, and $\frac{\hat{\rho}_g(C)}{M_C}$ towards $\frac{\hat{h}^2_{1g}}{M}$, $\frac{\hat{h}^2_{2g}}{M}$, and $\frac{\hat{\rho}_g}{M}$, respectively, where $\hat{h}^2_{g1}$, $\hat{h}^2_{g2}$, $\hat{\rho}_g$ are the genome-wide estimates, and M the total number of SNPs. We obtain the shrinkage as follows. Let $\gamma_1 = 1/\left(1 + \alpha\frac{\mathrm{Var}\left[\hat{h}^2_{g1}(C)\right]}{\mathrm{Var}\left[\hat{h}^2_{g1}\right]}\frac{M}{M_C}\right)$, $\gamma_2 = 1/\left(1 + \alpha\frac{\mathrm{Var}\left[\hat{h}^2_{g2}(C)\right]}{\mathrm{Var}\left[\hat{h}^2_{g2}\right]}\frac{M}{M_C}\right)$, and $\gamma_3 = 1/\left(1 + \alpha\frac{\mathrm{Var}[\hat{\rho}_g(C)]}{\mathrm{Var}[\hat{\rho}_g]}\frac{M}{M_C}\right)$ be the shrinkage obtained separately for $\hat{h}^2_{g1}(C)$, $\hat{h}^2_{g2}(C)$ and $\hat{\rho}_g(C)$, respectively, where $\alpha \in [0, 1]$ is the shrinkage parameter adjusting magnitude of

shrinkage. We then choose the most stringent shrinkage, $\gamma = \min\{\gamma_1, \gamma_2, \gamma_3\}$, as the final shared shrinkage for both heritability and trans-ethnic genetic covariance.

We shrink heritability and trans-ethnic genetic covariance of annotation C using γ as, $\bar{h}^2_{g1}(C) = M_C\left(\gamma\frac{\hat{h}^2_{g1}(C)}{M_C} + (1-\gamma)\frac{\hat{h}^2_{g1}}{M}\right)$, $\bar{h}^2_{g2}(C) = M_C\left(\gamma\frac{\hat{h}^2_{g2}(C)}{M_C} + (1-\gamma)\frac{\hat{h}^2_{g2}}{M}\right)$, and $\bar{\rho}_g(C) = M_C\left(\gamma\frac{\hat{\rho}_g(C)}{M_C} + (1-\gamma)\frac{\hat{\rho}_g}{M}\right)$, where $\bar{h}^2_{g1}(C)$, $\bar{h}^2_{g2}(C)$, and $\bar{\rho}_g(C)$ are the shrunk counterparts of $\hat{h}^2_{g1}(C)$, $\hat{h}^2_{g2}(C)$, and $\hat{\rho}_g(C)$, respectively. We shrink $\hat{r}^2_g(C)$ by substituting $\hat{h}^2_{g1}(C)$, $\hat{h}^2_{g2}(C)$, and $\hat{\rho}_g(C)$ with $\bar{h}^2_{g1}(C)$, $\bar{h}^2_{g2}(C)$, $\bar{\rho}_g(C)$, respectively, in Equation (9), to obtain its shrunk counterpart, $\bar{r}^2_g(C)$. Finally, we shrink $\hat{\lambda}^2(C)$, by plugging in $\bar{r}^2_g(C)$ in Equation (10) to obtain its shrunk counterpart, $\bar{\lambda}^2(C)$. We recommend $\alpha = 0.5$ as the default shrinkage parameter value, as this value provides robust estimates of $\lambda^2(C)$ in simulations. We note that S-LDXR does not use the shrinkage estimator when estimating genome-wide r_g and r^2_g.

Significance testing

To assess whether an annotation C is enriched or depleted of squared trans-ethnic genetic correlation for a trait, we test the null hypothesis $\hat{\lambda}^2(C) = 1$. Since $\hat{\lambda}^2(C)$ is not normally distributed,[80] we instead test the equivalent null hypothesis $\hat{D}^2(C) = \hat{\rho}^2_g(C) - \hat{r}^2_g\hat{h}^2_{g1}(C)\hat{h}^2_{g1}(C) = 0$, where $\hat{r}^2_g$ is the genome-wide squared trans-ethnic genetic correlation. We obtain test statistic as $\frac{\hat{D}^2(C)}{s.e.[\hat{D}^2(C)]}$, and obtain p-value under t-distribution with $B-1$ degrees of freedom, where B is the number of jackknife blocks. Since the $\hat{D}^2(C)$ statistic does not involve division by $\hat{h}^2_{g1}(C)\hat{h}^2_{g1}(C)$, we do not apply any shrinkage to $\hat{D}^2(C)$.

Baseline-LD-X model

We include a total of 54 binary functional annotations in the baseline-LD-X model. These include 53 annotations introduced in ref.,[21] which consists of 28 main annotations including conserved annotations (e.g. Coding, Conserved) and epigenomic annotations (e.g. H3K27ac, DHS, Enhancer) derived from ENCODE[81] and Roadmap,[82] 24 500-base-pair-extended main annotations, and 1 annotation containing all SNPs. We note that although chromatin accessibility can be population-specific, the fraction of such regions is small.[83] Following ref,[22] we created an additional annotation for all genomic positions with number of rejected substitutions[84] greater than 4. Further information for all functional annotations included in the baseline-LD-X model is provided in Table S1a.

We also include a total of 8 continuous-valued annotations in the baseline-LD-X model. First, we include 5 continuous-valued annotations introduced in ref.[22] (see URLs), without modification: background selection statistic,[30] CpG content (within a ± 50 kb window),

GERP (number of substitutation) score,[84] nucleotide diversity (within a ±10 kb window), and Oxford map recombination rate (within a ±10 kb window).[85] Second, we include 2 minor allele frequency (MAF) adjusted annotations introduced in ref.,[22] with modification: level of LD (LLD) and predicted allele age. We created analogous annotations applicable to both East Asian and European populations. To create an analogous LLD annotation, we estimated LD scores for each population using LDSC,[75] took the average across populations, and then quantile-normalized the average LD scores using 10 average MAF bins. We call this annotation "average level of LD". To create analogous predicted allele age annotation, we quantile-normalized allele age estimated by ARGweaver[86] across 54 multi-ethnic genomes using 10 average MAF bins. Finally, we include 1 continuous-valued annotation based on F_{ST} estimated by PLINK2,[87] which implements the Weir & Cockerham estimator of F_{ST}.[88] Further information for all continuous-valued annotations included in the baseline-LD-X model is provided in Table S1b.

Code and data availability

Python code implementing S-LDXR is available at https://github.com/huwenboshi/s-ldxr. Python code for simulating GWAS summary statistics under the baseline-LD-X model is available at https://github.com/huwenboshi/s-ldxr-sim. baseline-LD-X model annotations and LD scores are available at https://data.broadinstitute.org/alkesgroup/S-LDXR/.

Simulations

We used simulated East Asian (EAS) and European (EUR) genotype data to assess the performance our method, as we did not have access to real EAS genotype data at sufficient sample size to perform simulations with real genotypes. We simulated genotype data for 100,000 East-Asian-like and 100,000 European-like individuals using HAPGEN2[25] (see URLs), which preserves population-specific MAF and LD patterns, starting from phased haplotypes of 481 East Asians and 489 Europeans individuals available in the 1000 Genomes Project[16] (see URLs), restricting to ∼2.5 million SNPs on chromosome 1 – 3 with minor allele count greater than 5 in either population. Since direct output of HAPGEN2 includes substantial relatedness,[2] we used PLINK2[87] (see URLs) to remove simulated individuals with genetic relatedness greater than 0.05, resulting in 35,378 EAS-like and 36,836 EUR-like individuals. From the filtered set of individuals, we randomly selected 500 individuals in each simulated population to serve as reference panels. We used 18,418 EAS-like and 36,836 EUR-like individuals to simulate GWAS summary statistics, capturing the imbalance

in sample size between EAS and EUR GWAS in analysis of real traits. In our secondary simulations, we also decreased or increased the reference panel size or decreased the GWAS sample size, to evaluate the robustness of our method with respect to reference panel size and GWAS sample size.

We performed both null simulations, where enrichment of squared trans-ethnic genetic correlation, $\lambda^2(C)$, is 1 across all functional annotations, and causal simulations, where $\lambda^2(C)$ varies across annotations, under various degrees of polygenicity (1%, 10%, and 100% causal SNPs). In the null simulations, we set τ_{1C}, τ_{2C}, θ_C to be the meta-analyzed τ_C in real-data analyses of EAS GWASs, and followed Equation (3) to obtain variance, $\mathrm{Var}[\beta_{1j}]$ and $\mathrm{Var}[\beta_{2j}]$, and covariance, $\mathrm{Cov}[\beta_{1j}, \beta_{2j}]$, of per-SNP causal effect sizes β_{1j}, β_{2j}, setting all negative per-SNP variance and covariance to 0. In the causal simulations, we directly specified per-SNP causal effect size variances and covariances using self-devised τ_{1C}, τ_{2C}, and θ_C coefficients, to attain $\lambda^2(C) \neq 1$, as these were difficult to attain using the coefficients from analyses of real traits.

We randomly selected a subset of SNPs to be causal for both populations, and set $\mathrm{Var}[\beta_{1j}]$, $\mathrm{Var}[\beta_{2j}]$, and $\mathrm{Cov}[\beta_{1j}, \beta_{2j}]$ to be 0 for all remaining non-causal SNPs. We scaled the trans-ethnic genetic covariance to attain a desired genome-wide r_g. Next, we drew causal effect sizes of each causal SNP j in the two populations from the bi-variate Gaussian distribution,

$$\begin{bmatrix} \beta_{1j} \\ \beta_{2j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathrm{Var}[\beta_{1j}] & \mathrm{Cov}[\beta_{1j}, \beta_{2j}] \\ \mathrm{Cov}[\beta_{1j}, \beta_{2j}] & \mathrm{Var}[\beta_{2j}] \end{bmatrix} \right), \tag{12}$$

and scaled the drawn effect sizes to match the desired total heritability and trans-ethnic genetic covariance. We also performed null simulations in which imperfect genome-wide trans-ethnic genetic correlation is due to population-specific causal variants. In these simulations, we randomly selected 10% of the SNPs to be causal in each population, with 80% of causal variants in each population shared with the other population, and sampled perfectly correlated causal effect sizes for shared causal variants using Equation (12). We simulated genetic component of the phenotype in population p as $\boldsymbol{X}_p\boldsymbol{\beta}_p$, where $\boldsymbol{X}_p$ is column-centered genotype matrix, and drew environmental effects, $\boldsymbol{\epsilon}_p$, from the Gaussian distribution, $N\left(0, 1 - \mathrm{Var}[\boldsymbol{X}_p\boldsymbol{\beta}_p]\right)$, such that the total phenotypic variance in each population is 1. Finally, we simulated GWAS summary association statistics for population p, $\boldsymbol{Z}_p$, as $Z_{pj} = \frac{\boldsymbol{X}_{pj}^{\intercal}\boldsymbol{Y}_p}{\sqrt{N_p}\sigma_{pj}}$, where σ_{pj} is the standard deviation of SNP j in population p. We have publicly released Python code for simulating GWAS summary statistics for 2 populations (see URLs). Fifth, we performed additional null simulations with annotation-dependent

MAF-dependent genetic architectures,[26–28] defined as architectures in which the level of MAF-dependence is annotation-dependent.

We also performed null simulations with annotation-dependent MAF-dependent genetic architectures,[26–28] defined as architectures in which the level of MAF-dependence is annotation-dependent, to assess the impact on estimates of enrichment of stratified squared trans-ethnic genetic correlation, ($\lambda^2(C)$). In these simulations, we set the variance of causal effect size of each SNP j in both populations to be proportional to $[p_{j,\max}(1 - p_{j,\max})]^{\alpha}$, where $p_{j,\max}$ is the maximum MAF of SNP j in the two populations. (We elected to use maximum MAF because a SNP that is rare in one population but common in the other is less likely to be impacted by negative selection.) We set α to -0.38, as previously estimated for 25 UK Biobank diseases and complex traits in ref.[28]. We sampled causal effect sizes using Equation (12), with $\mathrm{Var}[\beta_{1j}]$, $\mathrm{Var}[\beta_{2j}]$, and $\mathrm{Cov}[\beta_{1j}, \beta_{2j}]$ scaled to attain a desired genome-wide heritability and trans-ethnic genetic correlation. We randomly selected 10% of SNPs to be causal in both populations. Additionally, in the top quintile of background selection statistic, we selected 1.8× more low-frequency causal variants ($p_{j,\max} < 0.05$) than common variants ($p_{j,\max} \geqslant 0.05$), capturing the action of negative selection across low-frequency and common variants.[27]

Summary statistics for 31 diseases and complex traits

We analyzed GWAS summary statistics of 31 diseases and complex traits, primarily from UK Biobank,[74] Biobank Japan,[20] and CONVERGE.[18] All summary statistics were based on genotyping arrays with imputation to an appropriate LD reference panel (e.g. Haplotype Reference Consortium[89] and UK10K[90] for UK Biobank,[74] the 1000 Genomes Project[16] for Biobank Japan[20]), except those of the MDD GWAS in the East Asian population, which was based on low-coverage whole genome sequencing data.[18] These include: atrial fibrillation (AF),[91,92] age at menarche(AMN),[93,94] age at menopause (AMP),[93,94] basophil count(BASO),[20,95] body mass index (BMI),[20,96] blood sugar(BS),[20,96] diastolic blood pressure (DBP),[20,96] eosinophil count(EO),[20,96] estimated glomerular filtration rate (EGFR),[20,97] hemoglobin A1c(HBA1C),[20,96] height (HEIGHT),[96,98] high density lipoprotein (HDL),[20,96] hemoglobin (HGB),[20,95] hematocrit (HTC),[20,95] low density lipoprotein (LDL),[20,96] lymphocyte count(LYMPH),[20,96] mean corpuscular hemoglobin (MCH),[20,96] mean corpuscular hemoglobin concentration (MCHC),[20,95] mean corpuscular volume (MCV),[20,95] major depressive disorder (MDD),[18,99] monocyte count (MONO),[20,96] neutrophil count(NEUT),[20,95] platelet count (PLT),[20,96] rheumatoid arthritis(RA),[100] red blood cell count (RBC),[20,96] systolic blood pressure (SBP),[20,96] schizophrenia (SCZ)[101], type 2 diabetes (T2D),[102,103] total

cholesterol (TC),[20,96] triglyceride (TG),[20,96] and white blood cell count (WBC).[20,96] Further information for the GWAS summary statistics analyzed is provided in Table S16. In our main analyses, we performed random-effect meta-analysis to aggregate results across all 31 diseases and complex traits. To test if the meta-analyzed $\hat{\lambda}^2(C)$ is significantly different from 1, we computed a test statistic as $\frac{\hat{\lambda}^2(C)-1}{s.e.(\hat{\lambda}^2(C))}$, where $s.e.(\hat{\lambda}^2(C))$ is the standard error of meta-analyzed $\hat{\lambda}^2(C)$, and obtained a p-value under the normal distribution. We also defined a set of 20 approximately independent diseases and complex traits with cross-trait r_g^2 (estimated using cross-trait LDSC[71]) less than 0.25 in both populations: AF, AMN, AMP, BASO, BMI, EGFR, EO, HBA1C, HEIGHT, HTC, LYMPH, MCHC, MCV, MDD, NEUT, PLT, RA, SBP, TC, TG.

Expected enrichment of stratified squared trans-ethnic genetic correlation from 8 continuous-valued annotations

To obtain expected enrichment of squared trans-ethnic genetic correlation of a binary annotation C, $\lambda^2(C)$, from 8 continuous-valued annotations, we first fit the S-LDXR model using these 8 annotations together with the base annotation for all SNPs, yielding coefficients, $\tau_{1C'}$, $\tau_{2C'}$, and $\theta_{C'}$, for a total of 9 annotations. We then use Equation (3) to obtain per-SNP variance and covariance of causal effect sizes, β_{1j} and β_{1j}, substituting τ_{1C}, τ_{2C}, θ_C with $\tau_{1C'}$, $\tau_{2C'}$, and $\theta_{C'}$, respectively. We apply shrinkage with default parameter setting ($\alpha = 0.5$), and use Equation (9) and (10) to obtain expected stratified squared trans-ethnic genetic correlation, $r_g^2(C)$, and subsequently $\lambda^2(C)$.

Analysis of specifically expressed gene annotations

We obtained 53 specifically expressed gene (SEG) annotations, defined in ref.[24] as $\pm$100k-base-pair regions surrounding genes specifically expressed in each of 53 GTEx[32] tissues. A list of the SEG annotations is provided in Table S2. Correlations between SEG annotations and the 8 continuous-valued annotations are reported in Figure S28 and Table S2. Most SEG annotations are moderately correlated with the background selection statistic and CpG content annotations.

To test whether there is heterogeneity in enrichment of squared trans-ethnic genetic correlation, $\lambda^2(C)$, across the 53 SEG annotations, we first computed the average $\lambda^2(C)$ across the 53 annotations, $\bar{\lambda}^2(C)$, using fixed-effect meta-analysis. We then computed the test statistic $\sum_{i=1}^{53} \frac{\left(\hat{\lambda}^2(C_i)-\bar{\lambda}^2(C_i)\right)^2}{\mathrm{Var}[\hat{\lambda}^2(C_i)]}$, where C_i is the i-th SEG annotation, and $\hat{\lambda}^2(C_i)$ the estimated $\lambda^2(C)$. We computed a p-value for this test statistic based on a χ^2 distribution with 53

degrees of freedom.

Analysis of distance to nearest exon annotation

We created a continuous-valued annotation, named "distance to nearest exon annotation", based on a SNP's physical distance (number of base pairs) to its nearest exon, using 233,254 exons defined on the UCSC genome browser[104] (see URLs). This annotation is moderately correlated with the background selection statistic annotation[22] ($R = -0.21$), defined as (1 - McVicker B statistic / 1000), where the McVicker B statistic quantifies a site's genetic distance to its nearest exon.[30] We have publicly released this annotation (see URLs).

To assess the informativeness of functionally important regions versus regions impacted by selection in explaining the depletions of squared trans-ethnic genetic correlation, we applied S-LDXR on the distance to nearest exon annotation together with the baseline-LD-X model annotations. We used both enrichment of squared trans-ethnic genetic correlation ($\lambda^2(C)$) and standardized annotation effect size (τ^*_{1C}, τ^*_{2C}, and θ^*_C) to assess informativeness.

Analysis of probability of loss-of-function intolerance decile gene annotations

We created 10 annotations based on genes in deciles of probability of being loss-of-function intolerant (pLI) (see URLs), defined as the probability of assigning a gene into haplosufficient regions, where protein-truncating variants are depleted.[56] Genes with high pLI (e.g. > 0.9) have higly constrained functionality, and therefore mutations in these genes are subject to negative selection. We included SNPs within a 100kb-base-pair window around each gene, following ref.[24] A correlation heat map between pLI decile gene annotations and the 8 continuous-valued annotations is provided in Figure S29. All pLI decile gene annotations are moderately correlated with the background selection statistic and CpG content annotations.

Analysis of the integrated haplotype score annotation

We created a binary annotation (proportion of SNPs: 6.3%) that includes all SNPs whose maximum absolute value of the integrated haplotype score (iHS)[43,44] (see URLs) across all 1000 Genomes Project EAS and EUR sub-populations are greater than 2.0, the recommended threshold to detect positive selection in ref.[43]. This annotation is positively correlated with the top quintile of the background selection statistic annotation ($R = 0.077$). We note that although the iHS is a recombination-rate-adjusted quantity to detect the action

of recent positive selection, it may also capture actions of negative selection.[43,44]

Acknowledgements

We are grateful to L. O'Connor, H. Finucane, D. Kassler, S. Mallick, N. Patterson, B. Neale, R. Walters, A. Martin, B. Brown, F. Hormozdiari, M. Hujoel, K. Burch, and B. Pasaniuc for helpful discussions. This research was conducted using the UK Biobank Resource under Application 16549 and was funded by NIH grants R01 HG006399, U01 HG009379, R37 MH107649, R01 MH101244, and R01 CA222147.

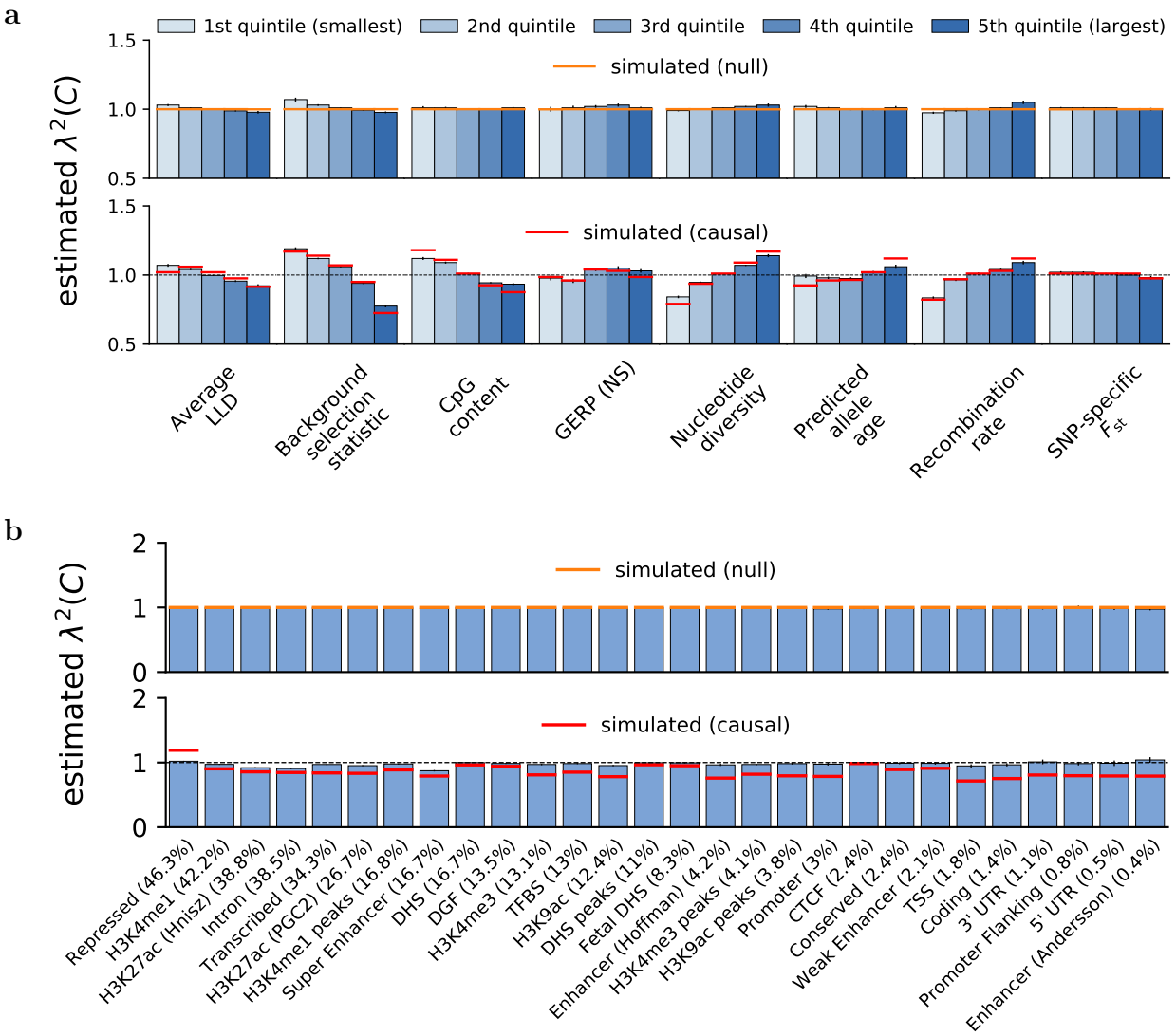

Figure 1: **Accuracy of S-LDXR in null and causal simulations.** We report estimates of the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$) in both null and causal simulations, for (a) quintiles of 8 continuous-valued annotations and (b) 28 main binary annotations (sorted by proportion of SNPs, displayed in parentheses). Results are averaged across 1,000 simulations. Error bars denote $\pm 1.96\times$ standard error. Numerical results are reported in Table S5 and S8.

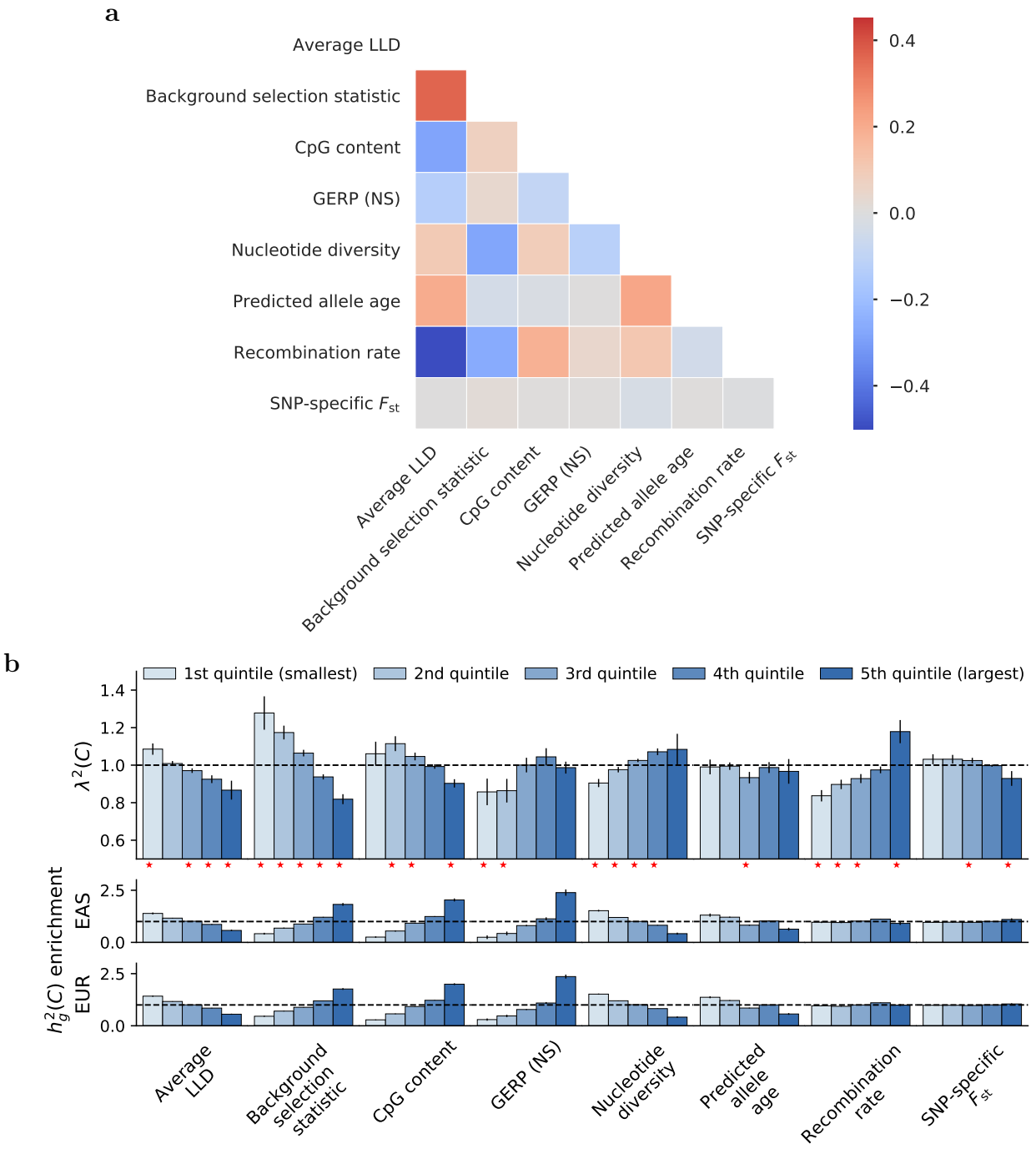

Figure 2: **S-LDXR results for quintiles of 8 continuous-valued annotations across 31 diseases and complex traits.** (a) We report correlations between each continuous-valued annotation; diagonal entries are not shown. Numerical results are reported in Table S1. (b) We report estimates of the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$), as well as population-specific estimates of heritability enrichment, for quintiles of each continuous-valued annotation. Results are meta-analyzed across 31 diseases and complex traits. Error bars denote $\pm 1.96\times$ standard error. Red stars ($\star$) denote two-tailed $p<0.05/40$. Numerical results are reported in Table S17.

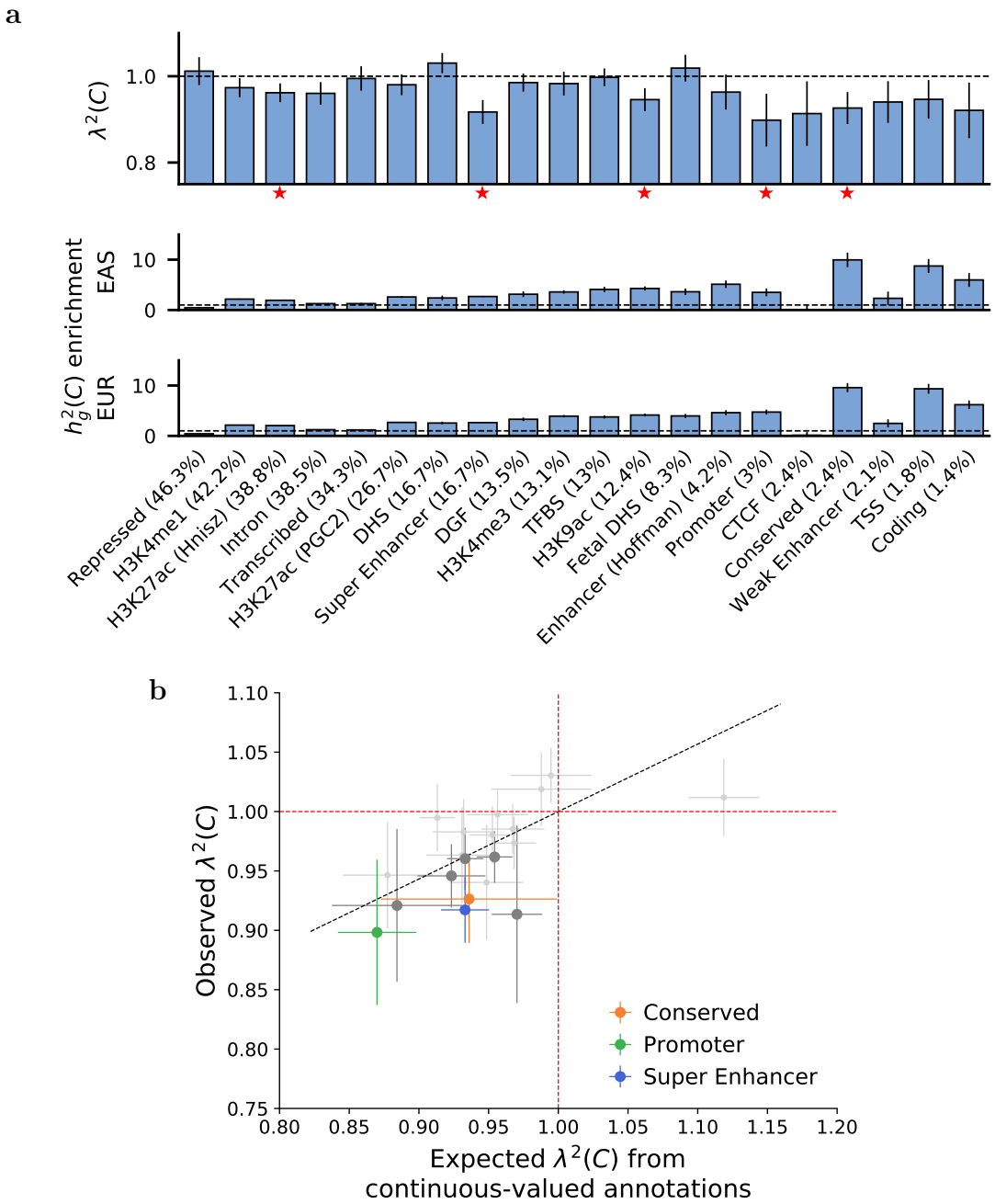

Figure 3: **S-LDXR results for 20 binary functional annotations across 31 diseases and complex traits.** (a) We report estimates of the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$), as well as population-specific estimates of heritability enrichment, for each binary annotation (sorted by proportion of SNPs, displayed in parentheses). Results are meta-analyzed across 31 diseases and complex traits. Error bars denote $\pm 1.96\times$ standard error. Red stars ($\star$) denote two-tailed p<0.05/20. Numerical results are reported in Table S18. (b) We report observed $\lambda^2(C)$ vs. expected $\lambda^2(C)$ based on 8 continuous-valued annotations, for each binary annotation. Results are meta-analyzed across 31 diseases and complex traits. Error bars denote $\pm 1.96\times$ standard error. Annotations for which $\lambda^2(C)$ is significantly different from 1 (p<0.05/20) are denoted in color (see legend) or dark gray. The dashed black line (slope=0.57) denotes a regression of observed $\lambda(C) - 1$ vs. expected $\lambda(C) - 1$ with intercept constrained to 0. Numerical results are reported in Table S20.

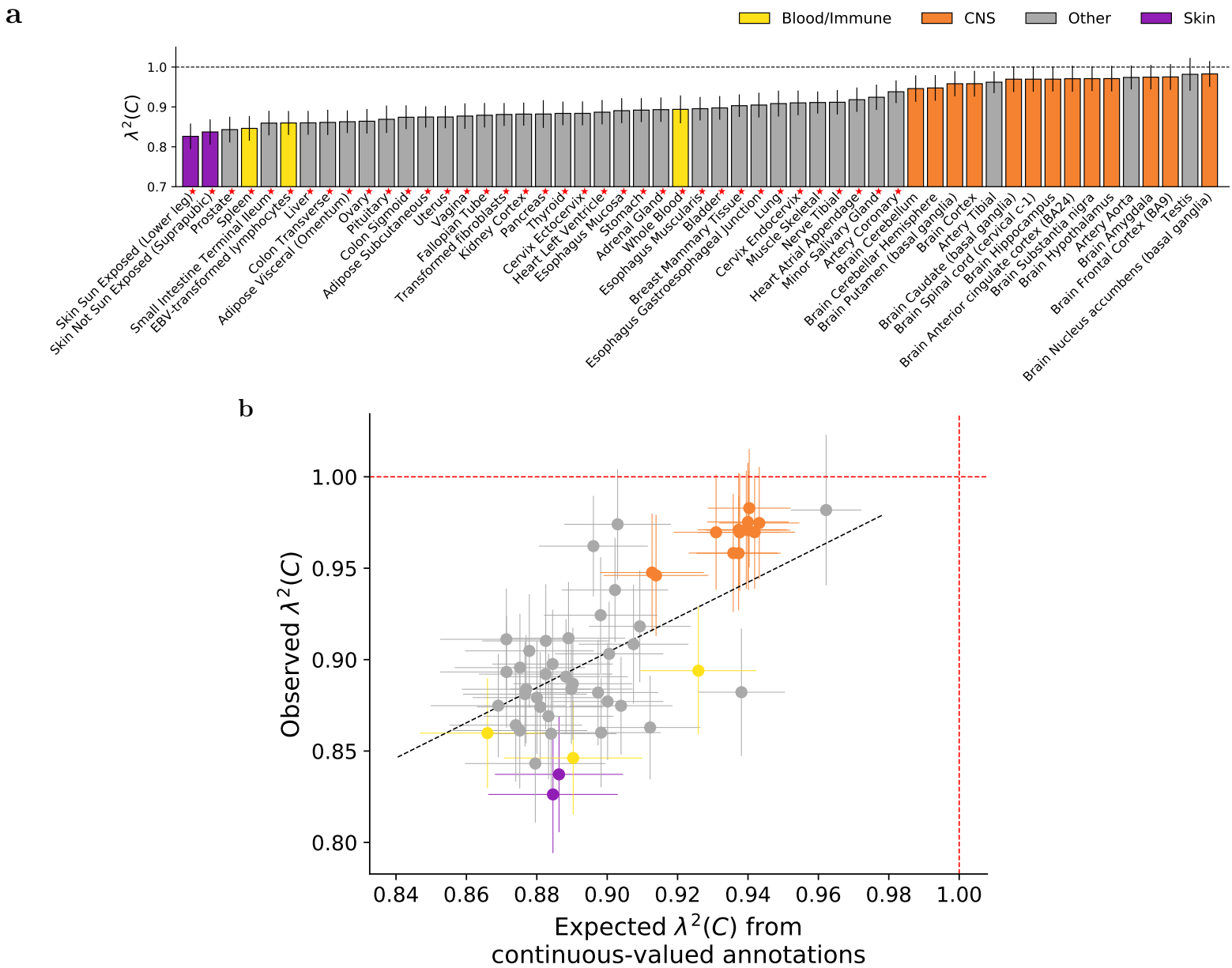

Figure 4: **S-LDXR results for 53 specifically expressed gene (SEG) annotations across 31 diseases and complex traits.** (a) We report estimates of the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$) for each SEG annotation (sorted by $\lambda^2(C)$). Results are meta-analyzed across 31 diseases and complex traits. Error bars denote $\pm 1.96\times$ standard error. Red stars (⋆) denote two-tailed p<0.05/53. Numerical results are reported in Table S21. (b) We report observed $\lambda^2(C)$ vs. expected $\lambda^2(C)$ based on 8 continuous-valued annotations, for each SEG annotation. Results are meta-analyzed across 31 diseases and complex traits. Error bars denote $\pm 1.96\times$ standard error. Annotations are color-coded as in (a). The dashed black line (slope=0.96) denotes a regression of observed $\lambda(C) - 1$ vs. expected $\lambda(C) - 1$ with intercept constrained to 0. Numerical results and population-specific heritability enrichment estimates are reported in Table S23.

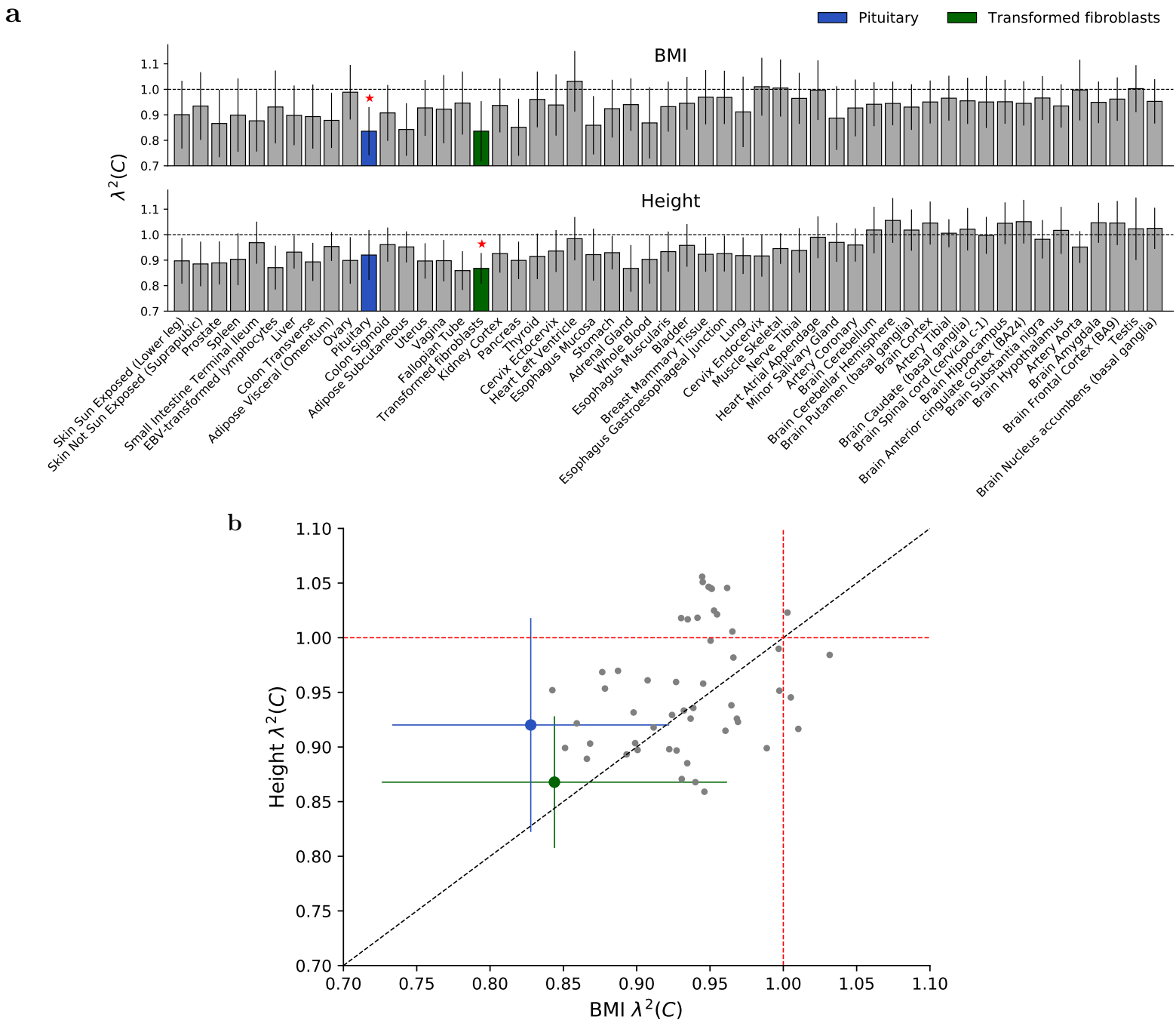

Figure 5: **S-LDXR results for 53 specifically expressed gene (SEG) annotations for BMI and height.** (a) We report estimates of the enrichment/depletion of squared trans-ethnic genetic correlation ($\lambda^2(C)$) for each SEG annotation for BMI and height. SEG annotations are ordered as in Figure 4. Error bars denote $\pm 1.96\times$ standard error. Red stars ($\star$) denote two-tailed $p < 0.05/53$ for each respective trait. (b) We report $\lambda^2(C)$ estimates for height vs. BMI for each SEG annotation. These estimates were moderately correlated ($R = 0.35$). Annotations are color-coded as in (a). For Pituitary, height $\lambda^2(C)$ is left-shifted by 0.008 and BMI $\lambda^2(C)$ is right-shifted by 0.008 for easier visualization of standard errors. Error bars denote $\pm 1.96\times$ standard error. The dashed black line denotes the y vs x line. Numerical results for all 31 diseases and complex traits are reported in Table S24.

References

[1] Teresa R de Candia et al. "Additive genetic variation in schizophrenia risk is shared by populations of African and European descent". In: *The American Journal of Human Genetics* 93.3 (2013), pp. 463–470.

[2] Brielin C Brown et al. "Transethnic genetic-correlation estimates from summary statistics". In: *The American Journal of Human Genetics* 99.1 (2016), pp. 76–88.

[3] Nicholas Mancuso et al. "The contribution of rare variation to prostate cancer heritability". In: *Nature genetics* 48.1 (2016), p. 30.

[4] Masashi Ikeda et al. "Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect". In: *Schizophrenia bulletin* 45.4 (2018), pp. 824–834.

[5] Kevin J Galinsky et al. "Estimating cross-population genetic correlations of causal effect sizes". In: *Genetic epidemiology* 43.2 (2019), pp. 180–188.

[6] Alicia R Martin et al. "Clinical use of current polygenic risk scores may exacerbate health disparities". In: *Nature genetics* 51.4 (2019), p. 584.

[7] Christopher S Carlson et al. "Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study". In: *PLoS biology* 11.9 (2013), e1001661.

[8] Alicia R Martin et al. "Human demographic history impacts genetic risk prediction across diverse populations". In: *The American Journal of Human Genetics* 100.4 (2017), pp. 635–649.

[9] Carla Márquez-Luna et al. "Multiethnic polygenic risk scores improve risk prediction in diverse populations". In: *Genetic epidemiology* 41.8 (2017), pp. 811–823.

[10] Genevieve L Wojcik et al. "Genetic analyses of diverse populations improves discovery for complex traits". In: *Nature* (2019).

[11] L Duncan et al. "Analysis of polygenic risk score usage and performance in diverse human populations". In: *Nature Communications* 10.1 (2019), p. 3328.

[12] Kevin L Keys et al. "On the cross-population portability of gene expression prediction models". In: *bioRxiv* (2019), p. 552042.

[13] Sang Hong Lee et al. "Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood". In: *Bioinformatics* 28.19 (2012), pp. 2540–2542.

[14] Giorgio Sirugo, Scott M Williams, and Sarah A Tishkoff. "The missing diversity in human genetic studies". In: *Cell* 177.1 (2019), pp. 26–31.

[15] Deepti Gurdasani et al. "Genomics of disease risk in globally diverse populations". In: *Nature Reviews Genetics* (2019).

[16] 1000 Genomes Project Consortium et al. "A global reference for human genetic variation". In: *Nature* 526.7571 (2015), p. 68.

[17] International HapMap 3 Consortium et al. "Integrating common and rare genetic variation in diverse human populations". In: *Nature* 467.7311 (2010), p. 52.

[18] Na Cai et al. "Sparse whole-genome sequencing identifies two loci for major depressive disorder". In: *Nature* 523.7562 (2015), p. 588.

[19] Akiko Nagai et al. "Overview of the BioBank Japan Project: study design and profile". In: *Journal of epidemiology* 27.Supplement_III (2017), S2–S8.

[20] Masahiro Kanai et al. "Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases". In: *Nature genetics* 50.3 (2018), p. 390.

[21] Hilary K Finucane et al. "Partitioning heritability by functional annotation using genome-wide association summary statistics". In: *Nature genetics* 47.11 (2015), p. 1228.

[22] Steven Gazal et al. "Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection". In: *Nature genetics* 49.10 (2017), p. 1421.

[23] Steven Gazal et al. "Reconciling S-LDSC and LDAK functional enrichment estimates". In: *Nature genetics* 51.8 (2019), pp. 1202–1204.

[24] Hilary K Finucane et al. "Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types". In: *Nature genetics* 50.4 (2018), p. 621.

[25] Zhan Su, Jonathan Marchini, and Peter Donnelly. "HAPGEN2: simulation of multiple disease SNPs". In: *Bioinformatics* 27.16 (2011), pp. 2304–2305.

[26] Jian Zeng et al. "Signatures of negative selection in the genetic architecture of human complex traits". In: *Nature genetics* 50.5 (2018), p. 746.

[27] S Gazal et al. "Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations." In: *Nature genetics* 50.11 (2018), p. 1600.

[28] Armin P Schoech et al. "Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection". In: *Nature communications* 10.1 (2019), p. 790.

[29] Na Cai, Kenneth Kendler, and Jonathan Flint. "Minimal phenotyping yields GWAS hits of low specificity for major depression". In: *BioRxiv* (2018), p. 440735.

[30] Graham McVicker et al. "Widespread genomic signatures of natural selection in hominid evolution". In: *PLoS genetics* 5.5 (2009), e1000471.

[31] Gleb Kichaev and Bogdan Pasaniuc. "Leveraging functional-annotation data in trans-ethnic fine-mapping studies". In: *The American Journal of Human Genetics* 97.2 (2015), pp. 260–271.

[32] GTEx Consortium et al. "Genetic effects on gene expression across human tissues". In: *Nature* 550.7675 (2017), p. 204.

[33] Soumya Raychaudhuri et al. "Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function". In: *PLoS genetics* 6.9 (2010), e1001097.

[34] Pardis C Sabeti et al. "Positive natural selection in the human lineage". In: *science* 312.5780 (2006), pp. 1614–1620.

[35] Rasmus Nielsen et al. "Recent and ongoing selection in the human genome". In: *Nature Reviews Genetics* 8.11 (2007), p. 857.

[36] John Novembre and Anna Di Rienzo. "Spatial patterns of variation due to natural selection in humans". In: *Nature Reviews Genetics* 10.11 (2009), p. 745.

[37] Kevin N Laland, John Odling-Smee, and Sean Myles. "How culture shaped the human genome: bringing genetics and the human sciences together". In: *Nature Reviews Genetics* 11.2 (2010), p. 137.

[38] Sandra Wilde et al. "Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y". In: *Proceedings of the National Academy of Sciences* 111.13 (2014), pp. 4832–4837.

[39] Harald von Boehmer. "Positive selection of lymphocytes". In: *Cell* 76.2 (1994), pp. 219–228.

[40] Jingjing Li et al. "Natural selection has differentiated the progesterone receptor among human populations". In: *The American Journal of Human Genetics* 103.1 (2018), pp. 45–57.

[41] Luke J O'Connor et al. "Extreme Polygenicity of Complex Traits Is Explained by Negative Selection". In: *The American Journal of Human Genetics* (2019).

[42] Krishna R Veeramah and Michael F Hammer. "The impact of whole-genome sequencing on the reconstruction of human population history". In: *Nature Reviews Genetics* 15.3 (2014), p. 149.

[43] Benjamin F Voight et al. "A map of recent positive selection in the human genome". In: *PLoS biology* 4.3 (2006).

[44] Kelsey Elizabeth Johnson and Benjamin F Voight. "Patterns of shared signatures of recent positive selection across human populations". In: *Nature ecology & evolution* 2.4 (2018), pp. 713–720.

[45] Jenny van Dongen and Dorret I Boomsma. "The evolutionary paradox and the missing heritability of schizophrenia". In: *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 162.2 (2013), pp. 122–136.

[46] Antonio F Pardiñas et al. "Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection". In: *Nature genetics* 50.3 (2018), pp. 381–389.

[47] Valentina Vicennati and Renato Pasquali. "Abnormalities of the hypothalamic-pituitary-adrenal axis in nondepressed women with abdominal obesity and relations with insulin resistance: evidence for a central and a peripheral alteration". In: *The Journal of Clinical Endocrinology & Metabolism* 85.11 (2000), pp. 4093–4098.

[48] AN Vgontzas et al. "Hypothalamic-pituitary-adrenal axis activity in obese men with and without sleep apnea: effects of continuous positive airway pressure therapy". In: *The Journal of Clinical Endocrinology & Metabolism* 92.11 (2007), pp. 4199–4207.

[49] Mousumi Bose, Blanca Oliván, and Blandine Laferrère. "Stress and obesity: the role of the hypothalamic–pituitary–adrenal axis in metabolic disease". In: *Current opinion in endocrinology, diabetes, and obesity* 16.5 (2009), p. 340.

[50] Nobuyuki Itoh and David M Ornitz. "Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease". In: *The Journal of Biochemistry* 149.2 (2011), pp. 121–130.

[51] Matthew R Robinson et al. "Genotype–covariate interaction effects and the heritability of adult body mass index". In: *Nature genetics* 49.8 (2017), p. 1174.

[52] William G Hill, Michael E Goddard, and Peter M Visscher. "Data and theory point to mainly additive genetic variance for complex traits". In: *PLoS genetics* 4.2 (2008), e1000008.

[53] Asko Mäki-Tanila and William G Hill. "Influence of gene interaction on complex trait variation with multilocus models". In: *Genetics* 198.1 (2014), pp. 355–367.

[54] Zhihong Zhu et al. "Dominance genetic variation contributes little to the missing heritability for human complex traits". In: *The American Journal of Human Genetics* 96.3 (2015), pp. 377–385.

[55] Maaike de Jong et al. "Natural variation in Arabidopsis shoot branching plasticity in response to nitrate supply affects fitness". In: *PLoS genetics* 15.9 (2019), e1008366.

[56] Monkol Lek et al. "Analysis of protein-coding genetic variation in 60,706 humans". In: *Nature* 536.7616 (2016), p. 285.

[57] Adam Eyre-Walker. "Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies". In: *Proceedings of the National Academy of Sciences* (2010), p. 200906182.

[58] International Schizophrenia Consortium et al. "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder". In: *Nature* 460.7256 (2009), p. 748.

[59] Eli A Stahl et al. "Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis". In: *Nature genetics* 44.5 (2012), p. 483.

[60] Po-Ru Loh et al. "Efficient Bayesian mixed-model analysis increases association power in large cohorts". In: *Nature genetics* 47.3 (2015), p. 284.

[61] Bjarni J Vilhjálmsson et al. "Modeling linkage disequilibrium increases accuracy of polygenic risk scores". In: *The american journal of human genetics* 97.4 (2015), pp. 576–592.

[62] Yiming Hu et al. "Leveraging functional annotations in genetic risk prediction for human complex diseases". In: *PLoS computational biology* 13.6 (2017), e1005589.

[63] Tian Ge et al. "Polygenic prediction via Bayesian regression and continuous shrinkage priors". In: *Nature communications* 10.1 (2019), pp. 1–10.

[64] Wonil Chung et al. "Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes". In: *Nature communications* 10.1 (2019), pp. 1–11.

[65] Luke R Lloyd-Jones et al. "Improved polygenic prediction by Bayesian multiple regression on summary statistics". In: *Nature communications* 10.1 (2019), pp. 1–11.

[66] Carla Márquez-Luna et al. "LDpred-funct: incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets". In: *bioRxiv* (2020), p. 375337.

[67] Daniel J Schaid, Wenan Chen, and Nicholas B Larson. "From genome-wide associations to candidate causal variants by statistical fine-mapping". In: *Nature Reviews Genetics* 19.8 (2018), pp. 491–504.

[68] Huwenbo Shi et al. "Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data". In: *The American Journal of Human Genetics* (2020).

[69] Andrew P Morris. "Transethnic meta-analysis of genomewide association studies". In: *Genetic epidemiology* 35.8 (2011), pp. 809–822.

[70] Patrick Turley et al. "Multi-trait analysis of genome-wide association summary statistics using MTAG". In: *Nature genetics* 50.2 (2018), p. 229.

[71] Brendan Bulik-Sullivan et al. "An atlas of genetic correlations across human diseases and traits". In: *Nature genetics* 47.11 (2015), p. 1236.

[72] Qiongshi Lu et al. "A powerful approach to estimating annotation-stratified genetic covariance via GWAS summary statistics". In: *The American Journal of Human Genetics* 101.6 (2017), pp. 939–964.

[73] Michael F Seldin, Bogdan Pasaniuc, and Alkes L Price. "New approaches to disease mapping in admixed populations". In: *Nature Reviews Genetics* 12.8 (2011), p. 523.

[74] Clare Bycroft et al. "The UK Biobank resource with deep phenotyping and genomic data". In: *Nature* 562.7726 (2018), p. 203.

[75] Brendan K Bulik-Sullivan et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". In: *Nature genetics* 47.3 (2015), p. 291.

[76] Yang Luo et al. "Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations". In: *bioRxiv* (2019), p. 503144.

[77] Alicia R Martin et al. "Transcriptome sequencing from diverse human populations reveals differentiated regulatory architecture". In: *PLoS genetics* 10.8 (2014), e1004549.

[78] Lauren S Mogil et al. "Genetic architecture of gene expression traits across diverse populations". In: *PLoS genetics* 14.8 (2018), e1007586.

[79] Arun Durvasula and Kirk E Lohmueller. "Negative selection on complex traits limits genetic risk prediction accuracy between populations". In: *bioRxiv* (2019), p. 721936.

[80] JH Curtiss. "On the distribution of the quotient of two chance variables". In: *The Annals of Mathematical Statistics* 12.4 (1941), pp. 409–421.

[81] ENCODE Project Consortium et al. "An integrated encyclopedia of DNA elements in the human genome". In: *Nature* 489.7414 (2012), p. 57.

[82] Anshul Kundaje et al. "Integrative analysis of 111 reference human epigenomes". In: *Nature* 518.7539 (2015), p. 317.

[83] Maya Kasowski et al. "Extensive variation in chromatin states across humans". In: *Science* 342.6159 (2013), pp. 750–752.

[84] Eugene V Davydov et al. "Identifying a high fraction of the human genome to be under selective constraint using GERP++". In: *PLoS computational biology* 6.12 (2010), e1001025.

[85] Simon Myers et al. "A fine-scale map of recombination rates and hotspots across the human genome". In: *Science* 310.5746 (2005), pp. 321–324.

[86] Matthew D Rasmussen et al. "Genome-wide inference of ancestral recombination graphs". In: *PLoS genetics* 10.5 (2014), e1004342.

[87] Christopher C Chang et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets". In: *Gigascience* 4.1 (2015), p. 7.

[88] Bruce S Weir and C Clark Cockerham. "Estimating F-statistics for the analysis of population structure". In: *evolution* 38.6 (1984), pp. 1358–1370.

[89] Shane McCarthy et al. "A reference panel of 64,976 haplotypes for genotype imputation". In: *Nature genetics* 48.10 (2016), pp. 1279–1283.

[90] UK10K consortium et al. "The UK10K project identifies rare variants in health and disease". In: *Nature* 526.7571 (2015), pp. 82–90.

[91] Siew-Kee Low et al. "Identification of six new genetic loci associated with atrial fibrillation in the Japanese population". In: *Nature genetics* 49.6 (2017), p. 953.

[92] Jonas B Nielsen et al. "Biobank-driven genomic discovery yields new insight into atrial fibrillation biology". In: *Nature genetics* 50.9 (2018), p. 1234.

[93] Momoko Horikoshi et al. "Elucidating the genetic architecture of reproductive ageing in the Japanese population". In: *Nature communications* 9.1 (2018), p. 1977.

[94] Felix R Day et al. "Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair". In: *Nature genetics* 47.11 (2015), p. 1294.

[95] William J Astle et al. "The allelic landscape of human blood cell trait variation and links to common complex disease". In: *Cell* 167.5 (2016), pp. 1415–1429.

[96] Po-Ru Loh et al. "Mixed-model association for biobank-scale datasets". In: *Nature genetics* 50.7 (2018), p. 906.

[97] Cristian Pattaro et al. "Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function". In: *Nature communications* 7 (2016), p. 10023.

[98] Masato Akiyama et al. "Characterizing rare and low-frequency height-associated variants in the Japanese population". In: *Nature Communications* 10.1 (2019), pp. 1–11.

[99] Naomi R Wray et al. "Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression". In: *Nature genetics* 50.5 (2018), p. 668.

[100] Yukinori Okada et al. "Genetics of rheumatoid arthritis contributes to biology and drug discovery". In: *Nature* 506.7488 (2014), p. 376.

[101] Max Lam et al. "Comparative genetic architectures of schizophrenia in East Asian and European populations". In: *Nature genetics* 51.12 (2019), pp. 1670–1678.

[102] Ken Suzuki et al. "Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population". In: *Nature genetics* 51.3 (2019), p. 379.

[103] Robert A Scott et al. "An expanded genome-wide association study of type 2 diabetes in Europeans". In: *Diabetes* 66.11 (2017), pp. 2888–2902.

[104] Donna Karolchik, Angie S Hinrichs, and W James Kent. "The UCSC genome browser". In: *Current protocols in bioinformatics* 40.1 (2012), pp. 1–4.